MEASURE–VALUED DIFFERENTIATION FOR FINITE PRODUCTS OF MEASURES: THEORY AND APPLICATIONS

ISBN 978 90 5170 905 6

Cover design: Crasborn Graphic Designers bno, Valkenburg a.d. Geul

This book is no. 428 of the Tinbergen Institute Research Series, established through cooperation between Thela Thesis and the Tinbergen Institute. A list of books which already appeared in the series can be found in the back.

VRIJE UNIVERSITEIT

MEASURE–VALUED DIFFERENTIATION FOR FINITE PRODUCTS OF MEASURES: THEORY AND APPLICATIONS

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. L. M. Bouter,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de faculteit der Economische Wetenschappen en Bedrijfskunde
op maandag 22 september 2008 om 13.45 uur
in de aula van de universiteit,
De Boelelaan 1105

door
Haralambie Leahu
geboren te Galaţi, Roemenië

promotor:	prof.dr. H.C. Tijms
copromotor:	dr. B.F. Heidergott

TO MY PARENTS

CONTENTS

PREFACE

A wide range of stochastic systems in the area of manufacturing, transportation, finance and communication can be modeled by studying cost-functions[1] over a finite collection of independent random variables, called *input variables*. From a probabilistic point of view such a system is completely determined by the distributions of the input variables under consideration, which will be called *input distributions*. Throughout this thesis we consider parameter-dependent stochastic systems, i.e., we assume that the input distributions depend on some real-valued parameter denoted by θ. More specifically, let $\Theta \subset \mathbb{R}$ denote an open, connected subset of real numbers and let $\mu_{i,\theta}$, for $1 \leq i \leq n$, be a finite family of probability measures (input distributions) on some state spaces $\mathbb{S}_i$, for $1 \leq i \leq n$, depending on some parameter $\theta \in \Theta$, such as, for example, the mean. We consider a stochastic system driven by the above specified distributions and we call a *performance measure* of such a system the expression

$$P_g(\theta) := \mathbb{E}_\theta[g(X_1, \ldots, X_n)] = \int \ldots \int g(x_1, \ldots, x_n)\Pi_\theta(dx_1, \ldots, dx_n), \tag{0.1}$$

for an arbitrary cost-function g, where the input variables X_i, for $1 \leq i \leq n$, are distributed according to $\mu_{i,\theta}$, respectively, and Π_θ denotes the product measure

$$\forall \theta \in \Theta : \ \Pi_\theta := \mu_{1,\theta} \times \ldots \times \mu_{n,\theta}. \tag{0.2}$$

This thesis is devoted to the analysis of performance measures modeled in (0.1). This class of models covers a wide area of applications such as queueing theory, project evaluation and review technique (PERT), which provide suitable models for manufacturing or transportation networks, and insurance models. Specifically, the following concrete models will be treated as examples: single-server queueing networks, stochastic activity networks and insurance models over a finite number of claims. Correspondingly, transient waiting times in queueing networks, completion times in stochastic activity networks or ruin probabilities in insurance models are examples of performance measures.

The main topic of research put forward in this thesis will be the study of analytical properties of the performance measures $P_g(\theta)$ such as continuity, differentiability and analyticity with respect to the parameter θ, for g belonging to some pre-specified class of cost-functions $\mathcal{D}$. This allows for a wide range of applications such as gradient estimation (which very often is an useful tool for performing stochastic optimization), sensitivity analysis (bounds on perturbations) or Taylor series approximations. To this end, we study the distribution Π_θ of the vector $(X_1, \ldots, X_n)$ rather than investigating each $P_g(\theta)$ individually, i.e., we study *weak properties* of the probability measure Π_θ. More specifically, if

[1] Real-valued functions designed to measure some specific performance of the system.

$\mathcal{D}$ is a set of cost-functions, we say that a property (**P**) (e.g., continuity, differentiability) holds weakly, in a $\mathcal{D}$-sense, for the measure-valued mapping $\theta \mapsto \mu_\theta$ if for each $g \in \mathcal{D}$ the mapping $\theta \mapsto \int g d\mu_\theta$ has the same property (**P**). It turns out that one can simultaneously handle the whole class of performance measures $\{P_g(\theta) : g \in \mathcal{D}\}$.

We propose here a modular approach to the analysis of Π_θ, explained in the following. Let us identify the original stochastic process with the product measure Π_θ defined in (0.2). Assume further that the input distributions $\mu_{i,\theta}$ are weakly $\mathcal{D}$-differentiable, for each $1 \le i \le n$. Then we show that the product probability measure Π_θ is weakly differentiable and it follows that $P_g(\theta)$ is differentiable with respect to θ, for each $g \in \mathcal{D}$. In addition, there exist a finite collection of "parallel processes", $\{\Pi^l_\theta : 1 \le l \le 2n\}$, which have the same physical interpretation as the original process but differ from that by (at most) one input distribution, such that for each $g \in \mathcal{D}$ we have

$$P'_g(\theta) = \frac{d}{d\theta} \int g(\mathbf{x}) \Pi_\theta(d\mathbf{x}) = \sum_{l=1}^{2n} \beta_{l,\theta} \int g(\mathbf{x}) \Pi^l_\theta(d\mathbf{x}) = \sum_{l=1}^{2n} \beta_{l,\theta} P^l_g(\theta), \tag{0.3}$$

for some constants $\beta_{l,\theta}$ which do not depend on g, where $\mathbf{x} := (x_1, \dots, x_n)$ denotes a sample path of the process and, for $1 \le l \le 2n$, $P^l_g(\theta)$ denotes the counterpart of $P_g(\theta)$ in the process driven by Π^l_θ. Therefore, in accordance with (0.3), one can evaluate the derivative of the performance measure $P_g(\theta)$ as a linear combination of the corresponding performance measures $P^l_g(\theta)$ in some parallel processes. In particular, if $\hat{P}^l_g$ is an unbiased estimator for $P^l_g(\theta)$, for each $1 \le l \le 2n$, then

$$\hat{\partial}_\theta(P_g) := \sum_{l=1}^{2n} \beta_{l,\theta} \hat{P}^l_g \tag{0.4}$$

provides an unbiased estimator for $P'_g(\theta)$. As it will turn out, a similar procedure can be applied for evaluating higher-order derivatives of $P_g(\theta)$.

The concept of weak differentiation has been first introduced in [47] for $\mathcal{D}$ consisting of bounded and continuous performance measures and studied further in [48]. Although consistent with classical convergence of probability measures, which induces convergence in distribution for random variables, this approach has a major pitfall. Namely, it can not deal with unbounded cost-functions such as, for instance, the identity mapping. Therefore, the concept was extended to general classes of cost-functions in [32] where it has been shown that weak differentiation provides unbiased gradient estimators.

In this thesis we aim to develop a weak differential calculus for measures (measure-valued differential calculus). More specifically, if $\mathcal{D}$ denotes a class of real-valued mappings on some "well-behaved" metric space $\mathbb{S}$ then for any continuous, non-negative mapping $v : \mathbb{S} \to \mathbb{R}$ one can define the subsequent class $[\mathcal{D}]_v$ of v-bounded mappings as follows:

$$[\mathcal{D}]_v := \{g \in \mathcal{D} : \ \exists c > 0 \ s.t. \ \forall s \in \mathbb{S} : \ |g(s)| \le c\, v(s)\}. \tag{0.5}$$

It turns out that if $\mathcal{D}$ denotes the class of either continuous or measurable mappings on $\mathbb{S}$ then $[\mathcal{D}]_v$ defined by (0.5) becomes a Banach space when endowed with the so-called *v-norm* $\|\cdot\|_v$ given by

$$\forall g \in \mathcal{D} : \ \|g\|_v := \sup_{s \in \mathbb{S}} \frac{|g(s)|}{v(s)}. \tag{0.6}$$

The pair $(\mathcal{D}, v)$ will be called a *Banach base* on $\mathbb{S}$ and will serve as a basis for defining weak differentiability and, more generally, weak properties. Therefore, in order to establish a solid mathematical background to support our theory, we appeal to a rather advanced mathematical machinery. More specifically, starting from the observation that regular measures on metric spaces appear as continuous linear functionals on some functional (Banach) spaces, e.g., $[\mathcal{D}]_v$, we apply standard results from functional analysis in order to derive fruitful results for weak differentiation theory. For instance, if we identify a measure with a linear functional on the Banach space $[\mathcal{D}]_v$ then weak convergence of measures is equivalent to the convergence in the weak topology induced by $[\mathcal{D}]_v$ on its topological dual $[\mathcal{D}]_v^*$. In addition one can define a strong (norm) topology on the space of measures by using the *operator v-norm* defined as

$$\forall \mu \in [\mathcal{D}]_v^* : \ \|\mu\|_v := \sup_{\|g\|_v \le 1} \left| \int g(s)\mu(ds) \right|, \tag{0.7}$$

where $\|g\|_v$ is defined by (0.6). It will turn out that classical theorems such as the Banach-Steinhaus Theorem and the Banach-Alaoglu Theorem will perfectly fit into this setting.

The material in this thesis is organized into five chapters and it is largely based on the results put forward in [22], [23], [26] and [28]. However, this dissertation does not reduce to a simple concatenation of the results in the above papers but it is rather a monograph on weak differentiation of measures, and applications, which, for the sake of the completeness of the theory, includes some results which were not presented in the aforementioned papers. In Chapter 1 we provide a detailed overview of basic concepts and preliminary results which are used to develop a weak differentiation theory. Although most of these facts can be found in any standard text book on topology, measure theory or functional analysis, we think that a small compendium of mathematical analysis would be helpful for the reader. Apart from that, some new concepts, such as Banach base, which will be later used to formalize the concept of weak differentiation, are introduced and studied. Moreover, the theory of weak convergence of sequences of signed measures is developed in Chapter 1. More specifically, sufficient conditions for both weak $[\mathcal{D}]_v$-convergence of measures and weak convergence of positive and negative parts of signed measures are treated.

In Chapter 2 several types of measure-valued differentiation, among which weak differentiation plays a key role, are discussed. It turns out that, in some situations, weak differentiability is equivalent to Fréchet (strong) differentiability. A key result in this chapter, which has been first established in [28], will show that the product of two weakly differentiable measures is again weakly differentiable. This leads one to conclude that the product measure Π_θ defined by (0.2) is weakly differentiable, provided that the input distributions $\mu_{i,\theta}$, for $1 \le i \le n$, are weakly differentiable. In addition, a result which shows that weak differentiability implies strong Lipschitz continuity, where "strong" means with respect to the operator v-norm defined by (0.7), will be provided. This will be the starting point for establishing strong bounds on perturbations in Chapter 3. Eventually, we investigate under which conditions weak differentiability of measures implies set-wise differentiability and we illustrate our theory with some elaborate gradient estimation examples. For instance, a ruin problem arising in insurance will be treated in Section 2.5.1 and the weak differentiability of the distribution of the transient waiting time will be analyzed in Section 2.5.2.

Chapter 3 deals with strong bounds on perturbations. That is, we establish bounds for expressions such as

$$\Delta^g(\theta_1, \theta_2) := |P_g(\theta_2) - P_g(\theta_1)| \tag{0.8}$$

where, for $\theta \in \Theta$, $P_g(\theta)$ is defined by (0.1). We establish bounds on the perturbations Δ^g in (0.8) by showing that the function $P_g(\theta)$ is Lipschitz continuous in θ and we extend our results to general Markov chains. A first attempt on this issue was made in [22] and further developed in [26]. The results presented in Chapter 3 basically rely on the theory developed in Chapter 2. Eventually, we illustrate the results by an application to both transient and steady-state waiting times in the G/G/1 queue. An important result, which shows that weak differentiability of the service-time distribution in a G/G/1 queue implies strong Lipschitz continuity of the stationary distribution of the queue, will indicate that weak differentiation techniques can be successfully applied when studying strong stability of Markov chains.

In Chapter 4 we extend the concept of weak differentiation to higher order derivatives and weak analyticity. It will turn out that differentiation of products of measures is rather similar to that of functions in classical analysis, i.e., a “Leibnitz-Newton” rule holds true. Moreover, we show that, just like in conventional analysis, the product of two weakly analytical measures is again weakly analytical. Eventually, we perform Taylor series approximations for parameter-dependent stochastic systems. These results were also established in [28].

Finally, in Chapter 5 we apply our measure-valued differential calculus developed in Chapter 4 to distributions of random matrices in some non-conventional algebras of matrices (e.g., max-plus and min-plus algebra). An elaborate example was treated in [23]. It will turn out that, by choosing the set $\mathcal{D}$ to be a class of polynomially bounded cost-functions, a formal calculus of weak differentiation can be introduced for random matrices as well. This appears to be useful in applications as it provides handy tools for computing algorithmically higher-order derivatives and, consequently, constructing Taylor series.

1. MEASURE THEORY AND FUNCTIONAL ANALYSIS

This preliminary chapter deals with basic concepts and results from both *measure theory* and *functional analysis* as much of the theory put forward in this thesis relies on standard results from these two highly inter-connected fields of mathematics.

1.1 Introduction

The connection between measure theory and functional analysis is very well known. Concepts like *duality* and *norm spaces* find a perfect justification in terms of measures. More specifically, measures can be viewed as elements in some particular linear space. It is well known that *Radon measures* appear as linear functionals on the space of continuous functions on some locally compact topological space. For a recent reference see, e.g., [10]. Therefore, one can derive interesting results by establishing structural properties for the space of measures using tools from functional analysis and then translating them in terms of measures. This is particulary useful when dealing with convergence issues on spaces of measures.

Throughout this chapter, particular attention will be paid to signed measures. This deviates from standard literature where convergence results are formulated for probability measures, only. While many properties of signed measures can be easily derived from similar properties of positive measures via the well known *Hahn-Jordan decomposition*, this is not straightforward when dealing with convergence issues, as will be illustrated in Section 1.2.4. This will lead us to introduce the concept of *regular convergence*.

Most likely, the reason why not many authors dealt with convergence of signed measures is its lack of applications. So why investing in such a topic? The answer is partly given in Chapter 2, where the concept of weak differentiation is introduced. As it will turn out, the weak derivative is a signed measure and for studying weak derivatives it will prove fruitful to extend standard results regarding weak convergence of probability measures to signed measures. However, to be able to use tools from functional analysis, like the *Banach-Steinhaus* and *Banach-Alaoglu* theorems, an appropriate mathematical setting is needed and this leads to the concept of *Banach base* introduced in Section 1.3.2.

Weak convergence of measures is one of the key topics of this chapter. It was originally introduced by P. Billingsley in [8] for probability measures in terms of bounded and continuous functions (test functions). Here we aim to extend the concept in the following directions: (1) by considering signed measures and (2) by considering a larger class of test functions. The main reason is that weak convergence as introduced in [8] is unable to handle unbounded performance measures, e.g., the mean and the deviation, which drastically reduces its area of applicability. The analysis of weak convergence of signed measures as put forward in this chapter is new. The theoretical work is a technical preliminary for our later results on weak differentiability.

The chapter is organized as follows. A brief introduction to topology and measure theory is provided in Section 1.2, where basic definitions and notations are presented. Section 1.3 deals with norm spaces of both functions and measures. In particular, the concept of *Banach base*, which will serve as a basis for developing our theory, will be introduced.

1.2 Elements of Topology and Measure Theory

This section is devoted to recall basic concepts related to topology and measure theory. In Section 1.2.1 metric spaces, which will be the basis for developing our theory, are discussed. Then, in Section 1.2.2 we discuss the concept of measure and particular attention will be paid to signed measures. Eventually, in Section 1.2.3 a special class of functional spaces is introduced to be used in Section 1.2.4 for defining weak convergence of measures.

1.2.1 Topological and Metric Spaces

Let $\mathbb{S}$ be a non-empty set. A family $\mathfrak{T}$ of subsets of $\mathbb{S}$ is called a *topology* on $\mathbb{S}$ if it satisfies the following requirements

- $\mathbb{S}$ and $\emptyset$ belong to $\mathfrak{T}$.
- Any union of elements from $\mathfrak{T}$ belongs to $\mathfrak{T}$.
- Any finite intersection of elements from $\mathfrak{T}$ belongs to $\mathfrak{T}$.

A sub-family $\mathfrak{B} \subset \mathfrak{T}$ is called a *base* for the topology $\mathfrak{T}$ if any set $A \in \mathfrak{T}$ can be expressed as a union of elements from $\mathfrak{B}$. Bases are useful because many properties of topologies can be reduced to statements about a base generating that topology and because many topologies are most easily defined in terms of a base which generates them.

If $\mathfrak{T}, \mathfrak{T}'$ are topologies on $\mathbb{S}$ we say that $\mathfrak{T}$ is *coarser* than $\mathfrak{T}'$ if $\mathfrak{T} \subset \mathfrak{T}'$. It can be easily seen that any arbitrary intersection of topologies on $\mathbb{S}$ is again a topology on $\mathbb{S}$. Therefore, for an arbitrary family $\mathfrak{A}$ of subsets of $\mathbb{S}$ one can define the *topology generated by* $\mathfrak{A}$ by taking the intersection of all topologies on $\mathbb{S}$ which contain $\mathfrak{A}$, i.e., the coarsest topology which contains $\mathfrak{A}$. Consequently, it can be shown that $\mathfrak{B}$ is a base for the topology $\mathfrak{T}$ if and only if

(i) there exist an arbitrary family $\{A_i : i \in I\} \subset \mathfrak{B}$ such that

$$\mathbb{S} = \bigcup_{i \in I} A_i,$$

(ii) for any $A_1, A_2 \in \mathfrak{B}$ and $s \in A_1 \cap A_2$ there exist $A_3 \in \mathfrak{B}$ such that

$$s \in A_3 \subset A_1 \cap A_2,$$

(iii) $\mathfrak{T}$ is the topology generated by the family $\mathfrak{B}$.

If $\mathfrak{T}$ is a topology on $\mathbb{S}$ then the pair $(\mathbb{S}, \mathfrak{T})$ is called a *topological space*. The elements of $\mathfrak{T}$ are called *open sets* and the *closed sets* are defined as the complements of the open sets. It follows that any union and any finite intersection of open sets is still an open set and any topology is determined by the open sets.

Let $(\mathbb{S}_1, \mathfrak{T}_1)$ and $(\mathbb{S}_2, \mathfrak{T}_2)$ be topological spaces. A mapping $f : \mathbb{S}_1 \to \mathbb{S}_2$ is said to be *continuous* if

$$\forall A \in \mathfrak{T}_2 : \ f^{-1}(A) \in \mathfrak{T}_1,$$

where $f^{-1}(A)$ denotes the pre-image of the set A through f, i.e.,

$$f^{-1}(A) = \{s \in \mathbb{S}_1 : \ f(s) \in A\}.$$

Note that the continuity property of f depends on the topologies $\mathfrak{T}_1$ and $\mathfrak{T}_2$. Moreover, f remains continuous if one enlarges $\mathfrak{T}_1$ but one can not draw the same conclusion if $\mathfrak{T}_1$ becomes coarser. Hence, we conclude that, for fixed $\mathfrak{T}_2$, there is a minimal (coarsest) topology which makes f continuous. This is generated by the family

$$\left\{f^{-1}(A) : \ A \in \mathfrak{T}_2\right\}$$

and it is called the topology generated by f. In the same way, one can define the topology generated by an arbitrary family of functions $\{f_i : i \in I\}$.

While many other concepts such as compactness, separability and completeness can be introduced at this abstract level we prefer to concentrate our attention on the special class of metric spaces to be introduced presently.

A mapping $d : \mathbb{S} \times \mathbb{S} \to [0, \infty)$ is said to be a *distance* (or *metric*) on $\mathbb{S}$ if

- $d(s,t) = 0$ if and only if $s = t$,
- it is symmetric, i.e.,
$$\forall s, t \in \mathbb{S} : \ d(s,t) = d(t,s),$$
- it satisfies the *triangle inequality*, i.e.,
$$\forall r, s, t \in \mathbb{S} : \ d(r,t) \leq d(r,s) + d(s,t).$$

If d is a metric on $\mathbb{S}$, then the pair $(\mathbb{S}, d)$ will be called a *metric space*.

In what follows, we assume that $(\mathbb{S}, d)$ is a metric space and we let

$$\forall s \in \mathbb{S}, \epsilon > 0 : \ B_\epsilon(s) := \{x \in \mathbb{S} : \ d(x,s) < \epsilon\}$$

denote the *open ball centered in s of radius ϵ*. $\mathbb{S}$ is endowed with the standard topology given by metric d, i.e., the topology generated by the base

$$\mathfrak{B} = \left\{B_\epsilon(s) : \ s \in \mathbb{S}, \epsilon > 0\right\}.$$

It turns out that the set $A \subset \mathbb{S}$ is open if for all $s \in A$ there exists $\epsilon > 0$ such that $B_\epsilon(s) \subset A$. The *closure* of a set A, denoted by $\bar{A}$, is defined as the smallest closed set which includes A. For instance, it can be shown that the closure of $B_\epsilon(s)$, denoted shortly by $\bar{B}_\epsilon(s)$, is given by

$$\bar{B}_\epsilon(s) = \{x \in \mathbb{S} : \ d(x,s) \le \epsilon\}.$$

An element $x \in \mathbb{S}$ is said to be an *adherent point* for the set $A \subset \mathbb{S}$ if $x \in \bar{A}$ and we call x an *accumulation point* for A if $x \in \bar{A} \setminus A$. If $A \subset B \subset \mathbb{S}$, we say that A is a *dense* subset of B if $\bar{A} = B$, i.e., B consists at all adherent points of A. $\mathbb{S}$ is said to be *separable* if there exists a dense countable subset $\{s_i : i \in I\} \subset \mathbb{S}$. It is known, for instance, that Euclidean spaces $\mathbb{R}^n$ are separable.

The set $A \subset \mathbb{S}$ is said to be bounded if

$$\sup_{s,t \in A} d(s,t) < \infty$$

and we call A *compact* if for each family $\{A_i : \ i \in I\}$ satisfying

$$A \subset \bigcup_{i \in I} A_i$$

there exist a finite set of indices $\{i_1, \ldots, i_n\} \subset I$, for some $n \ge 1$, such that

$$A \subset \bigcup_{i=1}^{n} A_i.$$

It turns out that every compact set is closed and bounded but the converse is, in general, not true[1].

The metric space $\mathbb{S}$ is said to be *locally compact* if for all $s \in \mathbb{S}$, there exists some $\epsilon > 0$ such that $\bar{B}_\epsilon(s)$ is a compact set. $\mathbb{S}$ is said to be *complete* if each Cauchy sequence $\{s_n\}_n \subset \mathbb{S}$ is convergent to some limit $s \in \mathbb{S}$. Note that compactness implies completeness while the converse is not true. For instance, $\mathbb{R}$ is complete but it fails to be compact. It is however locally compact. For more details on general topology we refer to [36].

On the metric space $\mathbb{S}$ we denote by $\mathcal{C}(\mathbb{S})$ the space of continuous, real-valued functions and by $\mathcal{C}_B(\mathbb{S})$ the subspace of continuous and bounded functions. The set $\mathcal{C}_B(\mathbb{S})$ becomes itself a metric space when endowed with the distance

$$\forall f, g \in \mathcal{C}_B(\mathbb{S}) : \ D(f,g) = \sup_{s \in \mathbb{S}} d(f(s), g(s)). \tag{1.1}$$

Since every continuous function maps compacts into compacts (in particular bounded sets) $\mathcal{C}_B(\mathbb{S}) = \mathcal{C}(\mathbb{S})$ provided that $\mathbb{S}$ is compact. Moreover, if $\mathbb{S}$ is complete then $\mathcal{C}_B(\mathbb{S})$ enjoys the same property. For later reference we denote by $\mathcal{C}^+(\mathbb{S})$ the cone of non-negative, continuous mappings on $\mathbb{S}$, i.e.,

$$\mathcal{C}^+(\mathbb{S}) = \{f \in \mathcal{C}(\mathbb{S}) : \ f(s) \ge 0, \forall s \in \mathbb{S}\}.$$

[1] For euclidian spaces, *compact* is actually equivalent to *closed and bounded*.

1.2.2 The Concept of Measure

We call a *σ-field* on $\mathbb{S}$ a family $\mathcal{S}$ of subsets of $\mathbb{S}$ satisfying

- $\emptyset \in \mathcal{S}$,
- if $A_n \in \mathcal{S}$, for each $n \in \mathbb{N}$, then
$$\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{S},$$
- for each $A \in \mathcal{S}$ it holds that $\complement A \in \mathcal{S}$,

where $\complement A$ denotes the complement of A, i.e., $\complement A = \mathbb{S} \setminus A$. Similar to topologies, the intersection of an arbitrary family of σ-fields is a σ-field and consequently we define the σ-field generated by a family $\mathfrak{A}$ as the intersection of all σ-fields containing $\mathfrak{A}$. On the metric space $\mathbb{S}$ we denote by $\mathcal{S}$ its *Borel field*, i.e., the smallest σ-field which contains the open sets. If $\mathcal{R}$ denotes the Borel field of $\mathbb{R}$, then we say that the mapping $f : \mathbb{S} \to \mathbb{R}$ is *measurable* if
$$\forall C \in \mathcal{R} : \; \{s \in \mathbb{S} : \; f(s) \in C\} \in \mathcal{S}.$$
Let $\mathcal{F}(\mathbb{S})$ denote the space of measurable functions on $\mathbb{S}$ and $\mathcal{F}_B(\mathbb{S}) \subset \mathcal{F}(\mathbb{S})$ denote the subspace of bounded mappings. Since continuity implies measurability it holds that
$$\mathcal{C}(\mathbb{S}) \subset \mathcal{F}(\mathbb{S}).$$
σ-fields are basic structures on which we define *measures*. A mapping $\mu : \mathcal{S} \to \mathbb{R} \cup \{\pm\infty\}$ is called a *signed measure* if $\mu(\emptyset) = 0$ and for each family $\{A_n\}_n \subset \mathcal{F}$ of mutually disjoint sets it holds that[2]
$$\mu\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \sum_{n\in\mathbb{N}} \mu(A_n).$$
If $\mu(A) \geq 0$, for each $A \in \mathcal{S}$, we call μ a *positive measure*, or simply a *measure*, when no confusion occurs. In standard terminology, a signed measure is a measure which is allowed to attain negative values.

Positive Measures

The positive measure μ is said to be *finite* if $\mu(A) < \infty$, for each $A \in \mathcal{S}$, i.e., $\mu(\mathbb{S}) < \infty$. A (positive) measure μ is said to be *locally finite* if for all $s \in \mathbb{S}$ there exists $\epsilon > 0$ such that $\mu(B_\epsilon(s)) < \infty$. We call μ a *Radon measure* if it is locally finite and *regular*, i.e.,

- μ is *outer regular*, i.e., each set $A \in \mathcal{S}$ satisfies
$$\mu(A) = \inf\{\mu(U) : \; A \subset U, \, U \text{ is open}\},$$

[2] The property is often referred to as σ-additivity. To avoid unnecessary complications we exclude the case when μ takes both $\pm\infty$ as values.

- μ is *inner regular*, i.e., each open subset $U \subset \mathbb{S}$ satisfies

$$\mu(U) = \sup\{\mu(K) : \ K \subset U, \ K \text{ is compact}\}.$$

We say that a family $\mathcal{P}$ of measures is *tight* if each $\mu \in \mathcal{P}$ is finite and for each $\epsilon > 0$ there exists a compact subset K of $\mathbb{S}$ such that

$$\forall \mu \in \mathcal{P} : \ \mu(\mathbb{S} \setminus K) < \epsilon.$$

Note that, if $\mathcal{P} = \{\mu\}$, i.e., $\mathcal{P}$ consists of a single element, then tightness is equivalent to inner regularity of μ, provided that μ is finite.

For a measure μ and $p \geq 1$ we denote by $\mathcal{L}^p(\mu)$ the family of measurable functions whose p^{th} power is Lebesgue integrable with respect to μ, i.e.,

$$\mathcal{L}^p(\mu) = \left\{ g \in \mathcal{F}(\mathbb{S}) : \ \int |g(s)|^p \mu(ds) < \infty \right\}.$$

For an arbitrary family of measures $\{\mu_i : i \in I\}$ we denote by $\mathcal{L}^p(\mu_i : \ i \in I)$ the family of measurable functions which are Lebesgue integrable with respect to μ_i, for all $i \in I$, i.e.,

$$\mathcal{L}^p(\mu_i : \ i \in I) = \bigcap_{i \in I} \mathcal{L}^p(\mu_i).$$

We say that $v \in \mathcal{F}$ is *uniformly integrable* with respect to the family $\{\mu_i : i \in I\}$ if

$$\lim_{x \uparrow \infty} \sup_{i \in I} \int |v(s)| \cdot \mathbb{I}_{\{|v| > x\}}(s) \mu_i(ds) = 0,$$

where $\mathbb{I}_{\{|v|>x\}}$ denotes the indicator function of the set $\{s \in \mathbb{S} : \ |v(s)| > x\}$. It is worth noting that uniform integrability of v with respect to the family $\{\mu_i : i \in I\}$ implies uniform integrability of v with respect to any sub-family $\{\mu_i : i \in J\}$, with $J \subset I$ and if v is uniformly integrable with respect to $\{\mu_i : i \in I\}$ it follows that $v \in \mathcal{L}^1(\mu_i : i \in I)$. However, the converse is true only when I is finite.

In general, checking uniform integrability of a function g with respect to some family $\{\mu_i : i \in I\} \subset \mathcal{M}^+$, by definition, might not be the most convenient method. In practice, a common way to prove uniform integrability is the following.

Lemma 1.1. *Let $g \in \mathcal{F}$, $\{\mu_i : i \in I\} \subset \mathcal{M}^+$. If there exists $\vartheta : [0, \infty) \to [0, \infty)$ satisfying*

$$M := \sup_{i \in I} \int \vartheta(|g(s)|) \mu_i(ds) < \infty, \qquad \lim_{x \to \infty} \frac{\vartheta(x)}{x} = \infty,$$

then g is uniformly integrable with respect to $\{\mu_i : i \in I\}$.

Proof. From the limit-relation we conclude that for arbitrarily small $\epsilon > 0$ there exists some $x_\epsilon > 0$ such that for each $x > x_\epsilon$ it holds that $\epsilon^{-1} < x^{-1}\vartheta(x)$. Hence, for each s,

$$|g(s)| > x_\epsilon \Rightarrow |g(s)| < \epsilon \, \vartheta(|g(s)|).$$

Therefore, for any $x > x_\epsilon$ it holds that

$$\forall i \in I: \int |g(s)| \cdot \mathbb{I}_{\{|g|>x\}}(s)\mu_i(ds) \le \epsilon \int \vartheta(|g(s)|)\mu_i(ds) \le \epsilon\, M.$$

Take in the above inequality the supremum with respect to $i \in I$ and the claim follows by letting $\epsilon \to 0$. □

A measure μ is said to be *absolutely continuous* with respect to another measure λ if for each $A \in \mathcal{S}$, $\lambda(A) = 0$ implies $\mu(A) = 0$. Two measures μ and κ are said to be *orthogonal* if there exists $A \in \mathcal{S}$ such that $\mu(A) = \kappa(\complement A) = 0$. If $\mathbb{S}$ is a Euclidean space and we denote by ℓ the Lebesgue measure on $\mathbb{S}$ then any measure which is absolutely continuous with respect to ℓ is referred to as *absolutely continuous*, or *continuous*, and any measure which is orthogonal with ℓ is referred to as *singular*.

Signed Measures

At a theoretical level, signed measures arise as natural extensions of measures because they can be organized as a linear space. This will be explained in Section 1.3.3. In practice, signed measures very often appear as differences between positive measures. In fact, any signed measure can be represented as the difference between two measures. This fact derives from the well known Hahn-Jordan decomposition theorem which states that any signed measure μ can be represented as

$$\forall A \in \mathcal{S}: \ \mu(A) = [\mu]^+(A) - [\mu]^-(A), \tag{1.2}$$

where $[\mu]^\pm$ are uniquely determined orthogonal measures called the *positive* (resp. *negative*) *part* of μ. The measure $|\mu|$ defined as

$$\forall A \in \mathcal{S}: \ |\mu|(A) = [\mu]^+(A) + [\mu]^-(A)$$

is called the *variation measure* of μ and the positive number

$$\|\mu\| = |\mu|(\mathbb{S}) = [\mu]^+(\mathbb{S}) + [\mu]^-(\mathbb{S}) \tag{1.3}$$

is called the the *total variation* (norm) of μ. Note, however, that a representation as in (1.2) is not unique if we drop the orthogonality condition. More specifically, it can be shown that $[\mu]^\pm$ satisfy

$$[\mu]^+(A) = \sup\{\mu(E): \ E \in \mathcal{S}, \ E \subset A\} \ge \max\{\mu(A), 0\}, \ [\mu]^-(A) = [\mu]^+(A) - \mu(A) \ge 0,$$

and any other decomposition $\mu = \mu^+ - \mu^-$ of μ satisfies $\mu^\pm = \nu + [\mu]^\pm$, for some (positive) measure ν. This means in particular that the orthogonal decomposition in (1.2) minimizes the sum $\mu^+ + \mu^-$, where the minimization has to be understood with respect to the order relation given by $\mu \ge \nu$ iff $\mu(A) \ge \nu(A)$, for all $A \in \mathcal{S}$. Therefore, it holds that

$$|\mu| = \inf\{\mu^+ + \mu^- : \ \mu^+ - \mu^- = \mu, \ \mu^\pm \text{ are positive measures}\}.$$

Throughout this thesis we will denote the orthogonal decomposition by $[\mu]^\pm$.

In what follows we assume that $\mathbb{S}$ is separable and locally compact, we denote by $\mathcal{M}(\mathcal{S})$ the space of signed Radon measures on $\mathcal{S}$ and denote by $\mathcal{M}_B(\mathcal{S})$ the subset of finite (bounded) measures. The cone of positive measures in $\mathcal{M}(\mathcal{S})$ is denoted by $\mathcal{M}^+(\mathcal{S})$ and we denote by $\mathcal{M}^1(\mathcal{S})$ the subset of probability measures, i.e.,

$$\mathcal{M}^1(\mathcal{S}) = \{\mu \in \mathcal{M}^+(\mathcal{S}) : \mu(\mathbb{S}) = 1\}.$$

Many properties of measures can be extended to signed measures by means of the variation measure. More specifically, we say that μ is locally finite, finite, regular or absolutely continuous with respect to some λ if $|\mu|$ is locally finite (resp. finite, regular or absolutely continuous with respect to λ). In all of these situations it turns out that both $[\mu]^+$ and $[\mu]^-$ enjoy the same property. Moreover, we say that a measurable function is integrable with respect to a signed measure μ if it is integrable with respect to the variation of μ or, equivalently, if it is integrable with respect to both $[\mu]^\pm$. In the same vein we say that the family $\mathcal{P}$ of signed measures is tight if the corresponding family of positive measures $\{|\mu| : \mu \in \mathcal{P}\}$ is tight, which is equivalent to the tightness of both families $\{[\mu]^\pm : \mu \in \mathcal{P}\}$. Consequently, some standard results from measure theory can be easily extended to signed measures. A list of a few standard results in measure theory can be found in Section C of the Appendix. For thorough treatment of signed measures, we refer to [14].

We conclude this section with a few remarks on *measure-valued mappings*. For a non-empty set $\Theta \subset \mathbb{R}$ let $\{\mu_\theta : \theta \in \Theta\} \subset \mathcal{M}(\mathcal{S})$ be an arbitrary family of signed measures and consider the mapping $\mu_* : \Theta \to \mathcal{M}(\mathcal{S})$ defined as

$$\forall \theta \in \Theta : \ \mu_*(\theta) = \mu_\theta,$$

i.e., $\{\mu_\theta : \theta \in \Theta\}$ is the *range* of μ_*. Provided that an appropriate topology is introduced on $\mathcal{M}(\mathcal{S})$, or some subset which includes the range of μ_*, continuity of measure-valued mappings is defined in an obvious way.

1.2.3 $\mathcal{C}_v$-spaces

Throughout this section we assume that v is a non-negative, continuous function on $\mathbb{S}$, i.e., $v \in \mathcal{C}^+(\mathbb{S})$ and we denote by $\mathbb{S}_v$ the *support* of v, i.e., the open set

$$\mathbb{S}_v := \{s \in \mathbb{S} : \ v(s) > 0\}.$$

We denote by $\mathcal{C}_v(\mathbb{S})$ the set of v-bounded, continuous functions, i.e.,

$$\mathcal{C}_v(\mathbb{S}) := \{g \in \mathcal{C}(\mathbb{S}) : \ \exists c > 0 \, s.t. \, |g(s)| \le c\, v(s), \forall s \in \mathbb{S}\}. \tag{1.4}$$

Note that if $v \in \mathcal{C}_B(\mathbb{S})$ then $\mathcal{C}_v(\mathbb{S}) = \mathcal{C}_B(\mathbb{S})$ and, in general, $\mathcal{C}_B(\mathbb{S}) \subset \mathcal{C}_v(\mathbb{S})$ provided that $\inf\{v(s) : s \in \mathbb{S}\} > 0$. In addition, if $g \in \mathcal{C}_v$ then $g(s) = 0$ for any $s \in \mathbb{S} \setminus \mathbb{S}_v$. A typical choice for $\mathcal{C}_v(\mathbb{S})$ is provided in the following example.

Example 1.1. *Let $v_\alpha(x) = e^{\alpha x}$, for some $\alpha \ge 0$, for $x \in \mathbb{S} = [0, \infty)$. Since for every polynomial p it holds that $\lim_{x\to\infty} e^{-\alpha x}|p(x)| = 0$ it turns out that the space $\mathcal{C}_{v_\alpha}([0,\infty))$ contains all (finite) polynomials. However, $\mathcal{C}_{v_\alpha}$ is not restricted to polynomials. Indeed, note that the mapping $x \mapsto \ln(1+x)$ also belongs to $\mathcal{C}_{v_\alpha}$. Moreover, for $\alpha < \beta$ we have $\mathcal{C}_{v_\alpha} \subset \mathcal{C}_{v_\beta}$ since $\alpha < \beta$ implies $\|g\|_{v_\beta} \le \|g\|_{v_\alpha}$, for any g.*

Remark 1.1. *A set $\mathcal{D}$ of measurable mappings is said to separate the points of a family $\mathcal{P} \subset \mathcal{M}(\mathcal{S})$ if for each $\mu^1, \mu^2 \in \mathcal{P}$, $\mu^1 \neq \mu^2$ there exists some $g \in \mathcal{D}$ such that*

$$\int g(s)\mu^1(ds) \neq \int g(s)\mu^2(ds).$$

This can be re-phrased by saying that "the family of integrals with integrands $g \in \mathcal{D}$ uniquely determines the measure in $\mathcal{P}$". It is known that $\mathcal{C}_B(\mathbb{S})$ enjoys this property while, in general, such a property fails to hold true when $\mathcal{D} = \mathcal{C}_v(\mathbb{S})$. Indeed, let us denote by v the identity mapping on $\mathbb{S} = [0, \infty)$, i.e., $v(s) = s$, for each $s \geq 0$. Then, for all $g \in \mathcal{C}_v(\mathbb{S})$ it holds that $|g(0)| \leq c\, v(0) = 0$ and if for $\alpha > 0$ we let

$$\forall A \in \mathcal{S} : \ \mu_\alpha(A) = \alpha \cdot \mathbb{I}_A(0),$$

i.e., the measure which assigns mass α to 0, we note that $\mathcal{C}_v$ does not separate the points of the family $\mathcal{P} := \{\mu_\alpha : \alpha > 0\}$. Indeed, for $\alpha \neq \beta$, it holds that

$$\forall g \in \mathcal{C}_v(\mathbb{S}) : \int g(s)\mu_\alpha(ds) = \int g(s)\mu_\beta(ds) = 0,$$

which stems from the fact that all measures in $\mathcal{P}$ assign mass exclusively to point $0 \notin \mathbb{S}_v$.

As detailed in Remark 1.1, $\mathcal{C}_v$-spaces fail to separate the points of $\mathcal{M}(\mathcal{S})$. However, in applications one is typically interested in evaluating the integrals $\int g d\mu$, for $g \in \mathcal{C}_v(\mathbb{S})$, rather than investigating the measure μ, itself. That is, we study the trace of a measure μ on $\mathbb{S}_v$, since any $g \in \mathcal{C}_v$ vanishes on $\mathbb{S} \setminus \mathbb{S}_v$. The following result will show that $\mathcal{C}_v$-spaces separate equivalence classes.

Lemma 1.2. *Let $\mu^1, \mu^2 \in \mathcal{M}(\mathcal{S})$ and let $v \in \mathcal{C}^+(\mathbb{S}) \cap \mathcal{L}^1(\mu^1, \mu^2)$ be such that*

$$\forall g \in \mathcal{C}_v : \int g(s)\mu^1(ds) = \int g(s)\mu^2(ds). \tag{1.5}$$

Then the traces of μ^1 and μ^2 on $\mathbb{S}_v$ coincide, i.e.,

$$\forall A \in \mathcal{S} : \ \mu^1(A \cap \mathbb{S}_v) = \mu^2(A \cap \mathbb{S}_v), \tag{1.6}$$

provided that $\min\{\mu^1(\mathbb{S}_v), \mu^2(\mathbb{S}_v)\} < \infty$.

Proof. Since $\mathcal{S}$ is the Borel σ-field of $\mathbb{S}$, we may assume without loss of generality that $A \in \mathcal{S}$ is an arbitrary non-empty, open set. For $n \geq 1$ consider the set

$$A_n := \{s \in A : \ d(s, \complement A) \geq 1/n\} \subset A,$$

where, for $E \subset \mathbb{S}$ we denote $d(s, E) = \inf\{d(s, x) : x \in E\}$. Note that, for sufficiently large n, A_n is a non-empty, closed set satisfying $A_n \cap \complement A = \emptyset$. Since A is an open set, i.e., $\complement A$ is closed, according to Urysohn's Lemma there exists a continuous function $f_n : \mathbb{S} \to [0, 1]$ such that $f_n(s) = 1$ for $s \in A_n$ and $f_n(s) = 0$, for $s \in \complement A$. On the other hand the family $\{A_n : n \geq 1\} \subset \mathcal{S}$ is increasing and $\cup_{n \geq 1} A_n = A$. Hence, f_n converges point-wise to $\mathbb{I}_A$ as $n \to \infty$.

Consider now for each $n \geq 1$ the mapping $g_n \in \mathcal{C}^+(\mathbb{S})$ defined by

$$g_n(s) = \min\{f_n(s), n \cdot v(s)\}.$$

Note that $g_n \in \mathcal{C}_v(\mathbb{S})$, for each $n \geq 1$, and by hypothesis it follows that

$$\forall n \geq 1 : \int g_n(s)\mu^1(ds) = \int g_n(s)\mu^2(ds). \tag{1.7}$$

Moreover, we have $g_n \leq \mathbb{I}_{A\cap\mathbb{S}_v}$ and $\lim_n g_n = \mathbb{I}_{A\cap\mathbb{S}_v}$, point-wise. Therefore, provided that $\min\{\mu^1(\mathbb{S}_v), \mu^2(\mathbb{S}_v)\} < \infty$, by letting $n \to \infty$ in (1.7) it follows from the Dominated Convergence Theorem that

$$\mu^1(A \cap \mathbb{S}_v) = \mu^2(A \cap \mathbb{S}_v),$$

which concludes the proof of (1.6). □

Remark 1.2. *If we denote by $\sim$ the equivalence relation on $\mathcal{M}(\mathcal{S})$ given by $\mu^1 \sim \mu^2$ if (1.6) holds true then Lemma 1.2 shows that if (1.5) holds true then $\mu^1 \sim \mu^2$. That is, $\mathcal{C}_v(\mathbb{S})$ separates the points of the quotient space $\mathcal{M}(\mathcal{S})/\sim$.*

For ease of notation, in the following we will omit specifying the space $\mathbb{S}$ or the σ-field $\mathcal{S}$, when no confusion occurs.

1.2.4 Convergence of Measures

Throughout this section we discuss the concept of weak convergence of measures. Formally, we say that a sequence of measures $\{\mu_n\}_n$ is weakly $\mathcal{D}$-convergent to some limit μ if the integrals of μ_n converge to those of μ for some predefined class of cost-functions $\mathcal{D}$. Weak convergence of measures was originally introduced in [8] in terms of continuous and bounded functions, i.e., $\mathcal{D} = \mathcal{C}_B$. The main reason for this is that $\mathcal{C}_B(\mathbb{S})$ separates the points of $\mathcal{M}(\mathcal{S})$ and, as a consequence, the weak limit is uniquely determined, provided that it exists.

A first step in extending this concept is by taking $\mathcal{D} = \mathcal{C}_v$ since, according to Lemma 1.2, $\mathcal{C}_v$-spaces posses satisfactory separation properties which make them suitable for defining weak convergence. Concurrently, the main result of this section will establish how general $\mathcal{C}_v$-convergence is related to classical $\mathcal{C}_B$-convergence. The following definition introduces the concept of *weak convergence* on $\mathcal{M}$.

Definition 1.1. *Let $\{\mu_n : n \in \mathbb{N}\} \subset \mathcal{M}$ and $\mathcal{D} \subset \mathcal{L}^1(\mu_n : n \in \mathbb{N})$. The sequence $\{\mu_n\}_n$ is weakly $\mathcal{D}$-convergent, if there exists $\mu \in \mathcal{M}$ such that*

$$\forall g \in \mathcal{D} : \lim_{n\to\infty} \int g(s)\mu_n(ds) = \int g(s)\mu(ds). \tag{1.8}$$

We write $\mu_n \overset{\mathcal{D}}{\Longrightarrow} \mu$ (or simply $\mu_n \Rightarrow \mu$ when no confusion occurs) and we call μ a weak $\mathcal{D}$-limit of the sequence $\{\mu_n\}_n$.

Note that a weak $\mathcal{D}$-limit is determined by the class of integrals with integrands $g \in \mathcal{D}$ and is not unique if $\mathcal{D}$ does not separate the points of $\mathcal{M}$; see Remark 1.2. However, $\mathcal{C}_v \subset \mathcal{L}^1(\mu_n : n \in \mathbb{N})$ is equivalent to $v \in \mathcal{L}^1(\mu_n : n \in \mathbb{N})$ and by letting $\mathcal{D} = \mathcal{C}_v$ the weak limit μ in (1.8) is unique in the sense specified by Lemma 1.2. Therefore, one obtains a sensible definition for $\mathcal{C}_v$-convergence by letting $\mathcal{D} = \mathcal{C}_v$, for some $v \in \mathcal{C}^+ \cap \mathcal{L}^1(\mu_n : n \in \mathbb{N})$, in Definition 1.1. The following example illustrates the dependence of $\mathcal{C}_v$-convergence of a sequence of measures $\{\mu_n\}_n$ on the choice of v.

Example 1.2. *On $\mathbb{S} = \mathbb{R}$ let us consider the family of probability densities*

$$\forall \theta \in (0,2), x \in \mathbb{R} : \ f(\theta, x) = \frac{\sin\left(\frac{\pi\theta}{2}\right)}{\pi} \cdot \frac{|x|^{\theta-1}}{1+x^2}.$$

If we consider the sequence of probability measures $\{\mu_n : n \geq 1\}$, given by

$$\forall n \geq 1, x \in \mathbb{R} : \ \mu_n(dx) = f\left(\frac{n-1}{n}, x\right) dx,$$

then $\mu_n \stackrel{\mathcal{C}_B}{\Longrightarrow} \mu$, where μ denotes the Cauchy distribution, i.e.,

$$\mu(dx) = f(1,x)dx.$$

Nevertheless, the sequence $\{\mu_n\}_n$ fails to be $\mathcal{C}_v$-convergent, when $v(x) = |x|$, although $v \in \mathcal{L}^1(\mu_n : n \geq 1)$. Indeed, we have

$$\forall n \geq 1 : \ \int |x| \, \mu_n(dx) < \infty \text{ but } \int |x| \, \mu(dx) = \infty.$$

Now the following question comes naturally: "When does $\mathcal{C}_B$-convergence of measures imply $\mathcal{C}_v$-convergence?" More specifically, which $g \in \mathcal{F}$ satisfy

$$\lim_{n\to\infty} \int g(s)\mu_n(ds) = \int g(s)\mu(ds), \tag{1.9}$$

provided that $\mu_n \stackrel{\mathcal{C}_B}{\Longrightarrow} \mu$? In the following we aim to answer to this question and investigate how general $\mathcal{C}_v$-convergence relates to classical convergence. A first step into that direction is the following result which has been proved in [8]; see Theorem F.2 in the Appendix.

Lemma 1.3. *Let $\{\mu_n : n \in \mathbb{N}\} \subset \mathcal{M}^+$ be such that $\mu_n \stackrel{\mathcal{C}_B}{\Longrightarrow} \mu$. The mapping $g \in \mathcal{C}^+$ satisfies equation (1.9) if and only if g is uniformly integrable with respect to the family $\{\mu_n : n \in \mathbb{N}\}$.*

Note that, in Example 1.2, $v(x) = |x|$ is not uniformly integrable with respect to the family $\{\mu_n : n \geq 1\}$ although $v \in \mathcal{L}^1(\mu_n : n \geq 1)$. The following result will establish a relationship between $\mathcal{C}_v$-convergence and classical weak convergence of positive measures.

Theorem 1.1. *Let $v \in \mathcal{C}^+$ and let $\{\mu_n : n \in \mathbb{N}\} \subset \mathcal{M}^+$ be a sequence of measures.*

(i) If $\mu_n \stackrel{\mathcal{C}_B}{\Longrightarrow} \mu$, i.e., μ is the classical weak limit of the sequence $\{\mu_n\}_n$ and v is uniformly integrable with respect to $\{\mu_n : n \in \mathbb{N}\}$ then $\mu_n \stackrel{\mathcal{C}_v}{\Longrightarrow} \mu$.

(ii) If $\mu_n \overset{\mathcal{C}_v}{\Longrightarrow} \mu$, $\mu_n(\mathbb{S} \setminus \mathbb{S}_v) = 0$, *for each* $n \in \mathbb{N}$, *and the family* $\{\mu_n : n \in \mathbb{N}\}$ *is tight*[3] *then* $\mu_n \overset{\mathcal{C}_B}{\Longrightarrow} \mu$.

Proof. (i) We have to show that the limit relation in (1.9) holds true for each $g \in \mathcal{C}_v$ and we can assume without loss of generality that

$$\forall x \in \mathbb{S} : \ 0 \leq g(x) \leq v(x). \tag{1.10}$$

Therefore, in accordance with Lemma 1.3 it suffices to show that each g satisfying (1.10) is uniformly integrable with respect to $\{\mu_n : n \in \mathbb{N}\}$, provided that v is. Now, this follows immediately from the inequality

$$\forall x \in \mathbb{S} : \ g(x) \cdot \mathbb{I}_{\{g>\alpha\}}(x) \leq v(x) \cdot \mathbb{I}_{\{v>\alpha\}}(x).$$

(ii) We have to show that (1.9) holds true for each $g \in \mathcal{C}_B$. We can assume without loss of generality that $0 \leq g(s) \leq 1$, for each $s \in \mathbb{S}$. For $m \geq 1$, let us define

$$\forall s \in \mathbb{S} : \ g_m(s) := \min\{g(s), m \cdot v(s)\}$$

and let us show that the double-indexed sequence $\{a_{m,n}\}_{m,n}$, defined as

$$\forall m \geq 1, n \in \mathbb{N} : \ a_{m,n} := \int g_m(s)\mu_n(ds)$$

satisfies the conditions of Theorem B.1 (see the Appendix).

First, note that, for $m \geq 1$, $g_m \in \mathcal{C}_v$ and, by hypothesis,

$$\forall m \geq 1 : \ \lim_{n\to\infty} a_{m,n} = b_m := \int g_m(s)\mu(ds).$$

On the other hand, since $\mu_n(\mathbb{S} \setminus \mathbb{S}_v) = 0$, for each $n \in \mathbb{N}$, by the Monotone Convergence Theorem (see Theorem C.2 in the Appendix) we conclude that

$$\forall n \in \mathbb{N} : \ \lim_{m\to\infty} a_{m,n} = c_n := \int g(s)\mu_n(ds).$$

Furthermore, the family $\{\mu_n : n \in \mathbb{N}\}$ being tight it follows that there exists some compact $K_\epsilon \subset \mathbb{S}_v$ such that $\mu_n(\mathbb{S}_v \setminus K_\epsilon) < \epsilon$, for each $n \in \mathbb{N}$, and $\mu(\mathbb{S}_v \setminus K_\epsilon) < \epsilon$. Furthermore, the function g/v being continuous, hence bounded on K_ϵ, it follows that

$$M := \sup_{s\in K_\epsilon} \frac{g(s)}{v(s)} < \infty.$$

Choosing now $m_\epsilon \geq M$, it follows that for $n \in \mathbb{N}$ and $m \geq m_\epsilon$ we have

$$|a_{m,n} - c_n| \leq \mu_n(\{s : g(s) > m \cdot v(s)\}) \leq \mu_n(\mathbb{S}_v \setminus K_\epsilon) \leq \epsilon, \tag{1.11}$$

since $\mu_n(\mathbb{S} \setminus \mathbb{S}_v) = 0$, for each $n \in \mathbb{N}$, and for $s \in K_\epsilon$ we have $g(s) \leq M \cdot v(s)$.

[3] Note that, if $\inf_s v(s) > 0$ then tightness of the family $\{\mu_n : n \in \mathbb{N}\}$ is a consequence of $\mu_n \overset{\mathcal{C}_v}{\Longrightarrow} \mu$.

Therefore, the sequence $\{a_{m,n}\}_{m,n}$ satisfies the conditions of Theorem C.1 and interchanging limits is justified, i.e.,

$$\lim_{n\to\infty} \int g(s)\mu_n(ds) = \lim_{n\to\infty} \lim_{m\to\infty} a_{m,n} = \lim_{m\to\infty} \lim_{n\to\infty} a_{m,n} = \int g(s)\mu(ds),$$

which concludes the proof. □

Theorem 1.1 provides the means for assessing $\mathcal{C}_v$-convergence when classical weak convergence of measures holds true and vice-versa. For instance, applying Theorem 1.1 to $\mathcal{C}_v$ defined in Example 1.1, we conclude that if the sequence $\{\mu_n\}_n$ converges $\mathcal{C}_B$-weakly to μ and (1.9) holds true for $v(s) = e^{\alpha s}$, for some $\alpha \geq 0$, then the moments of μ_n converge to those of μ.

We conclude this section by discussing the concept of regular convergence. Let the sequence $\{\mu_n\}_n$ be $\mathcal{C}_v$-convergent to some limit μ. We say that $\{\mu_n\}_n$ is *regularly $\mathcal{C}_v$-convergent* if

$$[\mu_n]^+ \stackrel{\mathcal{C}_v}{\Longrightarrow} [\mu]^+ \quad \text{and} \quad [\mu_n]^- \stackrel{\mathcal{C}_v}{\Longrightarrow} [\mu]^-,$$

i.e., the positive and negative parts of μ_n converge to the positive and negative parts of μ, respectively. A natural question that arises in the study of limits of signed measures is wether $\mathcal{C}_v$-convergence is equivalent to regular $\mathcal{C}_v$-convergence. Or, if the sequences $[\mu_n]^\pm$ converge at all. That would allow one to extend standard results regarding classical weak convergence of measures (e.g., Lemma 1.3 and Theorem 1.1) to general signed measures. Unfortunately, as the following example illustrates, this is not always the case.

Example 1.3. *Let $\xi_n = 1/n$, for each $n \geq 1$, and consider the sequence*

$$\mu_n = \begin{cases} \delta_{\xi_n} + \delta_{1+\xi_n} - \delta_1, & \text{for } n \text{ even;} \\ \delta_{\xi_n}, & \text{for } n \text{ odd,} \end{cases}$$

where, for $x \in \mathbb{S}$, we denote by δ_x the Dirac distribution, assigning mass at point x, i.e.,

$$\forall A \in \mathcal{S} : \delta_x(A) = \mathbb{I}_A(x).$$

Then $\mu_n \stackrel{\mathcal{C}_B}{\Longrightarrow} \delta_0$ but $[\mu_{2k+1}]^+ \stackrel{\mathcal{C}_B}{\Longrightarrow} \delta_0$ and $[\mu_{2k}]^+ \stackrel{\mathcal{C}_B}{\Longrightarrow} \delta_0 + \delta_1$.

However, it is worth noting that $\mathcal{C}_v$-convergence of the sequence $[\mu_n]^+$ is equivalent to that of $[\mu_n]^-$, provided that $\mu_n \stackrel{\mathcal{C}_v}{\Longrightarrow} \mu$. Moreover, $[\mu_n]^+ \stackrel{\mathcal{C}_v}{\Longrightarrow} [\mu]^+$ is equivalent to $[\mu_n]^- \stackrel{\mathcal{C}_v}{\Longrightarrow} [\mu]^-$. A sufficient condition for regular convergence will be given in Section 1.3.3.

1.3 Norm Linear Spaces

This section aims to illustrate the link between measure theory and functional analysis. More specifically, we show how both functions and measures can be treated as ordinary elements in some norm linear spaces. Moreover, powerful results can be derived by applying standard results from Banach spaces theory. To this end, we provide in Section 1.3.1 a brief overview of the basic concepts and tools from functional analysis which will be used throughout this thesis. In Section 1.3.2 we introduce the concept of *Banach base* and show, by means of an example, that this leads to a proper generalization of the $\mathcal{C}_v$-space introduced in Section 1.2.3. Spaces of measures are treated in Section 1.3.3 whereas Section 1.3.4 provides a method to construct Banach bases on product spaces.

1.3.1 Basic Facts from Functional Analysis

The central concept in functional analysis is the *linear (vector) space*. We say that $\mathcal{V}$ is a (real) linear space if there exist two binary operations

$$+ : \mathcal{V} \times \mathcal{V} \to \mathcal{V}, \ \cdot : \mathbb{R} \times \mathcal{V} \to \mathcal{V}$$

called *addition* and *scalar multiplication*, respectively, such that

- the addition is commutative and associative, i.e.

$$\forall \mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathcal{V} : \ \mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}, \ \mathbf{x} + (\mathbf{y} + \mathbf{z}) = (\mathbf{x} + \mathbf{y}) + \mathbf{z},$$

- there exists a *zero element* $\mathbf{0}$, i.e.,

$$\forall \mathbf{x} \in \mathcal{V} : \ \mathbf{x} + \mathbf{0} = \mathbf{x},$$

- for each $\mathbf{x} \in \mathcal{V}$ there exists an *inverse element* $-\mathbf{x} \in \mathcal{V}$, i.e.,

$$\forall \mathbf{x} \in \mathcal{V}, \exists - \mathbf{x} \in \mathcal{V} : \ \mathbf{x} + (-\mathbf{x}) = \mathbf{0},$$

- scalar multiplication is compatible with real number multiplication, i.e.,

$$\forall \alpha, \beta \in \mathbb{R}, \mathbf{x} \in \mathcal{V} : \ (\alpha\beta) \cdot \mathbf{x} = \alpha \cdot (\beta \cdot \mathbf{x}),$$

- 1 acts as an identity element for scalar multiplication, i.e.,

$$\forall \mathbf{x} \in \mathcal{V} : \ 1 \cdot \mathbf{x} = \mathbf{x},$$

- scalar multiplication distributes over both vector and real numbers addition, i.e.,

$$\forall \alpha, \beta \in \mathbb{R}, \mathbf{x}, \mathbf{y} \in \mathcal{V} : \ \alpha \cdot (\mathbf{x} + \mathbf{y}) = \alpha \cdot \mathbf{x} + \alpha \cdot \mathbf{y}, \ (\alpha + \beta) \cdot \mathbf{x} = \alpha \cdot \mathbf{x} + \beta \cdot \mathbf{x}.$$

A subset $\mathcal{W} \subset \mathcal{V}$ is called *stable*, or *linear subspace* if

$$\forall \alpha, \beta \in \mathbb{R}, \mathbf{x}, \mathbf{y} \in \mathcal{W} : \ \alpha \cdot \mathbf{x} + \beta \cdot \mathbf{y} \in \mathcal{W}.$$

We say that the mapping $\| \cdot \| : \mathcal{V} \to [0, \infty)$ is a *semi-norm* on $\mathcal{V}$ if

- $\| \cdot \|$ is sub-additive, i.e.,

$$\forall \mathbf{x}, \mathbf{y} \in \mathcal{V} : \ \|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|,$$

- $\| \cdot \|$ is positively homogenous, i.e.,

$$\forall \alpha \in \mathbb{R}, \mathbf{x} \in \mathcal{V} : \ \|\alpha \cdot \mathbf{x}\| = |\alpha| \, \|\mathbf{x}\|.$$

In particular, from the last property we conclude that $\|\mathbf{0}\| = 0$, by letting $\alpha = 0$. A family of semi-norms $\{\|\cdot\|_i : i \in I\}$ is said to be *separating* if for each $\mathbf{x} \in \mathcal{V}$, $\mathbf{x} \neq \mathbf{0}$, there exists some $i \in I$ such that $\|\mathbf{x}\|_i > 0$. A separating family of semi-norms induces a topology on $\mathcal{V}$ if we consider as a base the class of finite intersections from the family

$$\mathfrak{B}_0 = \{V_i(\mathbf{x}, \epsilon) : \mathbf{x} \in \mathcal{V}, \epsilon > 0, i \in I\},$$

where, for each $\mathbf{x} \in \mathcal{V}$, $\epsilon > 0$ and $i \in I$ we set

$$V_i(\mathbf{x}, \epsilon) := \{\mathbf{y} : \|\mathbf{y} - \mathbf{x}\|_i < \epsilon\}.$$

A topology generated in this way will be called a *locally convex topology* and this topology is the coarsest topology on $\mathcal{V}$ which makes the mappings $\|\cdot\|_i$ continuous, for each $i \in I$. For a full treatment of locally convex topologies we refer to [54].

If, in addition, $\|\mathbf{x}\| = 0$ implies that $\mathbf{x} = \mathbf{0}$, we say that $\|\cdot\|$ is a *norm*. A norm $\|\cdot\|$ induces a metric d on $\mathcal{V}$, as follows:

$$\forall \mathbf{x}, \mathbf{y} \in \mathcal{V} : \ d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|. \tag{1.12}$$

Therefore, any norm induces a topology on $\mathcal{V}$ by means of the metric d, given by (1.12), and the topology induced by the metric d will be called the *norm topology* on $\mathcal{V}$. Note that, if $\|\cdot\|$ is a norm on $\mathcal{V}$ then the single-element-family $\{\|\cdot\|\}$ is a separating family of semi-norms and the corresponding locally convex topology coincides with the norm topology, i.e., the norm topology is a particular case of locally convex topology.

We say that the linear norm space $(\mathcal{V}, \|\cdot\|)$ is a *Banach space* if it is complete under the norm topology. The simplest examples of Banach spaces are euclidian spaces $\mathbb{R}^k$, for $k \geq 1$, with the *uniform topology*, induced by the norm

$$\forall \mathbf{x} = (x_1, \ldots, x_k) \in \mathbb{R}^k : \ \|\mathbf{x}\| = \max\{|x_1|, \ldots, |x_k|\}.$$

A standard non-elementary Banach space is the space of bounded and continuous functions $\mathcal{C}_B(\mathbb{S})$ endowed with the *supremum norm*, i.e.,

$$\forall f \in \mathcal{C}_B(\mathbb{S}) : \ \|f\| = \sup_{s \in \mathbb{S}} |f(s)|. \tag{1.13}$$

If $(\mathcal{U}, \|\cdot\|_{\mathcal{U}})$ and $(\mathcal{V}, \|\cdot\|_{\mathcal{V}})$ are norm spaces we say that the mapping $\Phi : \mathcal{V} \to \mathcal{U}$ is a *linear operator* from $\mathcal{V}$ onto $\mathcal{U}$ if it is additive and homogeneous, i.e.,

$$\forall \alpha, \beta \in \mathbb{R}; \mathbf{x}, \mathbf{y} \in \mathcal{V} : \ \Phi(\alpha \cdot \mathbf{x} + \beta \cdot \mathbf{y}) = \alpha \cdot \Phi(\mathbf{x}) + \beta \cdot \Phi(\mathbf{y}).$$

The linear operator Φ is said to be a *bounded* if there exists $M > 0$ such that

$$\forall \mathbf{x} \in \mathcal{V} : \ \|\Phi(\mathbf{x})\|_{\mathcal{U}} \leq M \|\mathbf{x}\|_{\mathcal{V}} \tag{1.14}$$

and Φ is said to be an *isometric operator* or *isometry*, for short, if

$$\forall \mathbf{x} \in \mathcal{V} : \ \|\Phi(\mathbf{x})\|_{\mathcal{U}} = \|\mathbf{x}\|_{\mathcal{V}}. \tag{1.15}$$

It is a standard fact that a linear operator is continuous if and only if it is bounded. Moreover, any isometry is a continuous operator, since (1.14) holds true for any $M \geq 1$ and if $\mathcal{V}$ is a Banach space and Φ is a bijective isometry it follows that $\mathcal{U}$ is a Banach space, as well.

The minimal $M > 0$ for which (1.14) holds true is called the *operator norm* of Φ and is denoted by $\|\Phi\|$; in formula,

$$\|\Phi\| = \inf\left\{M > 0 : \|\Phi(\mathbf{x})\|_{\mathcal{U}} \leq M\|\mathbf{x}\|_{\mathcal{V}}, \forall \mathbf{x} \in \mathcal{V}\right\}. \tag{1.16}$$

If we denote by $\mathfrak{L}(\mathcal{V},\mathcal{U})$ the class of linear operators form $\mathcal{V}$ onto $\mathcal{U}$ then $\mathfrak{L}(\mathcal{V},\mathcal{U})$ is a linear space and $\|\cdot\|$ defined by (1.16) is a proper norm on the subspace of linear bounded operators, denoted by $\mathfrak{L}_B(\mathcal{V},\mathcal{U})$. In addition, for each $\Phi \in \mathfrak{L}_B(\mathcal{V},\mathcal{U})$ it holds that

$$\|\Phi\| = \sup\left\{\|\Phi(\mathbf{x})\|_{\mathcal{U}} : \|\mathbf{x}\|_{\mathcal{V}} \leq 1\right\} = \sup\left\{\|\Phi(\mathbf{x})\|_{\mathcal{U}} : \|\mathbf{x}\|_{\mathcal{V}} = 1\right\}.$$

If $(\mathcal{U}, \|\cdot\|_{\mathcal{U}})$ is a Banach space then $\mathfrak{L}_B(\mathcal{V},\mathcal{U})$ is a Banach space as well. Furthermore, if $\mathcal{U} = \mathbb{R}$ then $\mathfrak{L}(\mathcal{V},\mathbb{R})$ is called the *algebraic dual* of $\mathcal{V}$, its elements are called *linear functionals* and $\mathfrak{L}_B(\mathcal{V},\mathbb{R})$ is called the *topological dual* of $\mathcal{V}$, typically denoted by $\mathcal{V}^*$. Therefore, we conclude that the topological dual of a norm space is a Banach space. For more details on continuous linear operators we refer to [19].

Topological duality plays an important role in functional analysis and it provides the means for constructing new topologies on norm spaces. In some situations, the new topologies appear more natural for applications. That is why we briefly explain the concept of duality in the following. Let $\mathcal{V}$ and $\mathcal{U}$ be a pair of topological linear spaces and let $<\cdot,\cdot>: \mathcal{V} \times \mathcal{U} \to \mathbb{R}$ be a bilinear mapping such that

$$<\mathbf{x},\mathbf{y}>= 0, \forall \mathbf{x} \in \mathcal{V} \Rightarrow \mathbf{y} = \mathbf{0}, \text{ and } <\mathbf{x},\mathbf{y}>= 0, \forall \mathbf{y} \in \mathcal{U} \Rightarrow \mathbf{x} = \mathbf{0}.$$

Then one can define on $\mathcal{V}$ a minimal, locally convex topology, denoted by $\sigma(\mathcal{U},\mathcal{V})$, which makes the projection (linear) mappings

$$\{<\cdot,\mathbf{y}>: \mathbf{y} \in \mathcal{U}\}$$

continuous. This is the topology induced by the family of semi-norms

$$\{|<\cdot,\mathbf{y}>| : \mathbf{y} \in \mathcal{U}\}.$$

In addition, one can define by symmetry a minimal topology on $\mathcal{U}$, denoted by $\sigma(\mathcal{V},\mathcal{U})$, which makes the mappings

$$\{<\mathbf{x},\cdot>: \mathbf{x} \in \mathcal{V}\}$$

continuous. The topologies $\sigma(\mathcal{U},\mathcal{V})$ and $\sigma(\mathcal{V},\mathcal{U})$ are called dual topologies.

An interesting situation arises when considering the norm spaces $\mathcal{V}$ and $\mathcal{V}^*$, both endowed with the corresponding norm topology, and the continuous, bilinear mapping $<\cdot,\cdot>$ defined as

$$\forall \mathbf{x} \in \mathcal{V}, \Phi \in \mathcal{V}^* : <\mathbf{x},\Phi>= \Phi(\mathbf{x}).$$

In this case, the dual topologies are called *weak topologies*. More specifically, $\sigma(\mathcal{V}^*,\mathcal{V})$ is called the *weak topology* and $\sigma(\mathcal{V},\mathcal{V}^*)$ is called the *weak-* topology*.

Note that, in general, the weak topology is coarser than the norm topology. Consequently, continuity in norm topology implies continuity in weak topology whereas, in general, the converse is not true. This justifies the name "weak topology" and the fact that the norm topology is typically called "strong topology". For details on dual topologies we refer to [9], [19].

1.3.2 Banach Bases

In this section we provide a general method to construct spaces of measurable functions. These are norm spaces (in some cases even Banach spaces) and extend the concept $\mathcal{C}_v$-space introduced in Section 1.2.3.

For some $v \in \mathcal{C}^+$ let us consider the so-called *v-norm* on $\mathcal{F}$, as follows

$$\|g\|_v = \sup_{s \in \mathbb{S}} \frac{|g(s)|}{v(s)} = \inf\{c > 0 : \ |g(s)| \leq c \cdot v(s), \forall s \in \mathbb{S}\}.$$

In particular, for each $g \in \mathcal{F}$ it holds that[4]:

$$\forall s \in \mathbb{S} : \ |g(s)| \leq \|g\|_v \cdot v(s). \tag{1.17}$$

Example 1.4. *Let $\mathcal{C}_v$ be defined as in Example 1.1, for $\alpha = 1$. That is, $v(x) = e^x$, for $x \geq 0$. If $f(x) = 1 + x$, for $x \geq 0$, we have $f(x) \leq e^x$, for all $x \geq 0$ and*

$$\sup_{x \geq 0} f(x) e^{-x} = \lim_{x \downarrow 0} (1 + x) e^{-x} = 1.$$

Hence, $\|f\|_v = 1$. On the other hand, if $g(x) = x$ then $\|g\|_v = e^{-1}$ since

$$\sup_{x \geq 0} x e^{-x} = e^{-1}.$$

Remark 1.3. *The v-norm is also known as weighted supremum norm in the literature. An early reference is [42]. The v-norm is frequently used in Markov decision analysis. First traces date back to the early eighties, see [16] and the revised version which was published as [17]. It was originally used in analysis of Blackwell optimality; see [17], and [34] for a recent publication on this topic. Since then, it has been used in various forms under different names in many subsequent papers; see, for example, [35] and [44]. For the use of v-norm in the theory of measure-valued differentiation of Markov chains; see, e.g., [24]. For the use of v-norm in the context of strong stable Markov chains we refer to [35].*

For an arbitrary subset $\mathcal{D} \subset \mathcal{F}$ and $v \in \mathcal{C}^+$ let us denote by $[\mathcal{D}]_v$ the set of elements of $\mathcal{D}$ with finite v-norm, i.e.,

$$[\mathcal{D}]_v = \{g \in \mathcal{D} : \|g\|_v < \infty\} \tag{1.18}$$

and extend Definition 1.1 by calling the sequence $\{\mu_n\}_{n \in \mathbb{N}}$ *weakly $[\mathcal{D}]_v$-convergent* if there exists μ such that

$$\forall g \in [\mathcal{D}]_v : \ \lim_{n \to \infty} \int g(s) \mu_n(ds) = \int g(s) \mu(ds). \tag{1.19}$$

[4] Note that inequality in (1.17) still holds true if $\|g\|_v = \infty$.

The set $\mathcal{D}$ in (1.18) is called the *base set* of $[\mathcal{D}]_v$ and note that it can be chosen, without loss of generality, to be a linear subspace of $\mathcal{F}$. Moreover, the set $\mathcal{C}_v$ defined in (1.4) can be written as $[\mathcal{C}]_v$, i.e., $\mathcal{C}_v$-convergence introduced in Definition 1.1 is in fact $[\mathcal{C}]_v$-convergence and $[\mathcal{C}]_v = \mathcal{C}_B$, for any $v \in \mathcal{C}_B$, i.e., for $v \in \mathcal{C}_B$ we recover the classical weak convergence. In particular, if $v \equiv 1$ then the v-norm coincides with the supremum norm on $\mathcal{C}_B$.

As it will turn out, powerful results on convergence, continuity and differentiability of product measures can be established if the base set in (1.18) is such that $[\mathcal{D}]_v$ becomes a Banach space when endowed with the appropriate v-norm. This gives rise to the following definition.

Definition 1.2. *The pair $(\mathcal{D}, v)$ is called a **Banach base** on $\mathbb{S}$ if:*

(i) $\mathcal{D}$ is a linear space such that $\mathcal{C} \subset \mathcal{D} \subset \mathcal{F}$,

(ii) $v \in \mathcal{C}^+$ and the set $[\mathcal{D}]_v$ in (1.18) endowed with the v-norm is a Banach space.

In the following we present two examples of Banach bases that arise in applications.

Example 1.5. *The* continuity *paradigm: $\mathcal{D} = \mathcal{C}$. Taking $v \in \mathcal{C}^+$ we obtain $[\mathcal{C}]_v$ as the set of all continuous mappings bounded by v. It can be shown that $(\mathcal{C}, v)$ is a Banach base on $\mathbb{S}$. Indeed, the mapping*[5] *$\Phi : [\mathcal{C}(\mathbb{S})]_v \to \mathcal{C}_B(\mathbb{S}_v)$ defined as*

$$\forall s \in \mathbb{S}_v, g \in [\mathcal{C}(\mathbb{S})]_v : \ (\Phi g)(s) = \frac{g(s)}{v(s)} \tag{1.20}$$

establishes a linear bijection between two norm spaces and the inverse Φ^{-1} is given by

$$\forall s \in \mathbb{S}, g \in \mathcal{C}_B(\mathbb{S}_v) : \ (\Phi^{-1} g)(s) = g(s) \cdot v(s).$$

Furthermore, Φ is an isometry as it satisfies

$$\forall g \in [\mathcal{C}(\mathbb{S})]_v : \ \|\Phi g\| = \|g\|_v.$$

Since $\mathcal{C}_B(\mathbb{S}_v)$ is a Banach space when equipped with the supremum-norm, $[\mathcal{C}(\mathbb{S})]_v$ inherits the same property; see [56].

The measurability *paradigm: $\mathcal{D} = \mathcal{F}$. Taking $v \in \mathcal{C}^+$, we obtain $[\mathcal{F}]_v$ as the set of all measurable mappings bounded by v. Again, the linear mapping $\Phi : [\mathcal{F}(\mathbb{S})]_v \to \mathcal{F}_B(\mathbb{S}_v)$ defined by (1.20) is an isometry and we conclude that $(\mathcal{F}, v)$ is a Banach base on $\mathbb{S}$.*

As the above example shows, the functional spaces $[\mathcal{C}]_v$ and $[\mathcal{F}]_v$ are Banach bases for each $v \in \mathcal{C}^+$. Note that, the condition $\mathcal{C} \subset \mathcal{D}$ is a minimal prerequisite for developing our theory, since by Lemma 1.2 the space $[\mathcal{C}]_v$ posses satisfactory separation properties, while the condition $\mathcal{D} \subset \mathcal{F}$ comes naturally since we only deal with measurable functions. Therefore, if $(\mathcal{D}, v)$ is a Banach base then we have

$$[\mathcal{C}]_v \subset [\mathcal{D}]_v \subset [\mathcal{F}]_v.$$

[5] The assumption $v \in \mathcal{C}$ guarantees that the transformation Φ preserves continuity.

Remark 1.4. *Theorem F.2 (see the Appendix) shows that, for $\mathcal{D} = \mathcal{C}$, the set of functions g satisfying (1.19) includes some significant class of non-continuous, measurable mappings. Namely, if the sequence $\{\mu_n\}_n \subset \mathcal{M}^1$ is weakly $\mathcal{C}_B$-convergent to μ, i.e., (1.19) holds true for each $g \in \mathcal{C}_B$, then the class of functions g which satisfy (1.19) can be extended to $[\mathcal{C}(\mu)]_v$, for some v which is uniformly integrable with respect to the family $\{\mu_n : n \in \mathbb{N}\}$, where $\mathcal{C}(\mu)$ denotes the space of functions which are continuous μ-a.e.*

In the remainder of this thesis, we will impose the following assumption:

Whenever a Banach base $(\mathcal{D}, v)$ is considered, $\mathcal{D}$ is either $\mathcal{C}$ or $\mathcal{F}$.

The idea behind this assumption is that one should think of $\mathcal{D}$ as a class of functions enjoying some topological property rather than a simple set of functions. This is no severe restriction with respect to our applications; see Remark 1.4. In this setting, $[\mathcal{D}]_v$ spaces enjoy an important property which will be used in many proofs. Namely, if the function g belongs to the class $\mathcal{D}$ then a continuous transformation of g, i.e., the composition $f \circ g$ or the product $f \cdot g$, with f continuous, belongs also to the class $\mathcal{D}$.

Many statements in this thesis will be formulated in terms of $[\mathcal{D}]_v$ spaces, which means that they hold true for both $\mathcal{D} = \mathcal{C}$ and $\mathcal{D} = \mathcal{F}$, i.e., they generate two statements which are obtained by replacing $\mathcal{D}$ by $\mathcal{C}$ and $\mathcal{F}$, respectively. In most of the cases the proof does not distinguish between these two situations but, when necessary, the proof will be modified accordingly. As a final remark, since a weak $[\mathcal{F}]_v$ property implies the corresponding weak $[\mathcal{C}]_v$ property, in some statements we will replace $\mathcal{D}$ by $\mathcal{C}$, if possible, in order to make the result stronger.

1.3.3 Spaces of Measures

In functional analysis, signed measures often appear as continuous linear functionals on spaces of functions. More precisely, by the Riesz Representation Theorem (see Theorem F.3 in the Appendix) a space of measures can be seen as the *topological dual* of a certain space of functions. Throughout this section we aim to exploit this fact in order to derive new results using specific tools from Banach space theory.

Let $(\mathcal{D}, v)$ be a Banach base on $\mathbb{S}$ and let

$$\mathcal{M}_v := \left\{\mu \in \mathcal{M} : v \in \mathcal{L}^1(\mu)\right\}.$$

If $\alpha, \beta \in \mathbb{R}$ and $\mu, \nu \in \mathcal{M}_v$ then $\alpha \cdot \mu + \beta \cdot \nu \in \mathcal{M}_v$, where

$$\forall A \in \mathcal{S} : \ (\alpha \cdot \mu + \beta \cdot \nu)(A) = \alpha\mu(A) + \beta\nu(A).$$

Hence, $\mathcal{M}_v$ can be organized as a linear space. Moreover, note that we have

$$[\mathcal{D}]_v \subset \mathcal{L}^1(\mu_\theta : \theta \in \Theta) \Leftrightarrow v \in \mathcal{L}^1(\mu_\theta : \theta \in \Theta) \Leftrightarrow \{\mu_\theta : \theta \in \Theta\} \subset \mathcal{M}_v$$

and for $v = 1$ we have $\mathcal{M}_v = \mathcal{M}_B$, i.e., $\mathcal{M}_v$ consists of finite elements. The subset of $\mathcal{M}_v$ which consists of probability measures is denoted by $\mathcal{M}_v^1$, i.e.,

$$\mathcal{M}_v^1 := \mathcal{M}_v \cap \mathcal{M}^1$$

and note that if $v = 1$ then $\mathcal{M}_v^1 = \mathcal{M}_B \cap \mathcal{M}^1 = \mathcal{M}^1$.

For $\mu \in \mathcal{M}_v$ consider the Hahn-Jordan decomposition $\mu = [\mu]^+ - [\mu]^-$ and define the *weighted total variation norm* of μ with respect to v (shortly: v-norm), as follows:

$$\|\mu\|_v = \int v(s)|\mu|(ds) = \int v(s)[\mu]^+(ds) + \int v(s)[\mu]^-(ds). \tag{1.21}$$

In particular, a *Cauchy-Schwartz Inequality* holds for v-norms. In formula:

$$\forall g \in [\mathcal{D}]_v, \forall \mu \in \mathcal{M}_v : \left|\int g(s)\mu(ds)\right| \leq \|g\|_v \cdot \|\mu\|_v. \tag{1.22}$$

Note that, using the v-norm, the space $\mathcal{M}_v$ can be alternatively described as

$$\mathcal{M}_v = \{\mu \in \mathcal{M} : \ \|\mu\|_v < \infty\}$$

and for $v \equiv 1$ one recovers the total variation norm, given by (1.3).

On the other hand, for $\mu \in \mathcal{M}_v$ the application $\Phi_\mu^{\mathcal{D}} : [\mathcal{D}]_v \to \mathbb{R}$ defined as

$$\forall g \in [\mathcal{D}]_v : \ \Phi_\mu^{\mathcal{D}}(g) = \int g(s)\mu(ds)$$

is a linear functional on the space $[\mathcal{D}]_v$, whose operator norm satisfies

$$\left\|\Phi_\mu^{\mathcal{D}}\right\|_v = \sup\left\{\left|\Phi_\mu^{\mathcal{D}}(g)\right| : \ g \in [\mathcal{D}]_v, \|g\|_v \leq 1\right\} = \|\mu\|_v.$$

To see this, note that if A is a set such that $[\mu]^+(\complement A) = [\mu]^-(A) = 0$ and

$$\forall s \in \mathbb{S} : \ g^*(s) := v(s)\mathbb{I}_A(s) - v(s)\mathbb{I}_{\complement A}(s)$$

it follows that g^* is measurable, $\|g^*\|_v = 1$ and $\|\mu\|_v = |\int g^*(s)\mu(ds)|$. Hence,

$$\|\mu\|_v = \left|\int g^*(s)\mu(ds)\right| \leq \left\|\Phi_\mu^{\mathcal{F}}\right\|_v. \tag{1.23}$$

Moreover, using Urysohn's Lemma it can be shown that there exists some sequence of continuous functions $\{f_n\}_n$ such that $|f_n(s)| \leq 1$, for all n and s and such that

$$\forall s \in \mathbb{S} : \ \lim_{n\to\infty} f_n(s) = \mathbb{I}_A(s) - \mathbb{I}_{\complement A}(s).$$

Hence, if we define $g_n(s) = f_n(s)v(s)$, for each n and s, we have $g_n \in \mathcal{C}$, $\|g_n\|_v \leq 1$, for each n, and by the Dominated Convergence Theorem (see Theorem C.1 in the Appendix) we have

$$\|\mu\|_v = \left|\int g^*(s)\mu(ds)\right| = \lim_{n\to\infty} \left|\int g_n(s)\mu(ds)\right| \leq \left\|\Phi_\mu^{\mathcal{C}}\right\|_v. \tag{1.24}$$

On the other hand, from the Cauchy-Schwarz Inequality we conclude that

$$\left\|\Phi_\mu^{\mathcal{D}}\right\|_v \leq \|\mu\|_v,$$

which, together with (1.23) and (1.24) leads to

$$\|\Phi_\mu^{\mathcal{D}}\|_v = \|\mu\|_v.$$

Therefore, any element μ in $\mathcal{M}_v$ can be identified by a continuous linear functional $\Phi_\mu^{\mathcal{D}}$ on the space $[\mathcal{D}]_v$ and the operator norm of $\Phi_\mu^{\mathcal{D}}$ coincides with the weighted total variation norm of μ, given by (1.21). It follows that $\mathcal{M}_v$ is a subset of the topological dual of $[\mathcal{D}]_v$ and the weak $[\mathcal{D}]_v$-convergence is in fact the convergence given by the trace of the weak-* topology on $\mathcal{M}_v$. However, for ease of exposure, we agree to call it "weak" since we will not make any reference to the actual weak topology, induced on $[\mathcal{D}]_v$ by its topological dual, so no confusion occurs.

As discussed in Section 1.3.1, norm convergence on $\mathcal{M}_v$ implies weak convergence. In this case, this is a consequence of the Cauchy-Schwartz Inequality. Indeed, if μ_n converges in v-norm to μ, letting $\nu = \mu_n - \mu$ in (1.22), it follows that (1.19) holds true for all $g \in [\mathcal{D}]_v$. The converse is, however, not true as detailed in the following example.

Example 1.6. *Consider the convergent sequence $\{x_n\}_n \subset \mathbb{R}$ having limit $x \in \mathbb{R}$. It is known that the sequence of corresponding Dirac distributions $\{\delta_{x_n}\}_n \subset \mathcal{M}$ is weakly $\mathcal{C}_B$-convergent to δ_x. However, norm convergence does not hold since*

$$\lim_{n\to\infty} \|\delta_{x_n} - \delta_x\| = \lim_{n\to\infty} \sup_{|g|\le 1} |g(x_n) - g(x)| = 2 \neq 0.$$

In the following, we endow $\mathcal{M}_v$ with the weak-* topology given by $[\mathcal{D}]_v$-convergence (we omit specifying $[\mathcal{D}]_v$ when not relevant) and refer to the v-norm convergence as *strong convergence* . Consequently, by continuity we mean weak continuity, i.e., with respect to weak-* topology and by strong continuity we mean continuity with respect to the v-norm convergence.

We continue our analysis by presenting a few results which can be easily derived by using a functional analytic approach to spaces of measures. For instance, the Banach-Steinhaus Theorem can be applied to a convergent sequence μ_n of measures which allows to deduce that the family $\{\mu_n\}_n$ is strongly bounded in $\mathcal{M}_v$. For later reference we formalize this statement in the following lemma.

Lemma 1.4. *Let $(\mathcal{D}, v)$ be a Banach base and let the sequence $\{\mu_n\}_n$ converge to some limit μ in $\mathcal{M}_v$. Then, it holds that*

$$\sup_{n\in\mathbb{N}} \|\mu_n\|_v < \infty.$$

Proof. Under the assumption in the lemma, the set $\{\mu_n : n \in \mathbb{N}\}$ is bounded in the weak sense, i.e., for each $g \in [\mathcal{D}]_v$, the set $\{\int g d\mu_n : n \in \mathbb{N}\}$ is bounded in $\mathbb{R}$. The claim then follows from the Banach-Steinhaus Theorem (see Theorem G.1 in the Appendix). □

Recall now the definition of regular convergence given in Section 1.2.4. As illustrated by Example 1.3, convergence of a sequence μ_n towards some limit μ does not imply regular convergence. The following result will show that under some additional condition the positive part of μ_n converge to the positive parts of μ.

Lemma 1.5. *Let $(\mathcal{D}, v)$ be a Banach base and let the sequence $\{\mu_n\}_n$ converge to some limit μ in $\mathcal{M}_v$. Then, the sequence $\{\mu_n\}_n$ converges regularly to μ if and only if*

$$\lim_{n\to\infty} \|\mu_n\|_v = \|\mu\|_v. \tag{1.25}$$

Proof. The direct implication is immediate. Assume now that the sequence $\{\mu_n\}_n$ converges to μ and (1.25) holds true. Lemma 1.4 implies that the family $\{\mu_n : n \in \mathbb{N}\}$ is strongly bounded in $\mathcal{M}_v$ and so is $\{[\mu_n]^+ : n \in \mathbb{N}\}$, since $\|[\mu_n]^+\|_v \leq \|\mu_n\|_v$. Therefore, in accordance with the Banach-Alaoglu Theorem (see Theorem G.2), it follows that the closure of the set $\{[\mu_n]^+ : n \in \mathbb{N}\}$ is compact in the weak-* topology and there exist a subsequence $\{n_k\}_{k\geq 1} \subset \mathbb{N}$ such that $\{[\mu_{n_k}]^+\}_k$ converges in $\mathcal{M}_v$.

Next, we show that any convergent subsequence of $\{[\mu_n]^+ : n \in \mathbb{N}\}$ converges to $[\mu]^+$. Indeed, choose an arbitrary convergent subsequence $\{[\mu_{n_k}]^+\}_k$ and denote by $\lambda \in \mathcal{M}^+$ its limit. Since $[\mu_{n_k}]^- = [\mu_{n_k}]^+ - \mu_{n_k}$ it follows that $[\mu_{n_k}]^-$ converges to $\lambda - \mu$. Moreover, $(\lambda - \mu) \in \mathcal{M}^+$ since it is the limit of a sequence of positive measures. The uniqueness of the limit implies that $\mu = \lambda - (\lambda - \mu)$ and from the minimality property of the Hahn-Jordan decomposition we conclude that there exists some $\nu \in \mathcal{M}^+$ such that $\lambda = \nu + [\mu]^+$ and $\lambda - \mu = \nu + [\mu]^-$. Consequently,

$$\lim_{k\to\infty} \|\mu_{n_k}\|_v = \|\mu\|_v + 2\|\nu\|_v$$

and by hypothesis it follows that $\|\nu\|_v = 0$. Therefore, from Lemma 1.2 it follows that ν is the null measure, i.e., $\lambda = [\mu]^+$, which concludes the proof. □

Remark 1.5. *The proof of Lemma 1.5 indicates that if μ_n converges to μ then it holds that*

$$\|\mu\|_v \leq \liminf_n \|\mu_n\|_v.$$

Therefore, another equivalent condition for regular convergence is

$$\|\mu\|_v \geq \limsup_n \|\mu_n\|_v.$$

An immediate consequence of Lemma 1.5 is the following result.

Corollary 1.1. *Under the conditions put forward in Lemma 1.5, if the sequence $\{\mu_n\}_n$ converges strongly to μ then it converges regularly to μ.*

Proof. First, note that the following inequality holds true:

$$\forall n \in \mathbb{N} : \ |\|\mu_n\|_v - \|\mu\|_v| \leq \|\mu_n - \mu\|_v.$$

Now the proof follows from Lemma 1.5 □

We say that the continuous measure-valued mapping μ_* is *regularly continuous* at θ if the mapping $[\mu_*]^+$ is continuous at θ. It follows that $[\mu_*]^-$ is continuous at θ, as well. The statements in Lemma 1.4 and Lemma 1.5 can be easily extended to arbitrary families of measures. More specifically, the following statement holds true.

Theorem 1.2. *Let $\mu_* : \Theta \to \mathcal{M}_v$ be a continuous measure-valued mapping.*

(i) Then for each compact $K \subset \Theta$ it holds that

$$\sup_{\theta \in K} \|\mu_\theta\|_v < \infty.$$

(ii) In addition, if the real-valued mapping $\|\mu_*\|_v$ is continuous at θ then the measure-valued mapping μ_* is regularly continuous at θ. In particular, the same conclusion holds true when μ_* is strongly continuous.

Proof. (i) By hypothesis, for each compact $K \subset \Theta$ it holds that

$$\forall g \in [\mathcal{D}]_v : \sup_{\theta \in K} \left| \int g(s) \mu_\theta(ds) \right| < \infty.$$

Assuming that there exists some compact $K' \subset \Theta$ such that $\sup_{K'} \|\mu_\theta\|_v = \infty$ it follows that there exists a sequence $\{\theta_n\}_n$ in K' such that $\sup_n \|\mu_{\theta_n}\|_v = \infty$, which contradicts Lemma 1.4.

(ii) Assuming, for instance, that $[\mu_*]^+$ is not continuous at θ it follows that there exists a sequence $\xi_n \to 0$ such that $[\mu_{\theta+\xi_n}]^+$ does not converge to $[\mu_\theta]^+$, which contradicts Lemma 1.5. A similar reasoning as in Corollary 1.1 concludes the proof. □

1.3.4 Banach Bases on Product Spaces

Let $\mathbb{S}, \mathbb{T}$ be separable complete metric spaces endowed with Borel fields $\mathcal{S}$ and $\mathcal{T}$ and Banach bases, $(\mathcal{D}(\mathbb{S}), v)$ and $(\mathcal{D}(\mathbb{T}), u)$, respectively and consider the class of mappings $g : \mathbb{S} \times \mathbb{T} \to \mathbb{R}$ satisfying

$$\forall s \in \mathbb{S}, t \in \mathbb{T} : \ g(s, \cdot) \in \mathcal{D}(\mathbb{T}), g(\cdot, t) \in \mathcal{D}(\mathbb{S}). \tag{1.26}$$

In addition, let us define the *tensor* product $v \otimes u : \mathbb{S} \times \mathbb{T} \to \mathbb{R}$ as follows:

$$\forall s \in \mathbb{S}, t \in \mathbb{T} : \ (v \otimes u)(s, t) = v(s) \cdot u(t). \tag{1.27}$$

Let us denote by $\mathcal{D}(\mathbb{S}) \otimes \mathcal{D}(\mathbb{T})$ the class of functions $g \in \mathcal{F}(\mathbb{S} \times \mathbb{T})$ satisfying condition (1.26) which, as the following example shows, imposes no restriction in applications.

Example 1.7. *We revisit the Banach bases introduced in Example 1.5.*

- *Let $g \in \mathcal{C}(\mathbb{S} \times \mathbb{T})$. Then*

$$\forall s \in \mathbb{S}, t \in \mathbb{T} : \ g(s, \cdot) \in \mathcal{C}(\mathbb{T}), \ g(\cdot, t) \in \mathcal{C}(\mathbb{S})$$

and it follows that

$$\mathcal{C}(\mathbb{S} \times \mathbb{T}) \subset \mathcal{C}(\mathbb{S}) \otimes \mathcal{C}(\mathbb{T}). \tag{1.28}$$

- *Let $g \in \mathcal{F}(\mathbb{S} \times \mathbb{T})$. Then*

$$\forall s \in \mathbb{S}, t \in \mathbb{T} : \ g(s, \cdot) \in \mathcal{F}(\mathbb{T}), \ g(\cdot, t) \in \mathcal{F}(\mathbb{S})$$

and it follows that

$$\mathcal{F}(\mathbb{S} \times \mathbb{T}) \subset \mathcal{F}(\mathbb{S}) \otimes \mathcal{F}(\mathbb{T}), \tag{1.29}$$

We define now the product of $(\mathcal{D}(\mathbb{S}), v)$ and $(\mathcal{D}(\mathbb{T}), u)$ as follows:

$$(\mathcal{D}(\mathbb{S}) \otimes \mathcal{D}(\mathbb{T}), v \otimes u).$$

The next result shows that products of Banach bases are again Banach bases, where the above definitions are extended to finite products in the obvious way.

Theorem 1.3. *Let $(\mathcal{D}(\mathbb{S}_i), v_i,)$ be Banach bases, respectively, for $1 \leq i \leq k$.*

(i) Then the pair

$$(\mathcal{D}(\mathbb{S}_1) \otimes \cdots \otimes \mathcal{D}(\mathbb{S}_k), v_1 \otimes \ldots \otimes v_k)$$

is a Banach base on $\mathbb{S}_1 \times \cdots \times \mathbb{S}_k$. In particular, for all $1 \leq i \leq k$

$$\forall s_j \in \mathbb{S}_j, j \neq i: \; g(s_1, \ldots, s_{i-1}, \cdot, s_{i+1}, \ldots, s_k) \in [\mathcal{D}(\mathbb{S}_i)]_{v_i},$$

provided that $g \in [\mathcal{D}(\mathbb{S}_1) \otimes \cdots \otimes \mathcal{D}(\mathbb{S}_k)]_{v_1 \otimes \ldots \otimes v_k}$.

(ii) If for each $1 \leq i \leq k$, $\mathcal{S}_i$ is the Borel field on $\mathbb{S}_i$ and $\mu_i \in \mathcal{M}_{v_i}(\mathcal{S}_i)$ then

$$\|\mu_1 \times \ldots \times \mu_k\|_{v_1 \otimes \ldots \otimes v_k} \leq \|\mu_1\|_{v_1} \cdot \ldots \cdot \|\mu_k\|_{v_k}.$$

In particular, $\mu_1 \times \ldots \times \mu_k \in \mathcal{M}_{v_1 \otimes \ldots \otimes v_k}(\sigma(\mathcal{S}_1 \times \ldots \times \mathcal{S}_k))$[6]*.*

Proof. (i) The proof follows by finite induction with respect to k and we only provide a proof for the case $k = 2$. More precisely, we prove the following: let $(\mathcal{D}(\mathbb{S}), v)$ and $(\mathcal{D}(\mathbb{T}), u)$ be Banach bases on $\mathbb{S}$ and $\mathbb{T}$, respectively, then $(\mathcal{D}(\mathbb{S}) \otimes \mathcal{D}(\mathbb{T}), v \o times u)$ is a Banach base on the product space $\mathbb{S} \times \mathbb{T}$; moreover, if $g \in [\mathcal{D}(\mathbb{S}) \otimes \mathcal{D}(\mathbb{T})]_{v \otimes u}$, then $g(s, \cdot) \in \mathcal{D}(\mathbb{T})$ and $g(\cdot, t) \in \mathcal{D}(\mathbb{S})$ for all $s \in \mathbb{S}, t \in \mathbb{T}$. To this end we verify the conditions in Definition 1.2. It is immediate that $\mathcal{D}(\mathbb{S}) \otimes \mathcal{D}(\mathbb{T})$ is a linear space, satisfying

$$\mathcal{C}_B(\mathbb{S} \times \mathbb{T}) \subset \mathcal{D}(\mathbb{S}) \otimes \mathcal{D}(\mathbb{T}) \subset \mathcal{F}(\mathbb{S} \times \mathbb{T}).$$

For the second part, one proceeds as follows: Let $g \in [\mathcal{D}(\mathbb{S}) \otimes \mathcal{D}(\mathbb{T})]_{v \otimes u}$. It follows that

$$\sup_{t \in \mathbb{T}} \frac{\|g(\cdot, t)\|_v}{u(t)} = \sup_{t \in \mathbb{T}} \sup_{s \in \mathbb{S}} \frac{|g(s, t)|}{v(s) \cdot u(t)} \leq \sup_{(s,t)} \frac{|g(s, t)|}{v(s) \cdot u(t)} = \|g\|_{v \otimes u} < \infty. \tag{1.30}$$

Thus, for $t \in \mathbb{T}$ we have $\|g(\cdot, t)\|_v \leq \|g\|_{v \otimes u} \cdot u(t) < \infty$ which shows that $g(\cdot, t) \in [\mathcal{D}(\mathbb{S})]_v$. By symmetry, we obtain $g(s, \cdot) \in [\mathcal{D}(\mathbb{T})]_u$, for all $s \in \mathbb{S}$.

Next, we show that $[\mathcal{D}(\mathbb{S}) \otimes \mathcal{D}(\mathbb{T})]_{v \otimes u}$ is a Banach space with respect to $v \otimes u$-norm. To this end, let $\{g_n\}_n$ be a Cauchy sequence in $[\mathcal{D}(\mathbb{S}) \otimes \mathcal{D}(\mathbb{T})]_{v \otimes u}$. That means that for each $\epsilon > 0$, there exist a rank $n_\epsilon \geq 1$, such that for all $j, k \geq n_\epsilon$ it holds that $\|g_j - g_k\|_{v \otimes u} \leq \epsilon$. Inserting now $g = g_j - g_k$ in (1.30) one obtains for $j, k \geq n_\epsilon$

$$\forall t \in \mathbb{T}: \; \|g_j(\cdot, t) - g_k(\cdot, t)\|_v \leq \|g_j - g_k\|_{v \otimes u} \cdot u(t) \leq \epsilon \cdot u(t).$$

[6] Here $\sigma(\mathcal{S}_1 \times \ldots \times \mathcal{S}_k)$ denotes the σ-field generated by the product $\mathcal{S}_1 \times \ldots \times \mathcal{S}_k$.

Hence, for all $t \in \mathbb{T}$, $\{g_n(\cdot,t)\}_n$ is a Cauchy sequence in the Banach space $[\mathcal{D}(\mathbb{S})]_v$, thus convergent to some limit $\bar{g}(\cdot,t) \in [\mathcal{D}(\mathbb{S})]_v$. Using again a symmetry argument we deduce that $\bar{g}(s,\cdot) \in [\mathcal{D}(\mathbb{T})]_u$, for all $s \in \mathbb{S}$, and we conclude that $\bar{g} \in \mathcal{D}(\mathbb{S}) \otimes \mathcal{D}(\mathbb{T})$.

Finally, we show that $\bar{g}$ is the $v \otimes u$-norm limit of the sequence $\{g_n\}_n$. Choose $\epsilon > 0$ and $n_\epsilon \geq 1$ such that for all $j, k \geq n_\epsilon$ we have $\|g_j - g_k\|_{v\otimes u} < \epsilon$; more explicitly:

$$\forall s \in \mathbb{S}, t \in \mathbb{T}: \ |g_j(s,t) - g_k(s,t)| < \epsilon \cdot v(s)u(t),$$

for all $j, k \geq n_\epsilon$. Letting now $k \to \infty$ in the above inequality yields

$$\forall s \in \mathbb{S}, t \in \mathbb{T}: \ |g_j(s,t) - \bar{g}(s,t)| \leq \epsilon \cdot v(s)u(t),$$

for all $j \geq n_\epsilon$, which is equivalent to $\|g_j - \bar{g}\|_{v\otimes u} \leq \epsilon$ for all $j \geq n_\epsilon$. Therefore, it follows that $\|\bar{g}\|_{v\otimes u} < \infty$, i.e., $\bar{g} \in [\mathcal{D}(\mathbb{S}) \otimes \mathcal{D}(\mathbb{T})]_{v\otimes u}$ and since ϵ was chosen arbitrarily we conclude that $\lim_{n\to\infty} \|g_n - \bar{g}\|_{v\otimes u} = 0$ which proves the claim.

(ii) To prove the second statement

$$\|\mu \times \eta\|_{v\otimes u} \leq \|\mu\|_v \, \|\eta\|_u, \tag{1.31}$$

To this end, let $\mu = [\mu]^+ - [\mu]^-$ and $\eta = [\eta]^+ - [\eta]^-$ be the Hahn-Jordan decompositions of μ and η, respectively. Then

$$\mu \times \eta = ([\mu]^+ \times [\eta]^+ + [\mu]^- \times [\eta]^-) - ([\mu]^+ \times [\eta]^- + [\mu]^- \times [\eta]^+)$$

is a decomposition of $\mu \times \eta$ and the minimality property of Hahn-Jordan decomposition ensures that

$$[\mu \times \eta]^+ \leq [\mu]^+ \times [\eta]^+ + [\mu]^- \times [\eta]^-, \quad [\mu \times \eta]^- \leq [\mu]^+ \times [\eta]^- + [\mu]^- \times [\eta]^+.$$

By adding up the above inequalities we obtain

$$|\mu \times \eta| \leq |\mu| \times |\eta|.$$

Thus, according to (1.21) it holds that (use Fubini Theorem; see Theorem E.1 in the Appendix)

$$\|\mu \times \eta\|_{v\otimes u} \leq \int (v \otimes u)(s,z) \, (|\mu| \times |\eta|) \, (ds, dz) = \|\mu\|_v \, \|\eta\|_u,$$

which establishes (1.31). □

1.4 Concluding Remarks

When the metric space $\mathbb{S}$ is compact, the Riesz Representation Theorem (see, e.g, [19]) asserts that the space $\mathcal{M}_B$ of finite Radon measures on $\mathcal{S}$ is isometric to the topological dual of $\mathcal{C}_B$, when endowed with the supremum norm defined by (1.13). Such a result does not hold true in general and $\mathcal{M}_B$ is isometric to a proper subspace of the topological dual space $(\mathcal{C}_B)^*$. Nevertheless, when $\mathbb{S}$ is locally compact, it has been shown in [11] that $\mathcal{M}_B$

is precisely the topological dual of $\mathcal{C}_B$ endowed with the so-called *strict* (compact open) topology, i.e., the locally convex topology generated by the family of semi-norms

$$\|f\|_K = \sup_{s\in K}|f(s)|,$$

where K ranges over the compact subsets[7] of $\mathbb{S}$. Moreover, the topological dual of $\mathcal{C}_B$, when endowed with the supremum norm topology, is the space of Radon measures on the *Stone compactification* of $\mathbb{S}$; see, e.g., [41], [60]. Therefore, tightness of a family of elements in $\mathcal{M}_B$ is a technical condition which ensures that all the limit points in the weak-* topology are contained in $\mathcal{M}_B$; see Prokhorov Theorem (Theorem F.3 in the Appendix). More specifically, if $\mathcal{P} \subset \mathcal{M}_B$ is tight then the closure $\overline{\mathcal{P}}$ of $\mathcal{P}$ in the weak-* topology satisfies $\overline{\mathcal{P}} \subset \mathcal{M}_B$. A standard example which illustrates this fact is the following.

Example 1.8. *Let us consider the family of Dirac measures $\{\delta_x : x \geq 0\}$. Then, the classical weak limit $\lim_{x\to\infty} \delta_x$ does not exist in $\mathcal{M}_B(\mathbb{R})$. Indeed,*

$$\forall g \in \mathcal{C}_B : \lim_{x\to\infty} \int g(s)\delta_x(ds) = \lim_{x\to\infty} g(x),$$

but the right-hand side limit above does not exist in general. Hence, the closure of the family $\mathcal{P}_{x_0} := \{\delta_x : x \geq x_0\}$ in the weak- topology is not contained in $\mathcal{M}_B(\mathbb{R})$, for any $x_0 \geq 0$. This stems from the fact that the family $\mathcal{P}_{x_0}$ is not tight. However, note that for $v(s) = 1/s$ the $[\mathcal{C}]_v$-limit of the family δ_x, for $x \to \infty$, is the null measure.*

$[\mathcal{C}]_v$-spaces appear as particular cases of *weighted spaces* which are introduced by means of the so-called *Nachbin families* of functions; see, e.g., [46]. A Nachbin family is, in fact, a family $\mathcal{N}$ of upper-semi-continuous functions which is upper directed, i.e., for each $v_1, v_2 \in \mathcal{N}$ there exist $\alpha > 0$ and $v \in \mathcal{N}$ such that

$$\forall s \in \mathbb{S} : \max\{v_1(s), v_2(s)\} \leq \alpha v(s).$$

Then, the weighted space generated by the family $\mathcal{N}$ is defined as the class of continuous functions g for which $g \cdot v$ is bounded for each $v \in \mathcal{N}$ and that becomes a topological vector space when endowed with the locally convex topology generated by the family semi-norms

$$\forall v \in \mathcal{N} : \|g\|_v := \sup_{s\in\mathbb{S}} v(s)|g(s)|.$$

Therefore, when $\mathcal{N} = \{\alpha/v : \alpha > 0\}$, for some $v \in \mathcal{C}^+$, one recovers the definition of the $[\mathcal{C}]_v$-space.

Weighted spaces have received a thorough treatment in [50], [51], [57], [58]. For instance, a result regarding the completeness of weighted spaces has been presented in [51] and an extension of the Stone-Weierstrass Theorem to weighted spaces has been discussed in [50]. Moreover, [57] addresses the problem of determining the topological dual of a weighted space. In particular, it turns out that the topological dual of a $[\mathcal{C}]_v$-space

[7] Local compactness of $\mathbb{S}$ implies that the above family of semi-norms is separating.

includes the space $\mathcal{M}_v$, which can alternatively described as the class of measures $\mu \in \mathcal{M}$ such that $v \cdot \mu$ is a finite, where, for arbitrary $\mu \in \mathcal{M}$, we define $v \cdot \mu \in \mathcal{M}$ as follows:

$$\forall s \in \mathbb{S}: \; (v \cdot \mu)(ds) = v(s)\mu(ds).$$

The reasoning essentially relies on the isometry between $[\mathcal{C}(\mathbb{S})]_v$ and $\mathcal{C}_B(\mathbb{S})$, defined by (1.20), which induces an isometry between corresponding spaces of measures. For later reference we synthesize these observations into the following remark.

Remark 1.6. *Inspired by Example 1.5 we note that $[\mathcal{C}(\mathbb{S})]_v$-convergence is equivalent to $\mathcal{C}_B(\mathbb{S}_v)$-convergence, i.e., the sequence $\{\mu_n\}_n$ is $[\mathcal{C}(\mathbb{S})]_v$-convergent to μ if and only if $\{v \cdot \mu_n\}_n$ is $\mathcal{C}_B(\mathbb{S}_v)$-convergent to $v \cdot \mu$*

The most important gain of strong convergence is that the limit relation in (1.8) holds uniformly in $g \in [\mathcal{C}]_v$, $\|g\|_v \leq 1$. Nevertheless, as shown in [52], on a $\mathcal{C}_v$-space weak convergence of measures is equivalent to uniform convergence of integrals with respect to equicontinuous families of functions, i.e., relatively compact subsets $\mathcal{K} \subset [\mathcal{C}]_v$. In general, weak differentiation is strictly weaker than strong differentiability, i.e., the weak*-topology is strictly coarser than the norm topology; see Example 1.6.

2. MEASURE-VALUED DIFFERENTIATION

This chapter is devoted to a detailed analysis of the concept of measure-valued differentiation and its applicability. New results will be established, by combining functional analytic and measure theoretical techniques and some applications will be provided.

2.1 Introduction

Measure-valued differentiation can be described in a general setting as follows: Consider a family $\{\Phi_\theta : \theta \in \Theta\}$ of linear functionals on some Banach space $\mathcal{V}$, where Θ is an open connected subset of $\mathbb{R}$. For fixed $\theta \in \Theta$, provided that for each $\mathbf{x} \in \mathcal{V}$ the limit

$$\Phi'_\theta(\mathbf{x}) := \lim_{\xi \to 0} \frac{\Phi_{\theta+\xi}(\mathbf{x}) - \Phi_\theta(\mathbf{x})}{\xi} \tag{2.1}$$

exists in $\mathbb{R}$, it follows that Φ'_θ is a linear operator on $\mathcal{V}$. Therefore, the formal derivative $\frac{d}{d\theta}\Phi_\theta$ has the following operator representation

$$\forall \mathbf{x} \in \mathcal{V}: \ \frac{d}{d\theta}\Phi_\theta(\mathbf{x}) = \Phi'_\theta(\mathbf{x}).$$

If $\mathcal{V}$ is a space of functions and Φ_θ is an integral operator, i.e., it can be represented as the integral with respect to some measure μ_θ, then we obtain a sensible concept of measure-valued differentiation.

Provided that the limit in (2.1) exists for each $\mathbf{x} \in \mathcal{V}$ it follows that

$$\lim_{\xi \to 0} \frac{\Phi_{\theta+\xi} - \Phi_\theta}{\xi} = \Phi'_\theta, \tag{2.2}$$

where the above convergence holds in the weak-* topology. Therefore, following the terminology in Definition 1.1 it is natural to call the differentiability concept described by (2.1) *weak differentiability*. This concept was first introduced in [47] for $\mathcal{V} = \mathcal{C}_B$. The general definition for $\mathcal{V} = [\mathcal{D}]_v$, is postponed to Section 2.2.1.

It is also possible to define a concept of *strong differentiability* by requiring that the limit relation in (2.2) holds in a strong (norm) sense. As explained in Section 1.4, strong differentiability, which relies on strong convergence, allows for a more powerful analysis. Nevertheless, weak differentiability is the minimal condition for (2.1) to hold true for each $\mathbf{x} \in \mathcal{V}$, which makes it attractive for applications. The aim of this chapter is to study both types of differentiability and their range of application. In addition, the concept of *regular differentiability* will be introduced to ensure a smooth extension of the properties of the classical weak convergence of positive measures to signed measures. As it will turn out, regular differentiability is a stronger property than weak differentiability, weaker than strong differentiability and it is fulfilled by the usual weakly differentiable distributions.

The chapter is organized as follows: In Section 2.2 the concept of measure-valued differentiation is discussed. In particular, we provide a representation of the weak derivative of a probability measure which will be crucial for our further analysis. Weak differentiability of product measures is treated in Section 2.3 while in Section 2.4 we investigate the relation between weak differentiability and set-wise differentiation. Eventually, in Section 2.5 we illustrate by means of two examples how weak derivatives lead to gradient estimators for some common applications.

2.2 The Concept of Measure-Valued Differentiation

In what follows we assume that $(\mathcal{D}, v)$ is a Banach base and $\{\mu_\theta : \theta \in \Theta\} \subset \mathcal{M}_v(\mathcal{S})$ is a family of (signed) measures, where Θ is an open connected subset of $\mathbb{R}$. In Section 2.2.1 we define and study several types of measure-valued differentiation and in Section 2.2.2 we discuss convenient representations of weak derivatives of probability measures. Eventually, in Section 2.2.3 we establish some results which prove to be useful when assessing weak differentiability and computing weak derivatives. We illustrate the results by several examples of weakly differentiable (usual) distributions.

2.2.1 Weak, Strong and Regular Differentiability

We now define the concept of weak differentiability.

Definition 2.1. *Let $(\mathcal{D}, v)$ be a Banach base on $\mathbb{S}$. We say that the mapping $\mu_* : \Theta \to \mathcal{M}_v$ is weakly $[\mathcal{D}]_v$-differentiable at θ or, μ_θ is weakly differentiable for short, if there exists $\mu'_\theta \in \mathcal{M}_v$ such that*

$$\forall g \in [\mathcal{D}]_v : \lim_{\xi \to 0} \frac{1}{\xi} \left(\int g(s)\mu_{\theta+\xi}(ds) - \int g(s)\mu_\theta(ds) \right) = \int g(s)\mu'_\theta(ds). \qquad (2.3)$$

Consequently, we call μ'_θ the weak derivative[1] *of μ_θ. If the left-hand side of the above equation equals zero for all $g \in [\mathcal{D}]_v$, then we say that the weak derivative of μ_θ is not significant.*

In addition, we say that μ_* is weakly $[\mathcal{D}]_v$-differentiable if μ_θ is weakly $[\mathcal{D}]_v$-differentiable for each $\theta \in \Theta$ and we denote by μ'_* the mapping

$$\forall \theta \in \Theta : \ \mu'_*(\theta) = \mu'_\theta.$$

Remark 2.1. *As mentioned before, differentiability of probability measures in the sense of Definition 2.1 was originally introduced for $[\mathcal{D}]_v = \mathcal{C}_B$ in [47] and received a thorough treatment in [48]. Other early traces are [39] and [40]. In [31], this concept is extended to general $[\mathcal{D}]_v$-differentiability and it is shown that $[\mathcal{D}]_v$-derivatives yield efficient unbiased gradient estimators. A recent result in this line of research shows that this class of gradient estimators can outperform single-run estimators such as those provided by' infinitesimal perturbation analysis; see [33].*

[1] Note that a weak derivative is unique in the sense specified by Lemma 1.2.

Note that in Definition 2.1 equation (2.3) is equivalent to

$$\frac{\mu_{\theta+\xi} - \mu_\theta}{\xi} \stackrel{[\mathcal{D}]_v}{\Longrightarrow} \mu'_\theta, \tag{2.4}$$

i.e., $(\mu_{\theta+\xi} - \mu_\theta)/\xi$ converges weakly, in $[\mathcal{D}]_v$ sense, to μ'_θ. Consequently, we say that μ_θ is *regularly* $[\mathcal{D}]_v$*-differentiable* (shortly: *regularly differentiable*) if the convergence in (2.4) is regular and we say that μ_θ is *strongly* $[\mathcal{D}]_v$*-differentiable* (shortly: *strongly differentiable*) if the convergence in (2.4) holds in the strong (v-norm) sense, i.e.,

$$\lim_{\xi\to 0} \left\| \frac{\mu_{\theta+\xi} - \mu_\theta}{\xi} - \mu'_\theta \right\|_v = 0 \tag{2.5}$$

Strong differentiability implies weak differentiability since (2.5) implies that (2.3) holds true for each $g \in [\mathcal{D}]_v$. However, strong differentiability is a more powerful tool since it implies that (2.3) holds true uniformly with respect to $g \in [\mathcal{D}]_v$, with $\|g\|_v \leq 1$. Indeed, (2.5) is equivalent to

$$\lim_{\xi\to 0} \sup_{\|g\|_v \leq 1} \left| \frac{1}{\xi} \left(\int g(s)\mu_{\theta+\xi}(ds) - \int g(s)\mu_\theta(ds) \right) - \int g(s)\mu'_\theta(ds) \right| = 0.$$

However, the converse is not true since, in general, there exist weakly differentiable distributions which are not strongly differentiable as will be illustrated by an example; see Example 2.6 later on in this section. Moreover, by Theorem 1.2 (*ii*) we conclude that regular differentiability is equivalent to

$$\lim_{\xi\to 0} \left\| \frac{\mu_{\theta+\xi} - \mu_\theta}{\xi} \right\|_v = \|\mu'_\theta\|_v. \tag{2.6}$$

and strong differentiability implies regular differentiability which, by definition, implies weak differentiability.

We continue our analysis by presenting two results which will establish connections between the three types of convergence/differentiability on $\mathcal{M}_v$.

The first result will show that weak differentiability implies strong continuity. This result will be particulary useful in Chapter 3 when we establish strong bounds on perturbations. The precise statement is as follows.

Theorem 2.1. *Let* $\mu_* : \Theta \to \mathcal{M}_v$ *be a* $[\mathcal{D}]_v$*-continuous measure-valued mapping such that* μ_θ *is* $[\mathcal{D}]_v$*-differentiable. Then for each closed neighborhood* V *of* 0*, such that* $\theta + \xi \in \Theta$ *for each* $\xi \in V$ *there exists some* $M > 0$ *such that*

$$\forall \xi \in V: \ \|\mu_{\theta+\xi} - \mu_\theta\|_v \leq M|\xi|.$$

In words, μ_θ *is* v*-norm continuous.*

Proof. For ξ such that $\theta + \xi \in \Theta$ let us define the measure-valued mapping

$$\bar{\mu}_\xi = \begin{cases} (\mu_{\theta+\xi} - \mu_\theta)/\xi, & \xi \neq 0; \\ \mu'_\theta, & \xi = 0. \end{cases}$$

By hypothesis, $\bar{\mu}_*$ is $[\mathcal{D}]_v$-continuous on V and Theorem 1.2 (*i*) concludes the proof. □

In general, checking strong and regular differentiability, as defined by (2.5) and (2.6), respectively, might be a very demanding task and it is desirable to have easily verifiable sufficient conditions instead. In the following, we express such sufficiency conditions by means of continuity of the weak derivative mapping μ'_*. More specifically, the following result shows that, provided that μ_* is weakly differentiable, strong (resp. regular) continuity of μ'_* at θ implies strong (resp. regular) differentiability of μ_*, at θ. The precise statement is as follows.

Theorem 2.2. *Let $\mu_* : \Theta \to \mathcal{M}_v$ be weakly $[\mathcal{D}]_v$-differentiable.*

(i) If μ'_* is strongly continuous at θ, then μ_θ is strongly differentiable.

(ii) If μ'_* is regularly continuous at θ, then μ_θ is regularly differentiable.

Proof. Applying the Mean Value Theorem to the mapping $\theta \mapsto \int g(s)\mu_\theta(ds)$ yields

$$\forall g \in [\mathcal{D}]_v : \int g(s)\mu_{\theta+\xi}(ds) - \int g(s)\mu_\theta(ds) = \xi \int g(s)\mu'_{\theta+\xi_g}(ds), \tag{2.7}$$

for some ξ_g depending on g and satisfying $0 < |\xi_g| < |\xi|$.

(i) Let $\epsilon > 0$ be arbitrary and choose $\zeta > 0$ such that

$$\forall \xi \in (-\zeta, \zeta) : \|\mu'_{\theta+\xi} - \mu'_\theta\|_v < \epsilon.$$

Hence, for all $g \in [\mathcal{D}]_v$ satisfying $\|g\|_v \leq 1$ and $\xi \in (-\zeta, \zeta)$ it holds that

$$\begin{aligned} \left| \int g(s)(\mu_{\theta+\xi} - \mu_\theta - \xi \cdot \mu'_\theta)(ds) \right| &= |\xi| \cdot \left| \int g(s)(\mu'_{\theta+\xi_g} - \mu'_\theta)(ds) \right| \\ &\leq |\xi| \cdot \|\mu'_{\theta+\xi_g} - \mu'_\theta\|_v \leq \epsilon|\xi|. \end{aligned}$$

Taking the supremmum with respect to $\|g\|_v \leq 1$ in the above inequality we conclude that

$$\|\mu_{\theta+\xi} - \mu_\theta - \xi \cdot \mu'_\theta\|_v \leq \epsilon|\xi|.$$

Since ϵ was arbitrary, dividing both sides in the above inequality by $|\xi|$ and letting $\xi \to 0$, proves the claim.

(ii) By hypothesis, the mapping $\|\mu'_*\|_v$ is continuous at θ and for arbitrary $\epsilon > 0$ choose $\zeta > 0$ such that

$$\forall \xi \in (-\zeta, \zeta) : \|\mu'_{\theta+\xi}\|_v \leq \|\mu'_\theta\|_v + \epsilon.$$

Therefore, from (2.7) we conclude that

$$\forall \xi \in (-\zeta, \zeta) : \left\| \frac{\mu_{\theta+\xi} - \mu_\theta}{\xi} \right\|_v \leq \|\mu'_\theta\|_v + \epsilon. \tag{2.8}$$

Since ϵ was arbitrary chosen, letting $\xi \to 0$ in (2.8) yields

$$\limsup_{\xi \to 0} \left\| \frac{\mu_{\theta+\xi} - \mu_\theta}{\xi} \right\|_v \leq \|\mu'_\theta\|_v.$$

Now, in accordance with (2.6), Remark 1.5 concludes the proof. □

Basic Rules of Weak Differentiation

In the following we discuss some basic rules of weak differentiation. More specifically, we are interested under which transformations weak differentiability of a measure-valued mapping is preserved. To this end, recall that if $\lambda \in \mathcal{M}$ and $f \in \mathcal{F}$ we define $f \cdot \lambda \in \mathcal{M}$ as follows:

$$\forall s \in \mathbb{S}: \ \mu(ds) := f(s)\lambda(ds).$$

Note that, if $\lambda \in \mathcal{M}_v$ and $f \in [\mathcal{F}]_v$, for some $v \in \mathcal{C}^+$, then $f \cdot \lambda$ is finite and if f is a constant function then we recover the scalar multiplication on the space of measures. The following two results are useful in applications. The first result shows that $[\mathcal{D}]_v$-differentiation acts as a linear operator.

Lemma 2.1. *If μ_θ and η_θ are $[\mathcal{D}]_v$-differentiable then any linear combination $\alpha \cdot \mu_\theta + \beta \cdot \eta_\theta$, with $\alpha, \beta \in \mathbb{R}$, is $[\mathcal{D}]_v$-differentiable and it holds that*

$$\forall \alpha, \beta \in \mathbb{R}: \ (\alpha \cdot \mu_\theta + \beta \cdot \eta_\theta)' = \alpha \cdot \mu'_\theta + \beta \cdot \eta'_\theta.$$

Proof. Basic properties of classical derivatives show that

$$\begin{aligned} \frac{d}{d\theta} \int g(s)(\alpha \cdot \mu_\theta + \beta \cdot \eta_\theta)(ds) &= \alpha \cdot \frac{d}{d\theta} \int g(s)\mu_\theta(ds) + \beta \cdot \frac{d}{d\theta} \int g(s)\eta_\theta(ds) \\ &= \alpha \int g(s)\mu'_\theta(ds) + \beta \int g(s)\eta'_\theta(ds) \\ &= \int g(s)(\alpha \cdot \mu'_\theta + \beta \cdot \eta'_\theta)(ds), \end{aligned}$$

holds true for any $g \in [\mathcal{D}]_v$. Therefore, Lemma 1.2 concludes the proof. □

Let $v, \vartheta \in \mathcal{C}^+$. The next result provides sufficient conditions such that $[\mathcal{D}]_\vartheta$-differentiability of a measure λ_θ implies $[\mathcal{D}]_v$-differentiability of the re-scaled measure $f_\theta \lambda_\theta$.

Lemma 2.2. *Let $\lambda_* : \Theta \to \mathcal{M}_\vartheta$ be a measure-valued mapping and consider a family of measurable functions h and f_θ, for $\theta \in \Theta$, such that $h \cdot v \in [\mathcal{F}]_\vartheta$ and $f_\theta \cdot v \in [\mathcal{F}]_\vartheta$, for each $\theta \in \Theta$, for some $v \in \mathcal{C}^+$. Assume further that the derivative $f'_\theta(s) := \frac{d}{d\theta} f_\theta(s)$ exists for each $s \in \mathbb{S}$ and satisfies*

$$\forall s \in \mathbb{S}: \ \sup_{\theta \in \Theta} |f'_\theta(s)| \le h(s).$$

If $\mu_\theta := f_\theta \cdot \lambda_\theta$, for $\theta \in \Theta$, then we have:

(i) If λ_θ is $[\mathcal{F}]_\vartheta$-differentiable then μ_θ is $[\mathcal{F}]_v$-differentiable and it holds that

$$\mu'_\theta = f'_\theta \cdot \lambda_\theta + f_\theta \cdot \lambda'_\theta. \tag{2.9}$$

(ii) If $f_\theta \in \mathcal{C}$ and λ_θ is $[\mathcal{C}]_\vartheta$-differentiable then μ_θ is $[\mathcal{C}]_v$-differentiable and (2.9) holds true.

(iii) *If $\lambda_\theta = \lambda$, for each $\theta \in \Theta$, then the conditions of the lemma can be relaxed to $h \cdot v \in \mathcal{L}^1(\lambda)$ and $f_\theta \cdot v \in \mathcal{L}^1(\lambda)$ and*

$$\mu'_\theta = f'_\theta \cdot \lambda.$$

Proof. *(i)* The conclusion is equivalent to

$$\frac{\mu_{\theta+\xi} - \mu_\theta}{\xi} \overset{[\mathcal{F}]_v}{\Longrightarrow} f'_\theta \cdot \lambda_\theta + f_\theta \cdot \lambda'_\theta,$$

for $\xi \to 0$. Moreover, simple algebra shows that

$$\frac{\mu_{\theta+\xi} - \mu_\theta}{\xi} = \frac{f_{\theta+\xi} - f_\theta}{\xi} \cdot \lambda_\theta + f_\theta \cdot \frac{\lambda_{\theta+\xi} - \lambda_\theta}{\xi} + (f_{\theta+\xi} - f_\theta) \cdot \frac{\lambda_{\theta+\xi} - \lambda_\theta}{\xi}. \tag{2.10}$$

We start by analyzing the first term in (2.10). According to the the Dominated Convergence Theorem,

$$\forall g \in [\mathcal{F}]_v : \lim_{\xi \to 0} \int g(s) \frac{f_{\theta+\xi}(s) - f_\theta(s)}{\xi} \lambda_\theta(ds) = \int g(s) f'_\theta(s) \lambda_\theta(ds). \tag{2.11}$$

Indeed, note that the integrand in the left-hand side satisfies

$$\forall s \in \mathbb{S} : \lim_{\xi \to 0} g(s) \frac{f_{\theta+\xi}(s) - f_\theta(s)}{\xi} = g(s) f'_\theta(s),$$

and by the Mean Value Theorem we have

$$\forall \xi : \left| g(s) \frac{f_{\theta+\xi}(s) - f_\theta(s)}{\xi} \right| \leq |g(s)| h(s) \leq \|g\|_v \, v(s) h(s).$$

We turn now to the second term in (2.10). Since $f_\theta \cdot v \in [\mathcal{F}]_\vartheta$, we conclude that

$$\forall g \in [\mathcal{F}]_v : \lim_{\xi \to 0} \int g(s) f_\theta(s) \left(\frac{\lambda_{\theta+\xi} - \lambda_\theta}{\xi} \right) (ds) = \int g(s) f_\theta(s) \lambda'_\theta(ds). \tag{2.12}$$

Finally, for arbitrary ξ and $g \in [\mathcal{F}]_v$ we have

$$\begin{aligned} \left| \int g(s) \frac{f_{\theta+\xi}(s) - f_\theta(s)}{\xi} (\lambda_{\theta+\xi} - \lambda_\theta)(ds) \right| &\leq \|g\|_v \int v(s) h(s) |\lambda_{\theta+\xi} - \lambda_\theta|(ds) \\ &\leq \|g\|_v \, \|h \cdot v\|_\vartheta \, \|\lambda_{\theta+\xi} - \lambda_\theta\|_\vartheta. \end{aligned}$$

Therefore, since $h \cdot v \in [\mathcal{F}]_\vartheta$, by letting $\xi \to 0$ in the above inequality, we conclude from Theorem 2.1 that

$$\forall g \in [\mathcal{F}]_v : \lim_{\xi \to 0} \int g(s) \frac{f_{\theta+\xi}(s) - f_\theta(s)}{\xi} (\lambda_{\theta+\xi} - \lambda_\theta)(ds) = 0, \tag{2.13}$$

which together with (2.11) and (2.12), concludes the proof.

(ii) If f_θ is continuous it follows that for any $g \in [\mathcal{C}]_v$ we have $g \cdot f_\theta \in [\mathcal{C}]_\vartheta$ and, consequently, (2.12) holds true.

(iii) The proof is similar to that of the first part and where we take into account that the expressions the left-hand side of (2.12) and (2.13) vanish. □

Remark 2.2. *The statement in Lemma 2.2 admits several variations which would make the result stronger. For instance, the condition "the derivative $f'_\theta(s)$ exists for each $s \in \mathbb{S}$" can be replaced by "both the right and the left-sided derivatives exist for each $s \in \mathbb{S}$ and the derivative $f'_\theta(s)$ exists μ_θ-a.e."*

Higher-order Differentiation

We conclude this section by discussing higher-order differentiation. To this end, note that (2.3) in Definition 2.1 is equivalent to

$$\forall g \in [\mathcal{D}]_v : \frac{d}{d\theta} \int g(s) \mu_\theta(ds) = \int g(s) \mu'_\theta(ds),$$

i.e., one can interchange integration with differentiation. In the same vein one can introduce higher-order differentiation. More specifically, for $n \geq 1$ we say that μ_θ in n-times weakly differentiable if there exist $\mu_\theta^{(n)} \in \mathcal{M}_v$ such that

$$\forall g \in [\mathcal{D}]_v : \frac{d^n}{d\theta^n} \int g(s) \mu_\theta(ds) = \int g(s) \mu_\theta^{(n)}(ds). \tag{2.14}$$

Remark 2.3. *Note that, just like in conventional analysis, higher-order derivatives satisfy*

$$\forall 0 \leq j \leq n-1 : \left(\mu_\theta^{(j)}\right)' = \mu_\theta^{(j+1)},$$

provided that μ_θ is n times weakly differentiable. Indeed, since μ_θ is $(j+1)$ times weakly differentiable it follows that

$$\begin{aligned} \forall g \in [\mathcal{D}]_v : \int g(s) \mu_\theta^{(j+1)}(ds) &= \frac{d^{j+1}}{d\theta^{j+1}} \int g(s) \mu_\theta(ds) \\ &= \frac{d}{d\theta} \left(\frac{d^j}{d\theta^j} \int g(s) \mu_\theta(ds) \right) \\ &= \frac{d}{d\theta} \int g(s) \mu_\theta^{(j)}(ds). \end{aligned}$$

Therefore, the measures $\left(\mu_\theta^{(j)}\right)'$ and $\mu_\theta^{(j+1)}$ agree when considering integrands from $[\mathcal{D}]_v$ and by Lemma 1.2 we conclude that they are equal in the sense of Remark 1.2.

2.2.2 Representation of the Weak Derivatives

In general, weak derivatives are abstract objects (that is, signed measures). For instance, if $\mu_* : \Theta \to \mathcal{M}^1$, i.e., μ_θ is a probability measure for each $\theta \in \Theta$ then there exists some (abstract) measurable space $(\Omega, \mathcal{K})$ and some measurable mapping (random variable) $X : \Omega \to \mathbb{S}$ such that for all $\theta \in \Theta$ we have

$$\forall g \in [\mathcal{D}]_v : \int g(s) \mu_\theta(ds) = \int g(X(\omega)) \mathbb{P}_\theta(d\omega),$$

where $\mathbb{P}_\theta$ is a probability measure on $(\Omega, \mathcal{K})$ satisfying

$$\forall A \in \mathcal{S} : \mathbb{P}_\theta(X \in A) = \mu_\theta(A), \tag{2.15}$$

i.e., X is a random variable distributed according to μ_θ. It follows that for each $\theta \in \Theta$ we have the following representation

$$\forall g \in [\mathcal{D}]_v : \int g(s) \mu_\theta(ds) = \mathbb{E}_\theta[g(X)], \tag{2.16}$$

where $\mathbb{E}_\theta$ denotes the expectation operator on the probability field $(\Omega, \mathcal{K}, \mathbb{P}_\theta)$. Moreover, the representation in (2.16) is valid whenever (2.15) holds true. Inspired by the above remarks, we give the following definition:

Definition 2.2. *Let $\mu_{i,*} : \Theta \to \mathcal{M}^1(\mathbb{S}_i)$, for $i \in I$, be an arbitrary family of measure-valued mappings. We say that $\mathbb{E}_\theta$ is an expectation operator consistent with $X_i \sim \mu_{i,\theta}$, for each $i \in I$, if there exists some probability field $(\Omega, \mathcal{K})$, on which random variables X_i are defined, for each $i \in I$, and there exists*[2] *a family of probability measures $\{\mathbb{P}_\theta : \theta \in \Theta\}$ on $(\Omega, \mathcal{K})$ satisfying*

$$\forall \theta \in \Theta, i \in I, A \in \mathcal{S} : \mathbb{P}_\theta(X_i \in A) = \mu_{i,\theta}(A)$$

and for each $\theta \in \Theta$, $\mathbb{E}_\theta$ coincides with the expectation operator on $(\Omega, \mathcal{K}, \mathbb{P}_\theta)$.

Therefore, weak differentiability of μ_θ provides the means of evaluating the derivatives of the expression $\mathbb{E}_\theta[g(X)]$, for $g \in [\mathcal{D}]_v$, provided that $\mathbb{E}_\theta$ is an expectation operator consistent with $X \sim \mu_\theta$. Note that the derivative of the right-hand side in (2.16) satisfies

$$\forall g \in [\mathcal{D}]_v : \int g(s) \mu_\theta^{(n)}(ds) = \frac{d^n}{d\theta^n} \mathbb{E}_\theta[g(X)]$$

but does not admit a representation as in (2.16) since $\mu_\theta^{(n)}$ fails to be a probability measure. Fortunately, if $\mu_\theta^{(n)}$ is a finite measure, a convenient representation for higher-order derivatives of probability measures in terms of random variables is possible via the Hahn-Jordan decomposition. This is useful in applications as it provides unbiased gradient estimators for $\mathbb{E}_\theta[g(X)]$.

For technical reasons we distinguish between the case

$$\inf\{v(s) : s \in \mathbb{S}\} > 0$$

which we will call *the standard case* and the case

$$\inf\{v(s) : s \in \mathbb{S}\} = 0$$

which will be referred to as *the non-standard case.*

The Standard Case

Note that, if $(\mathcal{D}, v)$ is a Banach base and $\inf\{v(s) : s \in \mathbb{S}\} > 0$ it holds that

$$\mathcal{C}_B \subset [\mathcal{C}]_v \subset [\mathcal{D}]_v.$$

For fixed $n \geq 1$, letting $g = \mathbb{I}_\mathbb{S}$ in (2.14) yields $\mu_\theta^{(n)}(\mathbb{S}) = 0$. Let

$$\mu_\theta^{(n)} = \left[\mu_\theta^{(n)}\right]^+ - \left[\mu_\theta^{(n)}\right]^-$$

be the Hahn-Jordan decomposition of $\mu_\theta^{(n)}$. It follows that

$$\left[\mu_\theta^{(n)}\right]^+ (\mathbb{S}) = \left[\mu_\theta^{(n)}\right]^- (\mathbb{S}), \tag{2.17}$$

[2] It can be shown that such objects always exist!

provided that $\mu_\theta^{(n)}$ is a finite measure. Denoting by $c_\theta^{(n)}$ the common value in (2.17) one can represent the n^{th}-order derivative $\mu_\theta^{(n)}$ as follows

$$\mu_\theta^{(n)} = c_\theta^{(n)} \left(\mu_\theta^{(n+)} - \mu_\theta^{(n-)} \right), \tag{2.18}$$

where $c_\theta^{(n)} > 0$ (if the n^{th} derivative is significant) and $\mu_\theta^{(n\pm)} \in \mathcal{M}^1$. Therefore, provided that $\mathbb{E}_\theta$ is an expectation operator consistent with $X \sim \mu_\theta$ and $X^{(n\pm)} \sim \mu_\theta^{(n\pm)}$, for $n \geq 1$, respectively, we have

$$\forall n \geq 1: \ \frac{d^n}{d\theta^n} \mathbb{E}_\theta[g(X)] = c_\theta^{(n)} \mathbb{E}_\theta \left[g\left(X^{(n+)}\right) - g\left(X^{(n-)}\right) \right]. \tag{2.19}$$

Note that a representation as in (2.18) is not unique. However, the representation provided by the Hahn-Jordan decomposition has the property that it minimizes the constant $c_\theta^{(n)}$ and we call it the *orthogonal representation.*

Therefore, one can identify the weak derivative $\mu_\theta^{(n)}$ with any triple

$$\left(c_\theta^{(n)}, \mu_\theta^{(n+)}, \mu_\theta^{(n-)} \right) \in \mathbb{R} \times \mathcal{M}^1 \times \mathcal{M}^1,$$

satisfying equation (2.18). This fact will be exploited in the following. For ease of writing, for $n = 1$, i.e., $\mu_\theta^{(n)} = \mu_\theta'$, we use the simplified notation

$$\left(c_\theta, \mu_\theta^+, \mu_\theta^- \right).$$

The Non-Standard Case

In the non-standard case, we drop the assumption $\inf\{v(s) : s \in \mathbb{S}\} > 0$, so we allow v to take very small values (close to, or even 0). However, we may assume without loss of generality that

$$\forall \theta \in \Theta: \ \mu_\theta(\mathbb{S} \setminus \mathbb{S}_v) = 0,$$

since within our theory we consider the trace of μ_θ on $\mathbb{S}_v$. Unfortunately, in this case $\mathbb{I}_\mathbb{S} \notin [\mathcal{D}]_v$ and a representation as in (2.18) can not be obtained in a straightforward way.

Example 2.1. *Let $v(s) = 1/s$, for $s > 0$, and consider the family*

$$\forall \theta \in [0,1]: \ \mu_\theta := \begin{cases} (1-\theta) \cdot \mu + \theta \cdot \delta_{1/\theta}, & \theta \in (0,1], \\ \mu, & \theta = 0, \end{cases}$$

for some $\mu \in \mathcal{M}_v^1$. Then μ_* is weakly $\mathcal{C}_B$-continuous at $\theta = 0$ but fails to be $\mathcal{C}_B$-differentiable since the family

$$\left\{ \frac{\mu_\xi - \mu_0}{\xi} : \xi > 0 \right\} = \left\{ \delta_{1/\xi} - \mu : \xi > 0 \right\}$$

is not tight and, consequently, the limit $\lim_{\xi \to 0}(\delta_{1/\xi} - \mu)$ does not exist in $\mathcal{M}_B$; see Example 1.8. However, it turns out that μ_* is $\mathcal{C}_v$-differentiable at $\theta = 0$ since

$$\forall g \in \mathcal{C}_v: \ \lim_{\xi \downarrow 0} \frac{1}{\xi} \left(\int g(s) \mu_\xi(ds) - \int g(s) \mu_0(ds) \right) = \lim_{\xi \downarrow 0} g\left(\frac{1}{\xi}\right) - \int g(s) \mu(ds),$$

which yields $\mu_0' = -\mu$. Therefore, a representation as in (2.18) is not possible for μ_0'. In addition, note that μ_* is strongly $\mathcal{C}_v$-differentiable. Indeed, we have

$$\lim_{\xi\downarrow 0} \left\| \frac{\mu_\xi - \mu_0}{\xi} - \mu_0' \right\|_v = \lim_{\xi\downarrow 0} \|\delta_{1/\xi} - \mu + \mu\|_v = \lim_{\xi\downarrow 0} \xi = 0.$$

Note that a representation as in (2.18) holds true whenever μ_θ is $\mathcal{C}_B(\mathbb{S}_v)$-differentiable. The following result shows that the representation in (2.18) is still possible, under slightly less restrictive conditions.

Lemma 2.3. *Let $\mu_* : \Theta \to \mathcal{M}_v$ be $[\mathcal{D}]_v$-differentiable at θ, such that $\mu_\theta(\mathbb{S}_v)$ is constant with respect to θ. If there exists a neighborhood V of 0 such that the family*

$$\left\{ \frac{\mu_{\theta+\xi} - \mu_\theta}{\xi} \,:\, \xi \in V \setminus \{0\} \right\}$$

is tight then it holds that $\mu_\theta'(\mathbb{S}_v) = 0$.

Proof. Let us define the sequence $f_n : V \setminus \{0\} \to \mathbb{R}$, for $n \geq 1$, as follows:

$$\forall n \geq 1, \xi \in V \setminus \{0\} :\ f_n(\xi) := \int \min\{1, n \cdot v(s)\} \left(\frac{\mu_{\theta+\xi} - \mu_\theta}{\xi} \right)(ds).$$

Formally, our statement is equivalent to

$$\mu_\theta'(\mathbb{S}_v) = \lim_{n\to\infty} \lim_{\xi\to 0} f_n(\xi) = \lim_{\xi\to 0} \lim_{n\to\infty} f_n(\xi) = 0. \tag{2.20}$$

In the following we show that the sequence $\{f_n\}_n$ satisfies the conditions of Theorem B.2 (see the Appendix) to prove that interchanging limit operations in (2.20) is justified.

First, note that $[\mathcal{D}]_v$-differentiability of μ_θ implies that

$$\forall n \geq 1 :\ \lim_{\xi\to 0} f_n(\xi) = L_n := \int \min\{1, n \cdot v(s)\}\, \mu_\theta'(ds), \tag{2.21}$$

since, for $n \geq 1$, the mapping $s \mapsto \min\{1, n \cdot v(s)\}$ is continuous and has finite v-norm; hence, belongs, by assumption, to $[\mathcal{D}]_v$.

On the other hand, we have

$$\forall \xi \in V \setminus \{0\} :\ \lim_{n\to\infty} f_n(\xi) = \left(\frac{\mu_{\theta+\xi} - \mu_\theta}{\xi} \right)(\mathbb{S}_v) = 0.$$

Moreover, by hypothesis, for each $\epsilon > 0$ there exists some compact set $K_\epsilon \subset \mathbb{S}_v$ such that

$$\forall \xi \in V \setminus \{0\} :\ \left| \frac{\mu_{\theta+\xi} - \mu_\theta}{\xi} \right| (\mathbb{S}_v \setminus K_\epsilon) < \epsilon. \tag{2.22}$$

Since v is continuous it follows that $1/v$ is bounded on K_ϵ, i.e.,

$$M := \sup_{s\in K_\epsilon} \frac{1}{v(s)} < \infty.$$

Choosing now some $n_\epsilon \geq M$ it follows that the following inclusion holds true:

$$\{s : n_\epsilon \cdot v(s) < 1\} \subset \{s : M \cdot v(s) < 1\} \subset \mathbb{S}_v \setminus K_\epsilon.$$

Therefore, for each $n \geq n_\epsilon$ and $\xi \in V \setminus \{0\}$ it holds that

$$\begin{aligned} |f_n(\xi)| &\leq \int |1 - \min\{1, n \cdot v(s)\}| \left| \frac{\mu_{\theta+\xi} - \mu_\theta}{\xi} \right| (ds) \\ &\leq \int \mathbb{I}_{\{s:\, n_\epsilon \cdot v(s) < 1\}}(s) \left| \frac{\mu_{\theta+\xi} - \mu_\theta}{\xi} \right| (ds) \\ &= \left| \frac{\mu_{\theta+\xi} - \mu_\theta}{\xi} \right| (\{s :\ n_\epsilon \cdot v(s) < 1\}) \leq \left| \frac{\mu_{\theta+\xi} - \mu_\theta}{\xi} \right| (\mathbb{S}_v \setminus K_\epsilon) \end{aligned}$$

and by (2.22) we conclude that the sequence $\{f_n\}_n$ converges to 0, uniformly with respect to $\xi \in V \setminus \{0\}$, i.e., for each $\epsilon > 0$ there exists $n_\epsilon \geq 1$ such that

$$\forall n \geq n_\epsilon, \xi \in V \setminus \{0\} :\ |f_n(\xi)| < \epsilon.$$

Applying now Theorem C.2 to the sequence $\{f_n\}_n$ concludes the proof. □

Note that, if in Lemma 2.3 μ_* is regularly $[\mathcal{D}]_v$-differentiable at θ then the conclusion is immediate. Indeed, by part *(ii)* of Theorem 1.1 we conclude that μ_* is regularly $\mathcal{C}_B(\mathbb{S}_v)$-differentiable at θ. The following representation result for the weak derivatives in the non-standard case is a consequence of Lemma 2.3.

Corollary 2.1. *Let $\mu_* : \Theta \to \mathcal{M}_v^1$ be n times $[\mathcal{D}]_v$-differentiable at θ, for some $n \geq 1$ and let k be such that $1 \leq k \leq n$. If there exists a neighborhood V_k of 0 such that the family*

$$\left\{ \frac{\mu_{\theta+\xi}^{(k-1)} - \mu_\theta^{(k-1)}}{\xi} \ :\ \xi \in V_k \setminus \{0\} \right\}$$

is tight then the k^{th}-order derivative $\mu_\theta^{(k)}$ admits a representation as in (2.18).

Proof. For $n = 1$ the proof follows from Lemma 2.3 by taking into account that $\mu_\theta(\mathbb{S}_v) = 1$, for each $\theta \in \Theta$. Indeed, it follows that μ'_θ is a finite measure such that $\mu'_\theta(\mathbb{S}_v) = 0$ and, consequently, admits a representation as in (2.18).

By finite induction, for $n \geq 2$, one can apply (again) Lemma 2.3 to $\mu_*^{(k-1)}$ which, by Remark 2.3, is $[\mathcal{D}]_v$-differentiable at θ and satisfies $\mu_\theta^{(k-1)}(\mathbb{S}_v) = 0$, for each θ for which the derivative exists. □

Therefore, we conclude that the "triple" representation of the weak derivatives in the non-standard case is still possible, under some additional conditions.

2.2.3 Computation of Weak Derivatives and Examples

We start with the following remark.

Remark 2.4. *It is worth noting that, in principle, weak derivatives can be computed in a straightforward way if it holds that*

$$\forall \theta \in \Theta : \; \mu_\theta(ds) = f_\theta(s) \cdot \lambda(ds),$$

i.e., if μ_θ has a density f_θ with respect to some $\lambda \in \mathcal{M}$. Indeed, by part (ii) of Lemma 2.2 we have

$$\forall g \in [\mathcal{D}]_v : \; \frac{d^n}{d\theta^n} \int g(s) f_\theta(s) \lambda(ds) = \int g(s) \frac{d^n}{d\theta^n} f_\theta(s) \lambda(ds), \tag{2.23}$$

provided that $f_\theta(s)$ is n-times differentiable at θ, for all $s \in S$, and interchanging differentiation and integral is justified. Hence, we have

$$\mu_\theta^{(n)}(ds) = \frac{d^n}{d\theta^n} f_\theta(s) \cdot \lambda(ds),$$

and a weak derivative can be easily computed by considering the positive and the negative parts of $\frac{d^n}{d\theta^n} f_\theta(s)$, i.e.,

$$\left[\mu_\theta^{(n)}\right]^+ (ds) = \left(\frac{d^n}{d\theta^n} f_\theta(s) \right)_+ \lambda(ds), \quad \left[\mu_\theta^{(n)}\right]^- (ds) = \left(\frac{d^n}{d\theta^n} f_\theta(s) \right)_- \lambda(ds),$$

where, for $a \in \mathbb{R}$, we set $a_+ := \max\{a, 0\}$ and $a_- := \max\{-a, 0\} = a_+ - a$.

We illustrate the concept of weak differentiation with a few families of measures that are of importance in applications. More examples can be found in Section H of the Appendix. For ease of exposition we agree on the following notations to be used throughout this thesis: Let ℓ denote the Lebesgue measure on $\mathbb{S} = \mathbb{R}^n$, for some $n \geq 1$, and for arbitrary $A \in \mathcal{S}$ we denote by $\mathcal{U}_A$ the uniform distribution on A, i.e.,

$$\forall x \in \mathbb{S} : \; \mathcal{U}_A(dx) := \frac{1}{\ell(A)} \mathbb{I}_A(x) dx.$$

Example 2.2. *Let $\mu \in \mathcal{M}_v$. If $\mu_\theta = \mu$, for all $\theta \in \Theta$, then μ_θ is obviously weakly $[\mathcal{F}]_v$-differentiable since*

$$\forall g \in [\mathcal{F}]_v : \; \frac{d}{d\theta} \int g(s) \mu_\theta(ds) = 0.$$

In this case the weak derivative is not significant and we set $\mu'_\theta = (1, \mu, \mu)$.

Example 2.3. *The Dirac distribution δ_θ, for $\theta \in [a, b] \subset \mathbb{R}$, fails to be weakly $[\mathcal{D}]_v$-differentiable for any sensible set $\mathcal{D}$. Indeed, $\int g(x) \delta_\theta(dx) = g(\theta)$ is differentiable at θ only if g is differentiable at θ. This however would impose quite strong restrictions on the performance measures to be analyzed.*

Nevertheless, the mapping δ_* is weakly $[\mathcal{C}]_v$-continuous for any $v \in \mathcal{C}^+$ and it is strongly continuous at θ only if $v(\theta) = 0$. Therefore, the Dirac distribution δ_θ is a standard example of distribution which is weakly continuous everywhere but nowhere weakly differentiable.

Example 2.4. *Let* $\mathbb{S} = \{x_1, x_2\}$*, with the discrete topology and for* $\theta \in [0,1]$ *let us consider*

$$\beta_\theta = (1-\theta)\cdot\delta_{x_1} + \theta\cdot\delta_{x_2},$$

i.e., the Bernoulli distribution with mass points $\{x_1, x_2\}$ *and probability weights* $\{1-\theta, \theta\}$*, respectively. To avoid trivialities we assume* $x_1 \neq x_2$*. Then it holds that*

$$\forall g \in \mathcal{F}: \ \frac{d}{d\theta}\int g(x)\beta_\theta(dx) = \frac{d}{d\theta}\Big((1-\theta)g(x_1) + \theta g(x_2)\Big) = g(x_2) - g(x_1)\,.$$

This means that β_θ *is weakly* $[\mathcal{F}]_v$*-differentiable, for any* $v \in \mathcal{C}^+$ *and*

$$\beta'_\theta = \delta_{x_2} - \delta_{x_1},$$

so that the weak derivative can be represented as $\beta'_\theta = (1, \delta_{x_2}, \delta_{x_1})$*. In addition, by Theorem 2.2 (i) it follows that* β_θ *is strongly differentiable.*

Furthermore, as it can be easily seen, higher-order derivatives exist but are not significant in this situation and we set $\beta_\theta^{(n)} = (1, \beta_\theta, \beta_\theta)$*, for* $n \geq 2$*.*

Example 2.5. *Let* $\mathbb{S} = [0,\infty)$ *with the usual topology,* $\Theta = (a,b)$*, for* $0 < a < b < \infty$*, and choose* $\mu_\theta(dx) = \theta\exp(-\theta x)\cdot\ell(dx)$*, i.e.,* μ_θ *denotes the exponential distribution with rate* θ*. Moreover, if* $v_p(x) = 1 + x^p$*, for some* $p \geq 0$*, then* μ_θ *is weakly* $[\mathcal{F}]_{v_p}$*-differentiable and its derivative satisfies*

$$\mu'_\theta(dx) = (1-\theta x)\exp(-\theta x)\ell(dx).$$

In addition, μ_θ *is n-times* $[\mathcal{F}]_{v_p}$*-differentiable, for all* $n \geq 1$*, and higher-order derivatives can be computed in the same way, by differentiating the density*

$$f_\theta(x) = \theta\exp(-\theta x)$$

in the classical sense. Consequently, for each $n \geq 1$ *we obtain*

$$\mu_\theta^{(n)}(dx) = (-1)^n x^{n-1}(\theta x - n)\exp(-\theta x)\ell(dx)$$

and an orthogonal representation can be obtained as explained in Remark 2.4. To see that, we show that the conditions of Lemma 2.2 are fulfilled. Indeed, note that $f_\theta \cdot v_p \in \mathcal{L}^1(\ell)$*, for each* $\theta \in (a,b)$ *and* $p \geq 0$*, and for* $n \geq 0$ *we have*

$$\forall\theta \in (a,b), x \geq 0: \ |x^{n-1}(\theta x - n)\exp(-\theta x)| \leq x^{n-1}(\theta x + n)\exp(-\theta x).$$

Therefore, if for $n \geq 0$ *we set*

$$\forall x \geq 0: \ h_n(x) := x^{n-1}(bx + n)\exp(-ax)$$

it follows that for each $n \geq 0$ *we have*

$$\forall x \geq 0: \ \sup_{\theta\in\Theta}\left|\frac{d^n}{d\theta^n}f_\theta(x)\right| \leq h_n(x)$$

and $h_n \cdot v_p \in \mathcal{L}^1(\ell)$, for each $p \geq 0$, and part (ii) of Lemma 2.2 concludes the proof. Furthermore, one can easily check that $\mu_*^{(n+1)}$ is strongly continuous on Θ, for $n \geq 1$, and it follows by Theorem 2.2 that μ_θ is n times strongly (in particular, regularly) differentiable, for each $n \geq 1$.

Finally, if we denote by $\varepsilon_{n,\theta}$ the Erlang distribution with parameters n, θ , i.e., the convolution[3] of n exponential distributions with rate θ, then we have

$$\mu_\theta^{(n)} = \begin{cases} \left(\frac{n!}{\theta^n}, \varepsilon_{n,\theta}, \varepsilon_{n+1,\theta}\right), & \text{if } n \text{ is odd,} \\ \left(\frac{n!}{\theta^n}, \varepsilon_{n+1,\theta}, \varepsilon_{n,\theta}\right), & \text{if } n \text{ is even} \end{cases}, n \geq 1,$$

which yields another representation for the higher-order derivatives of μ_θ, which is more convenient for applications.

Example 2.6. *Let $\mathbb{S} = [0, \infty)$, and denote by ψ_θ the uniform distribution on the interval $[0, \theta)$, i.e., $\psi_\theta = \mathcal{U}_{[0,\theta)}$, for $\theta \in (0, b)$, with $b > 0$. Note that, one can extend the measure-valued mapping ψ_* in 0, by setting $\psi_0 = \delta_0$. It turns out that ψ_* is weakly continuous at 0 and it is strongly continuous at 0 only if $v(0) = 0$. Therefore, by Theorem 2.1 we conclude that, in general, ψ_* is not weakly differentiable at $\theta = 0$.*

Take $\mathcal{D}$ as the set $\mathcal{C}(\mathbb{S})$. Since the density $\theta^{-1}\mathbb{I}_{[0,\theta)}(x)$ is not differentiable (not even continuous) with respect to θ, Lemma 2.2 does not apply in this situation and we calculate the weak derivative ψ'_θ, for $\theta > 0$, by definition. For each g continuous at θ, we have

$$\int g(s)\psi'_\theta(ds) = \lim_{\xi \to 0} \frac{1}{\xi} \left(\frac{1}{\theta + \xi} \int_0^{\theta+\xi} g(s)ds - \frac{1}{\theta} \int_0^\theta g(s)ds \right),$$

which yields

$$\forall g \in \mathcal{C} : \int g(s)\psi'_\theta(ds) = \frac{1}{\theta} g(\theta) - \frac{1}{\theta^2} \int_0^\theta g(s)ds.$$

Hence, ψ_θ is weakly $[\mathcal{C}]_v$-differentiable, for any $v \in \mathcal{C}^+$ and

$$\psi'_\theta = \frac{1}{\theta}\delta_\theta - \frac{1}{\theta}\psi_\theta,$$

or in triplet representation $\psi'_\theta = (\theta^{-1}, \delta_\theta, \psi_\theta)$.

It follows from Theorem 2.2 and Example 2.3 that ψ_θ is regularly differentiable and it is strongly differentiable at θ only if $v(\theta) = 0$. Indeed, one can check that

$$\lim_{\xi \to 0} \left\| \frac{\psi_{\theta+\xi} - \psi_\theta}{\xi} - \psi'_\theta \right\|_v = 2v(\theta).$$

Higher-order derivatives of ψ_θ do not exist. This stems form the fact that the Dirac measure δ_θ fails to be weakly differentiable; see Example 2.3.

The following example is rather technical and is intended to show that, in general, weak differentiability does not imply regular differentiability.

[3] Note that $\varepsilon_{1,\theta} = \mu_\theta$.

Example 2.7. *Let ψ_θ denote the uniform distribution on $[0,\theta)$, introduced in Example 2.6 and consider the following family of distributions*

$$\forall \theta \in [0,1]: \ \phi_\theta = \psi_1 + \theta \cdot (\delta_0 - \psi_\theta),$$

where, by convention, $\psi_0 = \delta_0$. Note that, for each $\theta \in [0,1]$, ϕ_θ is a probability measure and by Lemma 2.2 ϕ_θ is weakly $[\mathcal{C}]_v$-differentiable, for any $v \in \mathcal{C}^+$. Indeed, we have

$$\forall \theta > 0: \ \phi'_\theta = (\delta_0 - \psi_\theta) - \theta \cdot \psi'_\theta = (\delta_0 - \psi_\theta) - (\delta_\theta - \psi_\theta) = \delta_0 - \delta_\theta.$$

Furthermore, ϕ_θ has a right-hand side weak derivative at $\theta = 0$, which equals the null measure $\varnothing$, but fails to be regularly differentiable since

$$\lim_{\xi \downarrow 0} \left\| \frac{\phi_\xi - \phi_0}{\xi} \right\|_v = \lim_{\xi \downarrow 0} \|\delta_0 - \psi_\xi\|_v = 2v(0) \neq 0 = \|\varnothing\|_v,$$

provided that $v(0) > 0$.

Truncated Distributions

We conclude this section by treating a special class of weakly differentiable distributions. Truncated distributions play an important role in our analysis as they are typical examples of weakly, but not strongly, differentiable distributions. In particular, it will turn out that the uniform distribution presented in Example 2.6 belongs to this class.

Let X be a real-valued random variable and let $-\infty \leq a < b \leq \infty$ be such that $\mathbb{P}(\{a < X < b\}) > 0$. By a *truncation* $\mu|_{(a,b)}$ of the distribution μ of X we mean the conditional distribution of X on the event $\{a < X < b\}$. In formula:

$$\forall A: \ \mu|_{(a,b)}(A) := \frac{\mu(A \cap (a,b))}{\mu((a,b))} = \frac{\mathbb{P}(A \cap \{a < X < b\})}{\mathbb{P}(\{a < X < b\})}.$$

If X (resp. μ) has a probability density ρ, then the mapping

$$\forall x \in \mathbb{R}: \ f(x) := \frac{\rho(x)}{\int_a^b \rho(s)ds} \cdot \mathbb{I}_{(a,b)}(x) \tag{2.24}$$

is the probability density of the truncated distribution $\mu|_{(a,b)}$.

Truncated distributions arise naturally in applications. Indeed, consider a constant $a > 0$ modeling a traveling time in a transportation network. It is quite common to add a normally distributed noise, say Z, to a in order to model some intrinsic randomness; see [30]. Since, for practical reasons, it is important to ensure that $\mathbb{P}(a + Z < 0) = 0$ (so that traveling times stay larger than zero), one considers a truncated version of Z. In other words, the distribution of $a + Z$ is conditioned on the event $\{a + Z > \theta\}$ for $\theta > 0$ small.

Note that f as defined by (2.24) is still a probability density if we only require that ρ is a non-negative integrable function on (a,b), i.e., μ is a locally finite measure and not necessarily a probability measure on $\mathbb{R}$. For instance, in some models one can observe that some random variable takes values within some given interval, but its distribution density is proportional to a certain function which is not integrable on that interval. Therefore, one can obtain a truncated distribution out of any locally finite measure by an appropriate re-scaling (see, e.g., Pareto, uniform), as the following example illustrates.

Example 2.8. *In the following we provide several examples.*

(i) Letting $\rho(x) = x$, $a = 0$ *and* $b < \infty$ *in (2.24) one recovers the uniform distribution on* $(0, b)$, *cf. Example 2.6.*

(ii) Letting $\rho(x) = x^{-(\beta+1)}$, *for some* $\beta > 0$, $a > 0$ *and* $b = \infty$ *in (2.24) one obtains the Pareto distribution with density*

$$f(x) = \beta a^{\beta} x^{-(\beta+1)} \mathbb{I}_{(a,\infty)}(x).$$

(iii) For $\rho(x) = e^{-\lambda x}$, *for some* $\lambda > 0$, *and* $b = \infty$ *one obtains the shifted exponential distribution with density*

$$f(x) = \lambda e^{-\lambda(x-a)} \mathbb{I}_{(a,\infty)}(x).$$

In the setting of this section, the truncated density (2.24) is considered with $a = \theta$ and $a < b \leq \infty$; more formally, a parametric family of left-side truncated distributions $\mu|_{(\theta,b)}$ is introduced with density given by

$$f_\theta(x) = \frac{\rho(x)}{\int_\theta^b \rho(x)dx} \mathbb{I}_{(\theta,b)}(x). \tag{2.25}$$

The remainder of this section is devoted to computation of the weak derivative of a left-side truncated distribution $\mu|_{(\theta,b)}$ generated by a density ρ, i.e., $\mu|_{(\theta,b)}$ has a Lebesgue density f_θ given by (2.25). In words, we are interested in the sensitivity of $\mu|_{(\theta,b)}$ with respect to the point of truncation θ. To this end, let $v \in \mathcal{C}^+(\mathbb{R})$ be such that $\int v(x)\rho(x)dx < \infty$, i.e., $v \in \mathcal{L}^1(\mu|_{(\theta,b)})$, for any θ. Using standard computations we obtain

$$\begin{aligned}
\forall g \in [\mathcal{C}]_v : \ \frac{d}{d\theta} \frac{\int_\theta^\infty g(x)\rho(x)dx}{\int_\theta^\infty \rho(x)dx} &= \frac{\rho(\theta) \int_\theta^\infty g(x)\rho(x)dx}{\left(\int_\theta^\infty \rho(x)dx\right)^2} - \frac{g(\theta)\rho(\theta)}{\int_\theta^\infty \rho(x)dx} \\
&= \frac{\rho(\theta)}{\int_\theta^\infty \rho(x)dx} \left(\int g(x)\mu_\theta(dx) - \int g(x)\delta_\theta(dx) \right),
\end{aligned}$$

provided that ρ is continuous at θ. Hence, one can represent the derivative as follows:

$$(\mu|_{(\theta,b)})' = (c_\theta, \mu_{(\theta,b)}, \delta_\theta), \ c_\theta = \frac{\rho(\theta)}{\mu((\theta,b))}.$$

We conclude that a left-side truncated distribution $\mu|_{(\theta,b)}$ generated by a density ρ is weakly $\mathcal{C}_v$-differentiable, for $v \in \mathcal{L}^1(\mu)$, provided that ρ is continuous at θ, and its weak derivative can be represented as the re-scaled difference between the original truncated distribution $\mu|_{(\theta,b)}$ and the Dirac distribution δ_θ which assigns total mass to the point of truncation. Therefore, by Theorem 2.2 (*ii*) this implies that $\mu|_{(\theta,b)}$ is regularly differentiable, since $c.$ is continuous at θ and both $\mu|_{(\theta,b)}$ and δ_θ are weakly continuous. Moreover, a similar argument as in Example 2.6 shows that $\mu|_{(\theta,b)}$ is, in general, not strongly differentiable and higher-order derivatives do not exist.

A similar result holds true for right-side truncated distributions $\mu|_{(a,\theta)}$. Precisely, if μ has a density ρ then $\mu|_{(a,\theta)}$ is weakly $[\mathcal{C}]_v$-differentiable, for $v \in \mathcal{L}^1(\mu)$, provided that ρ is continuous at θ, and its weak derivative can be represented as follows; see Example 2.6:

$$(\mu|_{(a,\theta)})' = (c_\theta, \delta_\theta, \mu_{(a,\theta)}), \ c_\theta = \frac{\rho(\theta)}{\mu((a,\theta))}.$$

2.3 Differentiability of Product Measures

In this section we will establish sufficient conditions for weak differentiability of product measures. As it will turn out, the product of weakly differentiable measures is again weakly differentiable provided that the functional spaces are Banach bases. The main result is the following theorem.

Theorem 2.3. *Let $(\mathcal{D}(\mathbb{S}), v)$ and $(\mathcal{D}(\mathbb{T}), u)$ be Banach bases on $\mathbb{S}$ and $\mathbb{T}$, respectively. Let $\mu_\theta \in \mathcal{M}_v(\mathcal{S})$ be $[\mathcal{D}(\mathbb{S})]_v$-differentiable, $\eta_\theta \in \mathcal{M}_u(\mathcal{T})$ be $[\mathcal{D}(\mathbb{T})]_u$-differentiable. Then, the product measure $\mu_\theta \times \nu_\theta$ is $[\mathcal{D}(\mathbb{S}) \otimes \mathcal{D}(\mathbb{T})]_{v\otimes u}$-differentiable, and it holds that*

$$(\mu_\theta \times \eta_\theta)' = (\mu_\theta' \times \eta_\theta) + (\mu_\theta \times \eta_\theta').$$

Proof. For ξ such that $\theta + \xi \in \Theta$, set

$$\bar{\mu}_\xi = \frac{\mu_{\theta+\xi} - \mu_\theta}{\xi} - \mu_\theta'; \quad \bar{\eta}_\xi = \frac{\eta_{\theta+\xi} - \eta_\theta}{\xi} - \eta_\theta'.$$

By hypothesis, $\bar{\mu}_\xi \overset{[\mathcal{D}]_v}{\Longrightarrow} \varnothing$ and $\bar{\eta}_\xi \overset{[\mathcal{D}]_u}{\Longrightarrow} \varnothing$, for $\xi \to 0$, where $\varnothing$ denotes the null measure. Simple algebra shows that the proof of the claim follows from

$$\xi \cdot (\bar{\mu}_\xi + \mu_\theta') \times (\bar{\eta}_\xi + \eta_\theta') + \mu_\theta \times \bar{\eta}_\xi + \bar{\mu}_\xi \times \eta_\theta \overset{[\mathcal{D}]_{v\otimes u}}{\Longrightarrow} \varnothing, \tag{2.26}$$

for $\xi \to 0$. Hence, to conclude the proof, we show that each term on the left side of (2.26) converges weakly to null measure $\varnothing$.

Since $\bar{\mu}_\xi + \mu_\theta' \overset{[\mathcal{D}]_v}{\Longrightarrow} \mu_\theta'$ and $\bar{\eta}_\xi + \eta_\theta' \overset{[\mathcal{D}]_u}{\Longrightarrow} \eta_\theta'$, applying Theorem 1.2 yields

$$\sup_{\xi \in V\setminus\{0\}} \|\bar{\mu}_\xi + \mu_\theta'\|_v < \infty \quad \text{and} \quad \sup_{\xi \in V\setminus\{0\}} \|\bar{\eta}_\xi + \eta_\theta'\|_u < \infty,$$

for any compact neighborhood V of 0. Therefore, applying the Cauchy-Schwartz inequality (1.22) together with Theorem 1.3 yields

$$\left| \xi \int g(s,t)\big((\bar{\mu}_\xi + \mu_\theta') \times (\bar{\eta}_\xi + \eta_\theta')\big)(ds, dt) \right| \leq |\xi| \cdot \|g\|_{v\otimes u} \cdot \|\bar{\mu}_\xi + \mu_\theta'\|_v \cdot \|\bar{\eta}_\xi + \eta_\theta'\|_u.$$

Letting $\xi \to 0$ in the above inequality it follows that the first term in (2.26) converges weakly to $\varnothing$.

The second and the third terms in (2.26) are symmetric so they can be treated similarly. For instance, for the second term in (2.26) note that

$$\int g(s,t)(\mu_\theta \times \bar{\eta}_\xi)(ds, dt) = \int \int g(s,t)\mu_\theta(ds)\, \bar{\eta}_\xi(dt) = \int H_\theta(g,t)\bar{\eta}_\xi(dt),$$

where $H_\theta(g,t) = \int g(s,t)\mu_\theta(ds)$ for all t and for all g. Theorem 1.3 implies that the pair $(\mathcal{D}(\mathbb{S}) \otimes \mathcal{D}(\mathbb{T}), v \otimes u)$ is a Banach base and by applying the Chauchy-Schwartz Inequality yields

$$\forall t \in \mathbb{T}: \ \frac{|H_\theta(g,t)|}{u(t)} \leq \frac{\|g(\cdot,t)\|_v}{u(t)} \|\mu_\theta\|_v \leq \|g\|_{v\otimes u}\, \|\mu_\theta\|_v,$$

where the second inequality follows from (see (1.30) within the proof of Theorem 1.3)

$$\forall s \in \mathbb{S}, t \in \mathbb{T} : \ |g(s,t)| \leq \|g\|_{v \otimes u} v(s) u(t).$$

Consequently, $H_\theta(g, \cdot) \in [\mathcal{D}(\mathbb{T})]_u$, for $g \in [\mathcal{D}(\mathbb{S}) \otimes \mathcal{D}(\mathbb{T})]_{v \otimes u}$. We have assumed that η_θ is $[\mathcal{D}(\mathbb{T})]_u$-differentiable, which yields $\bar{\eta}_\xi \overset{[\mathcal{D}]_u}{\Longrightarrow} \varnothing$. Hence,

$$\lim_{\xi \to 0} \int H_\theta(g,t) \bar{\eta}_\xi(dt) \to 0,$$

which shows that the second term in (2.26) converges weakly to $\varnothing$. This concludes the proof of the statement. □

Remark 2.5. *It is worth noting that if we see $\mathcal{D}$ as a particular class of functions, i.e., continuous or measurable, then the condition*

$$\mathcal{D}(\mathbb{S} \times \mathbb{T}) \subset \mathcal{D}(\mathbb{S}) \otimes \mathcal{D}(\mathbb{T})$$

is satisfied for any $\mathcal{D} \in \{\mathcal{C}, \mathcal{F}\}$, i.e., one considers the same class of functions $\mathcal{D}$ on $\mathbb{S}$, $\mathbb{T}$ and on the product space $\mathbb{S} \times \mathbb{T}$; see (1.28) and (1.29) in Example 1.7. It follows from Theorem 2.3 that weak differentiability is preserved by the product measure in both continuity and measurability paradigms. Since in the context of our applications we will consider a particular class of functions, e.g., continuous, measurable, we will denote by $\mathcal{D}$ the corresponding class of functions on each space under consideration.

For instance, choosing $\mathcal{D}(\mathbb{S}) = \mathcal{C}(\mathbb{S})$, $\mathcal{D}(\mathbb{T}) = \mathcal{C}(\mathbb{T})$, $v \equiv 1$, $u \equiv 1$ in Theorem 2.3 we conclude from (1.28) that weak $\mathcal{C}_B$-differentiability is preserved by the product measure, i.e., for each $g \in \mathcal{C}_B(\mathbb{S} \times \mathbb{T})$ it holds that

$$\frac{d}{d\theta} \int g(s,t) \mu_\theta(ds) \eta_\theta(dt) = \int g(s,t) \mu'_\theta(ds) \eta_\theta(dt) + \int g(s,t) \mu_\theta(ds) \eta'_\theta(dt).$$

This is asserted in [48] but no proof is given.

Extension to Finite Products of Measures

In what follows we extend Theorem 2.3 to finite product measures. To this end, let us consider a finite family of positive mappings $v_i \in \mathcal{C}^+(\mathbb{S}_i)$, a finite family of measure-valued mappings $\mu_{i,*} : \Theta \to \mathcal{M}_{v_i}(\mathbb{S}_i)$, for $1 \leq i \leq n$, and define the product mapping

$$\Pi_* : \Theta \to \mathcal{M}(\sigma(\mathcal{S}_1 \times \ldots \times \mathcal{S}_n)),$$

where $\sigma(\mathcal{S}_1 \times \ldots \times \mathcal{S}_n)$ denotes the product Borel field on $\mathbb{S}_1 \times \ldots \times \mathbb{S}_n$, as follows:

$$\forall \theta \in \Theta : \ \Pi_\theta = \mu_{1,\theta} \times \ldots \times \mu_{n,\theta}. \tag{2.27}$$

Moreover, to simplify the notation, we denote the tensor product $v_1 \otimes \ldots \otimes v_n$ (see (1.27) for a definition) by $\vec{v}$. In formula:

$$\forall (s_1, \ldots, s_n) \in \mathbb{S}_1 \times \ldots \times \mathbb{S}_n : \ \vec{v}(s_1, \ldots, s_n) = v_1(s_1) \cdot \ldots \cdot v_n(s_n). \tag{2.28}$$

The following statement follows by finite induction from Theorem 2.3.

Theorem 2.4. *If $\mu_{i,\theta}$ is weakly $[\mathcal{D}(\mathbb{S}_i)]_{v_i}$-differentiable, for $1 \leq i \leq n$, then the product measure Π_θ is weakly $[\mathcal{D}(\mathbb{S}_1) \otimes \ldots \otimes \mathcal{D}(\mathbb{S}_n)]_{\vec{v}}$-differentiable and*

$$\Pi'_\theta = \sum_{i=1}^{n} \mu_{1,\theta} \times \ldots \times \mu'_{i,\theta} \times \ldots \times \mu_{n,\theta}.$$

Moreover, if for $\theta \in \Theta$, $\mu_{i,\theta} \in \mathcal{M}^1(\mathbb{S}_i)$ and $\mu'_{i,\theta} = (c_{i,\theta}, \mu^+_{i,\theta}, \mu^-_{i,\theta})$, for $1 \leq i \leq n$, then an instance of the weak derivative Π'_θ is given by $(C_\theta, \Pi^+_\theta, \Pi^-_\theta)$, where

$$C_\theta = \sum_{i=1}^{n} c_{i,\theta}; \quad \Pi^\pm_\theta = \sum_{i=1}^{n} \frac{c_{i,\theta}}{C_\theta} \cdot \mu_{1,\theta} \times \ldots \times \mu^\pm_{i,\theta} \times \ldots \times \mu_{n,\theta}. \tag{2.29}$$

The following two results are immediate consequences of Theorem 2.4.

Corollary 2.2. *Consider the Banach base $((\mathcal{D}(\mathbb{S}), v)$ and denote the k-fold product of μ_θ by $\Pi_\theta(k)$, for $k \geq 1$. Assume that μ_θ has $[\mathcal{D}(\mathbb{S})]_v$-derivative $(c_\theta, \mu^+_\theta, \mu^-_\theta)$. Then $\Pi_\theta(n)$ is $[\mathcal{D}(\mathbb{S}_1) \otimes \ldots \otimes \mathcal{D}(\mathbb{S}_n)]_{\vec{v}}$-differentiable and we have*[4]

$$\frac{d}{d\theta} \int g(x) \Pi_\theta(n, dx) = c_\theta \sum_{j=1}^{n} \int g(s,t,u) \Pi_\theta(j-1, ds)(\mu^+_\theta - \mu^-_\theta)(dt) \Pi_\theta(n-j, du).$$

Proof. In Theorem 2.4 we let $\mu_{1,\theta} = \ldots = \mu_{n,\theta} = \mu_\theta$. If $\mu'_\theta = (c_\theta, \mu^+_\theta, \mu^-_\theta)$ then the conclusion follows from the equality $\frac{d}{d\theta} \int g(x)\Pi_\theta(n, dx) = \int g(x)\Pi'_\theta(n, dx)$. □

Corollary 2.3. Random Variable Version of Theorem 2.4: *Let $X_i \in \mathbb{S}_i$, for $1 \leq i \leq n$, be independent random variables, having distribution $\mu_{i,\theta}$, for $1 \leq i \leq n$, respectively. If for $1 \leq i \leq n$ the distribution $\mu_{i,\theta}$ is $[\mathcal{D}(\mathbb{S}_i)]_{v_i}$-differentiable, having derivative $(c_{i,\theta}, \mu^+_{i,\theta}, \mu^-_{i,\theta})$, then for any $g \in [\mathcal{D}(\mathbb{S}_1) \otimes \ldots \otimes \mathcal{D}(\mathbb{S}_n)]_{\vec{v}}$ it holds that*

$$\frac{d}{d\theta} P_g(\theta) = \sum_{j=1}^{n} c_{j,\theta} \mathbb{E}_\theta \left[g(X_1, \ldots, X^+_j, \ldots, X_n) - g(X_1, \ldots, X^-_j, \ldots, X_n) \right], \tag{2.30}$$

where we denote $P_g(\theta) = \mathbb{E}_\theta [g(X_1, \ldots, X_n)]$ and $\mathbb{E}_\theta$ denotes an expectation operator consistent with $(X_1, \ldots, X^\pm_j, \ldots, X_n) \sim \mu_{1,\theta} \times \ldots \times \mu^\pm_{j,\theta} \times \ldots \times \mu_{n,\theta}$, for $1 \leq j \leq n$.

Remark 2.6. *Note that, in Corollary 2.3, for any fixed j, $X^\pm_j$ should be independent of $\{X_i : i \neq j\}$. Nevertheless, note that it is not crucial that X^+_j and X^-_j are mutually independent. In addition, if we set $P^{j\pm}_g(\theta) := \mathbb{E}_\theta \left[g(X_1, \ldots, X^\pm_j, \ldots, X_n) \right]$, for $1 \leq j \leq n$, then $P^{j\pm}_g$ denotes the counterpart of P_g in a system where the j^{th} input variable X_j has been replaced by $X^\pm_j$. Hence, according to (2.30) we have*

$$\frac{d}{d\theta} P_g(\theta) = \sum_{j=1}^{n} c_{j,\theta} P^{j+}_g(\theta) - \sum_{j=1}^{n} c_{j,\theta} P^{j-}_g(\theta);$$

compare to (0.3). Therefore, an unbiased estimator for the stochastic gradient $\frac{d}{d\theta} P_g(\theta)$ can be obtained according to (0.4).

[4] By convention we disregard the void product $\Pi_\theta(0)$.

2.4 Non-Continuous Cost-Functions and Set-Wise Differentiation

In the literature, differentiation of a probability measure μ_θ has also been defined as differentiability of the corresponding set function. That is

$$\frac{d}{d\theta}\mu_\theta(A) = \mu'_\theta(A), \tag{2.31}$$

for each $A \in \mathcal{S}$; see, e.g., [39]. In fact, this holds true in the standard case when $\mathcal{D} = \mathcal{F}$ (in particular, when μ_θ is strongly differentiable). However, this is not always the case and the following example illustrates this. Taking ψ_θ to be the uniform distribution on $[0, \theta)$ and denoting the Lebesgue measure by ℓ it holds that

$$\forall x > 0 : \ \psi_\theta([0, x)) = \frac{1}{\theta}\ell([0, x) \cap [0, \theta)) = \frac{1}{\theta}\min\{x, \theta\}.$$

At $\theta = x$, the left-sided derivative of $\psi_\theta([0, x))$ equals 0 whereas the right-sided derivative equals $-1/x$. Hence, ψ_θ fails to be weakly differentiable in the set-wise sense, whereas it is shown in Example 2.6 that ψ_θ is $[\mathcal{C}]_v$-differentiable.

Motivated by this remark, in this section we aim to identify the sets A which satisfy (2.31), provided that μ_θ is $[\mathcal{C}]_v$-differentiable, for some $v \in \mathcal{C}^+$. More generally, we investigate under which conditions $[\mathcal{C}]_v$-differentiability is suitable for differentiating performance measures generated by non-continuous cost-functions. Specifically, if μ_θ is $[\mathcal{C}]_v$-differentiable then, by Definition 2.1, we have

$$\forall g \in [\mathcal{C}]_v : \ \frac{d}{d\theta}\int g(s)\mu_\theta(ds) = \int g(s)\mu'_\theta(ds). \tag{2.32}$$

However, the elements of $[\mathcal{C}]_v$ are, in general, not the only ones satisfying (2.32) as there might be non-continuous functions $g \in [\mathcal{F}]_v$ which satisfy (2.32), as well; see Remark 1.4.

Starting point of our analysis is the Portmanteau Theorem (see Theorem F.1 in the Appendix) which asserts that the sequence $\{\mu_n\}_n$ is $\mathcal{C}_B$-convergent to μ if and only if $\mu_n(A) \to \mu(A)$ for each continuity set A of μ, i.e., $\mu(\partial A) = 0$. More generally, if for arbitrary $g \in \mathcal{F}$ we denote by $D_g \subset \mathbb{S}$ the set of discontinuities[5] of g, then for each bounded $g \in \mathcal{F}$, such that $\mu(D_g) = 0$, it holds that

$$\lim_{n\to\infty}\int g(s)\mu_n(ds) = \int g(s)\mu(ds); \tag{2.33}$$

see Theorem F.2 in the Appendix. The above result can be easily extended from probability measures to general positive measures, as follows:

Lemma 2.4. *Let the sequence $\{\mu_n : n \in \mathbb{N}\} \subset \mathcal{M}^+$ be $[\mathcal{C}]_v$-convergent to μ. Then, for each mapping $g \in [\mathcal{F}]_v$, such that $\mu(D_g) = 0$, (2.33) holds true.*

Proof. First, we show that the statement holds true for $v = 1$, i.e., $[\mathcal{C}]_v = \mathcal{C}_B$. Indeed, by hypothesis, (2.33) holds true for each $g \in \mathcal{C}_B$. Letting $g = \mathbb{I}_\mathbb{S}$ in (2.33) it follows that $\mu_n(\mathbb{S}) \to \mu(\mathbb{S})$, for $n \to \infty$. Moreover, this implies

$$\mu(\mathbb{S}) < \infty, \quad \sup_{n\in\mathbb{N}}\mu_n(\mathbb{S}) < \infty.$$

[5] Note that, if $g = \mathbb{I}_A$, for some $A \in \mathcal{S}$, then $D_g = \partial A$.

If $\mu(\mathbb{S}) = 0$, i.e., μ is the null measure, then the conclusion is immediate. Otherwise, if $\mu(\mathbb{S}) > 0$, we define $\bar{\mu} \in \mathcal{M}^1$ as follows:

$$\forall A \in \mathcal{S} : \ \bar{\mu}(A) = \frac{\mu(A)}{\mu(\mathbb{S})}.$$

It follows that the set $\mathbb{N}_0 := \{n \in \mathbb{N} : \mu_n(\mathbb{S}) = 0\}$ is finite and by considering the sequence $\{\bar{\mu}_n : n \in \mathbb{N} \setminus \mathbb{N}_0\} \subset \mathcal{M}^1$, defined as

$$\forall A \in \mathcal{S} : \ \bar{\mu}_n(A) := \frac{\mu_n(A)}{\mu_n(\mathbb{S})},$$

for each $n \in \mathbb{N} \setminus \mathbb{N}_0$, we conclude that $\bar{\mu}_n$ is $\mathcal{C}_B$-convergent to $\bar{\mu}$. Since $\mu(D_g) = 0$ if and only if $\bar{\mu}(D_g) = 0$, it follows that for each bounded $g \in \mathcal{F}$, such that $\mu(D_g) = 0$, we have

$$\lim_{n\to\infty} \int g(s)\bar{\mu}_n(ds) = \int g(s)\bar{\mu}(ds).$$

Therefore, since $\mu_n(\mathbb{S}) \to \mu(\mathbb{S})$, for $n \to \infty$, it follows that

$$\int g(s)\mu_n(ds) = \mu_n(\mathbb{S}) \int g(s)\bar{\mu}_n(ds) \to \mu(\mathbb{S}) \int g(s)\bar{\mu}(ds) = \int g(s)\mu(ds),$$

provided that $\mu(D_g) = 0$, which proves the claim for $v = 1$.

Let now $v \in \mathcal{C}^+$. According to Remark 1.6, $[\mathcal{C}(\mathbb{S})]_v$-convergence of μ_n towards μ is equivalent to $\mathcal{C}_B(\mathbb{S}_v)$-convergence of $v \cdot \mu_n$ towards $v \cdot \mu$, where

$$\forall \eta \in \mathcal{M}_v : \ (v \cdot \eta)(ds) = v(s)\eta(ds).$$

By hypothesis, μ and μ_n, for $n \in \mathbb{N}$, are v-finite measures, i.e., belong to $\mathcal{M}_v$, which implies that $v \cdot \mu$ and $v \cdot \mu_n$, for $n \in \mathbb{N}$ are finite measures. Moreover, if $\Phi : [\mathcal{F}(\mathbb{S})]_v \to \mathcal{F}_B(\mathbb{S}_v)$ denotes the isometry defined in Example 1.5, i.e.,

$$\forall s \in \mathbb{S}_v, g \in [\mathcal{F}(\mathbb{S})]_v : \ (\Phi g)(s) = \frac{g(s)}{v(s)},$$

then it holds that $D_{\Phi g} \subset D_g$ and it follows that $\mu(D_{\Phi g}) = 0$, which implies $(v \cdot \mu)(D_{\Phi g}) = 0$. Therefore, choose an arbitrary $g \in [\mathcal{F}(\mathbb{S})]_v$. It follows that $\Phi g \in \mathcal{F}_B(\mathbb{S}_v)$ and from the first part of the proof, for $v = 1$, we conclude that

$$\lim_{n\to\infty} \int g(s)\mu_n(ds) = \lim_{n\to\infty} \int (\Phi g)(s)(v \cdot \mu_n)(ds) = \int (\Phi g)(s)(v \cdot \mu)(ds) = \int g(s)\mu(ds),$$

provided that $(v \cdot \mu)(D_{\Phi g}) = 0$. This concludes the proof. $\square$

Lemma 2.4 is the main technical tool that we use to analyze non-continuous cost-functions from a weak $[\mathcal{C}]_v$-differentiation perspective. In the following we apply this result to our differentiability setting. In particular, it will turn out that if μ_θ is regularly $[\mathcal{C}]_v$-differentiable then (2.31) holds true for each continuity set A of μ'_θ. More specifically, the following statement holds true.

Theorem 2.5. *If $\mu_* : \Theta \to \mathcal{M}_v^1$ is a $[\mathcal{C}]_v$-continuous measure-valued mapping such that μ_θ is regularly $[\mathcal{C}]_v$-differentiable then:*

(i) for each $g \in [\mathcal{F}]_v$, such that $|\mu'_\theta|(D_g) = 0$, it holds that

$$\frac{d}{d\theta}\int g(s)\mu_\theta(ds) = \int g(s)\mu'_\theta(ds). \tag{2.34}$$

(ii) if $A \in \mathcal{S}$ such that $\bar{A} \subset \mathbb{S}_v$ and A is a continuity set of μ'_θ then A satisfies (2.31).

Proof. Regular $[\mathcal{C}]_v$-differentiability of μ_θ implies that

$$\left[\frac{\mu_{\theta+\xi} - \mu_\theta}{\xi}\right]^+ \stackrel{[\mathcal{C}]_v}{\Longrightarrow} [\mu'_\theta]^+, \quad \left[\frac{\mu_{\theta+\xi} - \mu_\theta}{\xi}\right]^- \stackrel{[\mathcal{C}]_v}{\Longrightarrow} [\mu'_\theta]^-$$

and, since $|\mu'_\theta|(D_g) = 0$ implies $[\mu'_\theta]^\pm(D_g) = 0$, Lemma 2.4 concludes the proof of (2.34). Since $\bar{A} \subset \mathcal{S}_v$ implies that $\|\mathbb{I}_A\|_v < \infty$, letting now $g = \mathbb{I}_A$ in (2.34) concludes *(ii)*. □

Therefore, although weaker than $[\mathcal{F}]_v$-differentiability, regular $[\mathcal{C}]_v$-differentiability of μ_θ is a still a strong hypothesis since it implies that (2.32) holds true for each $g \in [\mathcal{F}(\mu'_\theta)]_v$, where we denote by $\mathcal{F}(\mu'_\theta)$ the linear space of $|\mu'_\theta|$-a.e. continuous functions, i.e.,

$$\mathcal{F}(\mu'_\theta) := \{g \in \mathcal{F} : \; |\mu'_\theta|(D_g) = 0\}.$$

Note that regularity is a crucial assumption in Theorem 2.5. Indeed, let us revisit the parametric distribution ϕ_θ introduced in Example 2.7, which is $\mathcal{C}_B$-differentiable for $\theta = 0$. Since $\phi'_0 = \varnothing$ the set $A = (0, \infty)$ is a continuity set for ϕ'_0 but it holds that

$$\frac{d}{d\theta}\phi_\theta(A)\Big|_{\theta=0} = -1 \neq 0 = \phi'_0(A).$$

The following result is an immediate consequence of Theorem 2.5.

Corollary 2.4. *Under the conditions put forward in Theorem 2.5, if there exists a neighborhood V of 0 such that for each $\xi \in V$ we have $\theta + \xi \in \Theta$ and the family*

$$\left\{\frac{\mu_{\theta+\xi} - \mu_\theta}{\xi} \; : \; \xi \in V \setminus \{0\}\right\}$$

is tight then each continuity set A of μ'_θ satisfies (2.31).

Proof. Note that, by hypothesis, both families

$$\left\{\left(\frac{\mu_{\theta+\xi} - \mu_\theta}{\xi}\right)^\pm \; : \; \xi \in V \setminus \{0\}\right\}$$

consist of positive measures and are tight. By Theorem 1.1 *(ii)* it follows that μ_θ is regularly $\mathcal{C}_B$-differentiable. Now taking into account that $\mathcal{C}_B = [\mathcal{C}]_v$, for $v = 1$ and that for each $A \in \mathcal{S}$ $D_{\mathbb{I}_A} = \partial A$, the proof follows from Theorem 2.5. □

Set-Wise Differentiation for Product Measures

We conclude this section by extending Theorem 2.5 to product measures. Note, however, that the decomposition $(C_\theta, \Pi_\theta^+, \Pi_\theta^-)$ in Theorem 2.4 is not orthogonal, even though one uses the orthogonal decomposition of $\mu'_{i,\theta}$, for each $1 \le i \le n$. Therefore, regular differentiability of $\mu_{i,\theta}$, for each $1 \le i \le n$, does not imply in a straightforward way that Π_θ is regularly differentiable and, in order to apply Theorem 2.5 to Π_θ an additional argument is needed. The following (weaker) result turns out to be useful in applications.

Theorem 2.6. *If $\mu_{i,*} : \Theta \to \mathcal{M}_v^1$ are such that $\mu_{i,\theta}$ is regularly $[\mathcal{C}]_{v_i}$-differentiable, for $1 \le i \le n$, then for each measurable $g \in [\mathcal{F}]_{\vec{v}}$ satisfying*

$$\forall 1 \le i \le n: \ (|\mu_{1,\theta}| \times \ldots \times |\mu'_{i,\theta}| \times \ldots \times |\mu_{n,\theta}|)(D_g) = 0 \tag{2.35}$$

it holds that

$$\frac{d}{d\theta} \int g(x)\Pi_\theta(dx) = \int g(x)\Pi'_\theta(dx). \tag{2.36}$$

Proof. By Theorem 2.4, Π_θ is $[\mathcal{C}(\mathbb{S}_1 \times \ldots \times \mathbb{S}_n)]_{\vec{v}}$-differentiable; see Remark 2.5. Moreover, note that for any ξ such that $\theta + \xi \in \Theta$ we have

$$\frac{\Pi_{\theta+\xi} - \Pi_\theta}{\xi} = \sum_{i=1}^{n} \mu_{1,\theta+\xi} \times \ldots \times \mu_{i-1,\theta+\xi} \times \frac{\mu_{i,\theta+\xi} - \mu_{i,\theta}}{\xi} \times \mu_{i+1,\theta} \times \ldots \times \mu_{n,\theta}.$$

Hence, $(\Pi_{\theta+\xi} - \Pi_\theta)/\xi = \Upsilon_\xi^+ - \Upsilon_\xi^- -$, for some $\Upsilon_\xi^\pm \in \mathcal{M}^+$, i.e.,

$$\Upsilon_\xi^\pm := \sum_{i=1}^{n} \mu_{1,\theta+\xi} \times \ldots \times \mu_{i-1,\theta+\xi} \times \left[\frac{\mu_{i,\theta+\xi} - \mu_{i,\theta}}{\xi}\right]^\pm \times \mu_{i+1,\theta} \times \ldots \times \mu_{n,\theta}$$

and regular differentiability of $\mu_{i,\theta}$, for $1 \le i \le n$, ensures that for $\xi \to 0$ we have

$$\Upsilon_\xi^\pm \overset{[\mathcal{C}]_{\vec{v}}}{\Longrightarrow} \Upsilon^\pm := \sum_{i=1}^{n} \mu_{1,\theta} \times \ldots \times \mu_{i-1,\theta} \times \left[\mu'_{i,\theta}\right]^\pm \times \mu_{i+1,\theta} \times \ldots \times \mu_{n,\theta}.$$

Choose now $g \in [\mathcal{F}]_{\vec{v}}$ satisfying (2.35). It follows that

$$\Upsilon^\pm(D_g) \le \sum (|\mu_{1,\theta}| \times \ldots \times |\mu'_{i,\theta}| \times \ldots \times |\mu_{n,\theta}|)(D_g) = 0$$

and by Lemma 2.4 we obtain

$$\frac{d}{d\theta} \int g(x)\Pi_\theta(dx) = \lim_{\xi \to 0} \int g(x)(\Upsilon_\xi^+ - \Upsilon_\xi^-)(dx) = \int g(x)(\Upsilon^+ - \Upsilon^-)(dx). \tag{2.37}$$

Finally, by the uniqueness of the $[\mathcal{C}]_v$-limit it follows that $\Pi'_\theta = \Upsilon^+ - \Upsilon^-$ and using that in (2.37) concludes the proof of (2.36). □

2.5 Gradient Estimation Examples

In this section we present some basic applications of the weak differentiation theory. More specifically, if X is a random variable describing the state of a stochastic system in which the input distributions are weakly differentiable with respect to some design parameter θ, we provide an unbiased estimator for the gradient

$$\frac{d}{d\theta}\mathbb{E}_\theta[g(X)],$$

for a certain class of performance measures g for which the above expression makes sense. The main theoretical tool to be used will be Theorem 2.4 which provides a representation of the weak derivative of the product measure. In Section 2.5.1 we construct a gradient estimator for a ruin probability, i.e., X is the indicator of the ruin event in some insurance model whereas in Section 2.5.2 we let X be the transient waiting time in a G/G/1 queue.

2.5.1 The Derivative of a Ruin Probability

Let us consider the following example. An insurance company receives premiums from clients at some constant rate $r > 0$ while claims $\{Y_i : i \geq 1\}$ arrive according to a Poisson process with rate $\lambda > 0$. Let $\{X_i : i \geq 1\}$ denote the inter-arrival times of the Poisson process and let N_τ denote the number of claims recorded up to some fixed time horizon $\tau > 0$. Assume further that the values of the claims are i.i.d. random variables following a Pareto distribution π_θ, i.e.,

$$\pi_\theta(dx) = \frac{\beta\theta^\beta}{x^{\beta+1}}\mathbb{I}_{(\theta,\infty)}(x)dx,$$

for some $\beta > 0$ and assume that the claims are independent of the Poisson process.

Let $V(0) \geq 0$ denote the initial credit of the insurance company. The credit (resp. debt) of the company right after the n^{th} claim, denoted by $V(n)$, follows the recursion

$$\forall n \geq 0: \; V(n+1) = V(n) + rX_{n+1} - Y_{n+1}.$$

Ruin occurs before time τ if at least one $n \leq N_\tau$ exists such that $V(n) < 0$. See Figure 2.1.

We are interested in estimating the derivative with respect to θ of the probability of ruin up to time τ. To this end, we denote by $\mathfrak{R}_\tau$ the event that ruin occurs before time τ and note that

$$\mathfrak{R}_\tau \cap \{N_\tau = n\} = \complement\left(\bigcap_{k=1}^{n}\{V(k) > 0\}\right) = \complement\left\{r \cdot \sum_{i=1}^{j} X_i > \sum_{i=1}^{j} Y_i, \forall 1 \leq j \leq n\right\}.$$

Hence, considering the sequence $\{g_n : n \geq 1\}$, $g_n \in \mathcal{F}(\mathbb{R}^{2n})$ given by

$$g_n(x_1, \ldots, x_n, y_1, \ldots, y_n) = 1 - \prod_{j=1}^{n}\mathbb{I}_{\{r\cdot\sum_{i=1}^{j} x_i > \sum_{i=1}^{j} y_i\}}(x_1, \ldots, x_n, y_1, \ldots, y_n) \tag{2.38}$$

we can write for each $n \geq 1$

$$\mathbb{P}_\theta(\mathfrak{R}_\tau \cap \{N_\tau = n\}) = \mathbb{E}_\theta\left[\mathbb{I}_{\{N_\tau = n\}} g_n(X_1, \ldots, X_n, Y_1, \ldots, Y_n)\right], \tag{2.39}$$

where $\mathbb{E}_\theta$ is an expectation operator consistent with $(X_1, \ldots, X_n, Y_1, \ldots, Y_n) \sim \mu^n \times \pi_\theta^n$ and μ denotes the exponential distribution with rate λ.

As explained in Section 2.2.3, the truncated distribution π_θ is regularly $\mathcal{C}_B$-differentiable and its weak derivative satisfies

$$\pi'_\theta = \frac{\beta}{\theta}(\pi_\theta - \delta_\theta).$$

Therefore, if we let $v = 1$ and

$$\mu_{i,\theta} := \begin{cases} \pi_\theta, & 1 \leq i \leq n, \\ \mu, & n+1 \leq i \leq 2n, \end{cases}$$

we conclude by Theorem 2.4 that the product measure

$$\Pi_\theta = \pi_\theta^n \times \mu^n$$

is $\mathcal{C}_B$-weakly differentiable and one can, according to (2.29), derive an instance of the weak derivative of Π_θ using the following representation for the weak derivatives of the input distributions:

$$\mu'_{i,\theta} := \begin{cases} (\frac{\beta}{\theta}, \pi_\theta, \delta_\theta), & 1 \leq i \leq n, \\ (1, \mu, \mu), & n+1 \leq i \leq 2n. \end{cases}$$

Therefore, one can write (see Remark 2.5)

$$\forall g \in \mathcal{C}_B(\mathbb{S}^{2n}) : \; \frac{d}{d\theta} \int g(s) \Pi_\theta(ds) = \int g(s) \Pi'_\theta(ds), \tag{2.40}$$

where, according to (2.29), we have

$$\begin{aligned} \Pi'_\theta &= \frac{\beta}{\theta} \sum_{i=1}^n \pi_\theta^{i-1} \times (\pi_\theta - \delta_\theta) \times \pi_\theta^{n-i} \times \mu^n \\ &= \frac{\beta}{\theta} \sum_{i=1}^n \left(\Pi_\theta - \pi_\theta^{i-1} \times \delta_\theta \times \pi_\theta^{n-i} \times \mu^n \right). \end{aligned} \tag{2.41}$$

Note, however, that the cost-function g_n introduced for modeling the ruin probability is not continuous; in formula: $g_n \notin \mathcal{C}_B$. Fortunately, by virtue of Theorem 2.6, the equality in (2.40) still holds true if g satisfies

$$\forall 1 \leq i \leq n : \; (|\pi_\theta|^{i-1} \times |\pi'_\theta| \times |\pi_\theta|^{n-i} \times \mu^n)(D_g) = 0.$$

In our case, we have

$$D_{g_n} = \partial \left(\mathfrak{R}_\tau \cap \{N_\tau = n\} \right) \subset \bigcup_{i=1}^n \left\{ r \cdot \sum_{j=1}^i x_j = \sum_{j=1}^i y_j \right\},$$

which yields

$$(|\pi_\theta|^{i-1} \times |\pi'_\theta| \times |\pi_\theta|^{n-i} \times \mu^n)(D_{g_n}) \leq \sum_{i=1}^n \mathbb{P}_\theta \left(\left\{ r \cdot \sum_{j=1}^i X_j = \sum_{j=1}^i Y_j \right\} \right) = 0,$$

since X_i has a continuous (exponential) distribution, for each $1 \leq i \leq n$. Hence, (2.40) applies for $g = g_n$, even though $g_n \notin \mathcal{C}_B$.

Examining (2.41) we note that, while Π_θ represents the distribution of the initial process, the product measure

$$\pi_\theta^{i-1} \times \delta_\theta \times \pi_\theta^{n-i} \times \mu^n$$

represents the distribution of a modified process $V_i(\cdot)$, where the size of the i^{th} claim has been replaced by the constant θ. Consequently, if $\mathfrak{R}_\tau^i$ denotes the event that ruin occurs before time τ, when the value of the i^{th} claim is replaced by the constant θ, then, letting $g = g_n$ in (2.40), it follows from (2.39) that

$$\begin{aligned} \frac{d}{d\theta}\mathbb{P}_\theta\left(\mathfrak{R}_\tau \cap \{N_\tau = n\}\right) &= \frac{\beta}{\theta}\sum_{i=1}^{n}\left(\mathbb{P}_\theta\left(\mathfrak{R}_\tau \cap \{N_\tau = n\}\right) - \mathbb{P}_\theta\left(\mathfrak{R}_\tau^i \cap \{N_\tau = n\}\right)\right) \\ &= \frac{\beta}{\theta}\sum_{i=1}^{n}\mathbb{P}_\theta\left(\left(\mathfrak{R}_\tau \setminus \mathfrak{R}_\tau^i\right) \cap \{N_\tau = n\}\right), \end{aligned} \tag{2.42}$$

where the last equality follows from the observation that $Y_i > \theta$, which implies that $\mathfrak{R}_\tau^i \subset \mathfrak{R}_\tau$. Moreover, the difference $\mathfrak{R}_\tau \setminus \mathfrak{R}_\tau^i$ represents the event that ruin occurs up to time τ but it does not occur anymore if one reduces the value of the i^{th} claim by $Y_i - \theta$. A graphical representation of these facts can be found in Figure 2.1, where the dashed line represents the modified process $V_i(\cdot)$. One can easily note that the event is incompatible with $\{N_\tau < i\}$, i.e., if the "reduced claim" comes after time τ. Hence, it holds that

$$\forall i \geq 1 : \ \mathbb{P}_\theta\left(\mathfrak{R}_\tau \setminus \mathfrak{R}_\tau^i\right) = \mathbb{P}_\theta\left(\left(\mathfrak{R}_\tau \setminus \mathfrak{R}_\tau^i\right) \cap \{N_\tau \geq i\}\right). \tag{2.43}$$

Let us consider now the following elementary identity[6]

$$\mathbb{P}_\theta\left(\mathfrak{R}_\tau\right) = \sum_{n=1}^{\infty}\mathbb{P}_\theta\left(\mathfrak{R}_\tau \cap \{N_\tau = n\}\right).$$

Provided that interchanging infinite summation with differentiation is allowed, we obtain the following sequence of equalities

$$\begin{aligned} \frac{d}{d\theta}\mathbb{P}_\theta\left(\mathfrak{R}_\tau\right) &= \sum_{n=1}^{\infty}\frac{d}{d\theta}\mathbb{P}_\theta\left(\mathfrak{R}_\tau \cap \{N_\tau = n\}\right) &(2.44)\\ &\overset{(2.42)}{=} \sum_{n=1}^{\infty}\frac{\beta}{\theta}\sum_{i=1}^{n}\mathbb{P}_\theta\left(\left(\mathfrak{R}_\tau \setminus \mathfrak{R}_\tau^i\right) \cap \{N_\tau = n\}\right) &(2.45)\\ &\overset{(*)}{=} \frac{\beta}{\theta}\sum_{i=1}^{\infty}\mathbb{P}_\theta\left(\left(\mathfrak{R}_\tau \setminus \mathfrak{R}_\tau^i\right) \cap \{N_\tau \geq i\}\right) &(2.46)\\ &\overset{(2.43)}{=} \frac{\beta}{\theta}\sum_{i=1}^{\infty}\mathbb{P}_\theta\left(\mathfrak{R}_\tau \setminus \mathfrak{R}_\tau^i\right), &(2.47) \end{aligned}$$

[6] Note that ruin can not occur if $N_\tau = 0$, i.e., if no claim is recorded up to time τ.

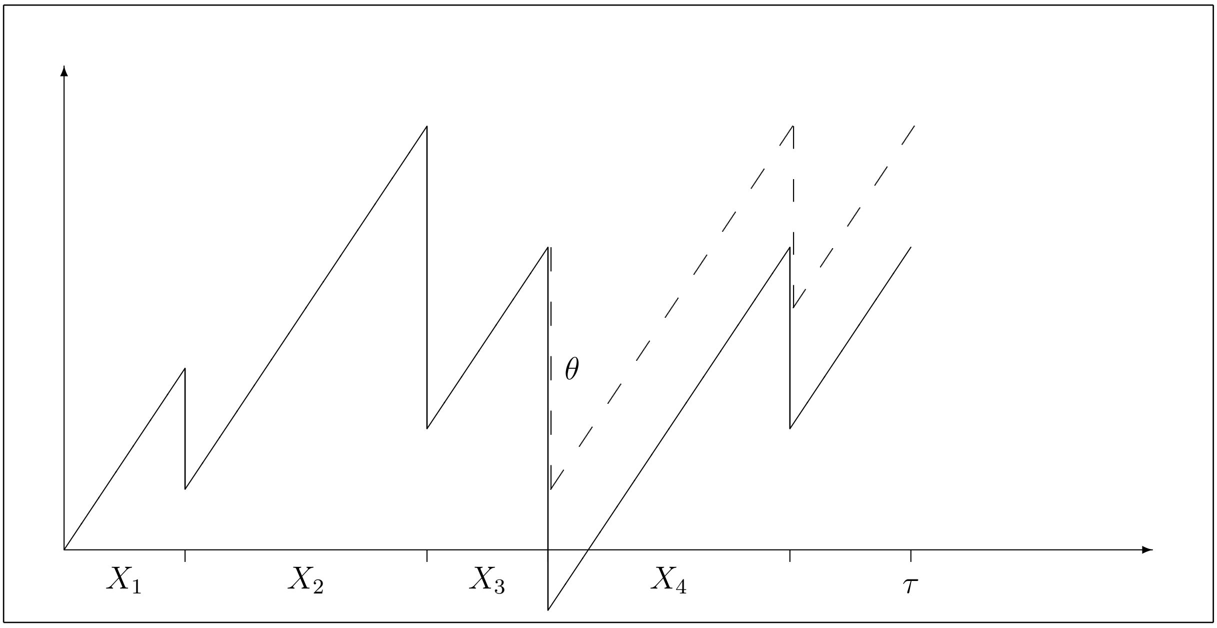

Fig. 2.1: An occurrence of the event $\mathfrak{R}_\tau \setminus \mathfrak{R}_\tau^3$ and $N_\tau = 4$. The dashed line represents a version of the process where the value of the 3^{rd} claim is reduced.

where the equality (*) follows by changing the summation order in (2.45), which is allowed because the series in (2.46) is absolutely convergent; see Theorem A.1 in the Appendix. Moreover, the k^{th} remainder term of the series in (2.46) can be bounded as follows:

$$\sum_{i=k+1}^{\infty} \mathbb{P}_\theta\left((\mathfrak{R}_\tau \setminus \mathfrak{R}_\tau^i) \cap \{N_\tau \geq i\}\right) \leq \sum_{i=k+1}^{\infty} \mathbb{P}_\theta(\{N_\tau \geq i\}) \leq e^{-\lambda\tau} \sum_{i=k+1}^{\infty} \sum_{j=i}^{\infty} \frac{(\lambda\tau)^j}{j!}. \tag{2.48}$$

Note that the above bound is independent of θ. Interchanging limit with differentiation in (2.44) is justified (see Theorem B.3 in the Appendix) provided that we deal with an uniformly convergent series of functions in θ. Hence, it suffices to show that the double sum in (2.48) converges to 0 as $k \to \infty$. To see that, choose $k \geq 1$ such that $(\lambda\tau)/(k+1) < q < 1$. In particular, it follows that for each $j \geq k+1$ it holds that $(\lambda\tau)/j < q < 1$. Then we have

$$\sum_{i=k+1}^{\infty} \sum_{j=i}^{\infty} \frac{(\lambda\tau)^j}{j!} \leq \frac{(\lambda\tau)^k}{k!} \sum_{i=k+1}^{\infty} \sum_{j=i}^{\infty} q^{j-k} = \frac{(\lambda\tau)^k}{k!} \frac{q^{-k}}{1-q} \sum_{i=k+1}^{\infty} q^i = \frac{(\lambda\tau)^k}{k!} \frac{q}{(1-q)^2}.$$

Now choose an arbitrary $\epsilon > 0$ and increase (if necessary) k in order to obtain

$$\frac{(\lambda\tau)^k}{k!} \leq \frac{\epsilon(1-q)^2}{q}.$$

Since ϵ was arbitrary chosen, we conclude that (2.47) holds true and the expression

$$\hat{\partial}_\theta(n) := \frac{\beta}{\theta} \sum_{j=1}^{n} \left(g_n(X_1, \ldots, X_n, Y_1, \ldots, Y_n) - g_n(X_1, \ldots, X_n, Y_1, \ldots, Y_{j-1}, \theta, Y_{j+1}, \ldots, Y_n)\right)$$

provides an asymptotically unbiased estimator, i.e., the sequence $\hat{\partial}_\theta(n)$ converges in mean to a unbiased estimator, as $n \to \infty$, for the derivative of the ruin probability.

2.5.2 Differentiation of the Waiting Times in a G/G/1 Queue

Let us consider a G/G/1 queue where the service times have distribution μ and inter-arrival times have distribution η. If $\{W_n : n \geq 1\}$ denotes the sequence of waiting times, then Lindley's recursion yields

$$\forall n \geq 1 : \; W_{n+1} = \max\{W_n + S_n - T_{n+1}, 0\},$$

where $\{S_n : n \geq 1\}$ and $\{T_n : n \geq 1\}$ denote the sequence of service and inter-arrival times, respectively.

Let us assume that the service time distribution $\mu = \mu_\theta$ depends on some parameter $\theta \in \Theta \subset \mathbb{R}$ and the inter-arrival time distribution η is independent of θ. We will investigate under which conditions the distribution of the $(n+1)^{st}$ waiting time W_{n+1}, which obviously will depend on θ, is weakly differentiable. To this end, we aim to apply Theorem 2.4 and we consider the sequence of mappings $\{w_n : \; n \geq 1\}$, $w_n : \mathbb{R}^{2n} \to \mathbb{R}$, defined as $w_1(s,t) = \max\{s-t, 0\}$ and

$$\forall n \geq 1, \sigma, \tau \in \mathbb{R}^n, s, t \in \mathbb{R} : \; w_{n+1}(\sigma, s, \tau, t) = \max\{w_n(\sigma, \tau) + s - t, 0\}. \tag{2.49}$$

Note that $w_n \in \mathcal{C}(\mathbb{R}^{2n})$ and from Lindley's recursion it follows that the waiting times satisfy

$$\forall n \geq 1 : \; W_{n+1} = w_n(S_1, \ldots, S_n, T_2, \ldots, T_{n+1}).$$

In what follows, we fix $n \geq 1$ and assume that μ_θ is $[\mathcal{D}]_v$-differentiable, having derivative $\mu'_\theta = (c_\theta, \mu^+_\theta, \mu^-_\theta)$. By letting

$$\mu_{i,\theta} := \begin{cases} \mu_\theta, & 1 \leq i \leq n, \\ \eta, & n+1 \leq i \leq 2n, \end{cases}$$

it follows that $\mu_{i,\theta}$ is $[\mathcal{D}]_v$-differentiable, for all $1 \leq i \leq 2n$, with derivatives

$$\mu'_{i,\theta} = \begin{cases} (c_\theta, \mu^+_\theta, \mu^-_\theta), & 1 \leq i \leq n, \\ (1, \eta, \eta), & n+1 \leq i \leq 2n, \end{cases}$$

provided that $v \in \mathcal{L}^1(\eta)$. Therefore, Theorem 2.4 applies and leads us to conclude that the distribution of W_{n+1} is weakly $[\mathcal{D}]_\vartheta$-differentiable, for all $\vartheta \in \mathcal{C}^+([0,\infty))$ satisfying

$$\|\vartheta \circ w_n\|_{\vec{v}} < \infty,$$

where, in this case, we have (see (2.28) for the definition of $\vec{v}$)

$$\forall s_1, \ldots, s_n, t_1, \ldots, t_n : \; \vec{v}(s_1, \ldots, s_n, t_1, \ldots, t_n) = \prod_{i=1}^{n} v(s_i) \prod_{i=1}^{n} v(t_i).$$

Now continuity of w_n implies that the distribution of W_{n+1} is weakly $[\mathcal{D}]_\vartheta$-differentiable, for any $\vartheta \in \mathcal{C}^+([0,\infty))$, satisfying

$$\sup_{\substack{s_1,\ldots,s_n\in\mathbb{R}\\ t_1,\ldots,t_n\in\mathbb{R}}} \frac{\vartheta(w_n(s_1,\ldots,s_n,t_1,\ldots,t_n))}{\prod_{i=1}^n v(s_i)\prod_{i=1}^n v(t_i)} < \infty. \tag{2.50}$$

Note that (2.50) holds true if ϑ is non-decreasing and satisfies

$$\forall x,y \geq 0:\ \vartheta(x+y) \leq \gamma v(x)v(y), \tag{2.51}$$

for some $\gamma > 0$. Indeed, note that for all $n \geq 1$, w_n in (2.49) satisfies

$$\forall s_1,\ldots,s_n,t_1,\ldots,t_n \in \mathbb{R}:\ w_n(s_1,\ldots,s_n,t_1,\ldots,t_n) \leq s_1+\ldots+s_n+t_1+\ldots+t_n.$$

Using monotonicity of ϑ, we conclude with (2.51) that

$$\sup_{\substack{s_1,\ldots,s_n\in\mathbb{R}\\ t_1,\ldots,t_n\in\mathbb{R}}} \frac{\vartheta(w_n(s_1,\ldots,s_n,t_1,\ldots,t_n))}{\prod_{i=1}^n v(s_i)\prod_{i=1}^n v(t_i)} \leq \gamma^{2n-1}.$$

In particular, if for all $x \geq 0$, $v(x) = \vartheta(x) = e^{\alpha x}$, for some $\alpha \geq 0$, then (2.51) is fulfilled for $\gamma = 1$. We conclude that if μ_θ is $[\mathcal{D}]_{v_\alpha}$-differentiable, then the distribution of W_{n+1} is $[\mathcal{D}]_{v_\alpha}$-differentiable as well. For later reference we synthesize our analysis into the following statement:

Theorem 2.7. *Let $v_\alpha(x) = e^{\alpha x}$, for all $x \geq 0$, for some $\alpha \geq 0$. If the service times distribution μ_θ is $[\mathcal{D}]_{v_\alpha}$-differentiable then the distribution of the $(n+1)^{st}$ waiting time is $[\mathcal{D}]_{v_\alpha}$-differentiable, for each $n \geq 1$, and it holds that*

$$\forall f \in [\mathcal{D}]_{v_\alpha}:\ \frac{d}{d\theta}\mathbb{E}_\theta[f(W_{n+1})] = c_\theta \sum_{k=1}^n \mathbb{E}_\theta[f(W_{n+1}^{k+}) - f(W_{n+1}^{k-})], \tag{2.52}$$

where in accordance with Corollary 2.3 we have

$$W_{n+1}^{k\pm} = w_n(S_1,\ldots,S_k^\pm,\ldots,S_n,T_2,\ldots,T_{n+1}) \tag{2.53}$$

and $\mathbb{E}_\theta$ is an expectation operator consistent with

$$\forall 1 \leq k \leq n:\ (S_1,\ldots,S_k^\pm,\ldots,S_n,T_2,\ldots,T_{n+1}) \sim \mu_\theta^{k-1}\times\mu_\theta^\pm\times\mu_\theta^{n-k}\times\eta^n.$$

Proof. We apply Corollary 2.3 to the family of random variables $\{X_i : 1 \leq i \leq 2n\}$ defined as

$$X_i := \begin{cases} S_i, & 1 \leq i \leq n; \\ T_{i-n+1}, & n+1 \leq i \leq 2n. \end{cases}$$

Since, by hypothesis, μ_θ is $\mathcal{C}_{v_\alpha}$-differentiable, having weak derivative $\mu_\theta' = (c_\theta, \mu_\theta^+, \mu_\theta^-)$ and η is trivially $\mathcal{C}_B$-differentiable, i.e., $\mathcal{C}_{v_\alpha}$-differentiable for $\alpha = 0$, its derivative being nonsignificant, then (2.52) follows from (2.30) by letting $g = w_n$.

To complete the proof, one has to show that for each $\alpha \geq 0$ v_α satisfies

$$\forall s_1, \ldots, s_n, t_1, \ldots, t_n : \; v_\alpha(w_n(s_1, \ldots, s_n, t_1, \ldots, t_n)) \leq \prod_{i=1}^{n} v_\alpha(s_i). \tag{2.54}$$

To see that, note that, for $n \geq 1$, w_n in (2.49) satisfies

$$\forall s_1, \ldots, s_n, t_1, \ldots, t_n : \; w_n(s_1, \ldots, s_n, t_1, \ldots, t_n) \leq s_1 + \ldots + s_n;$$

the proof follows by induction. Now monotonicity of v_α concludes the proof of (2.54). □

Analyzing equation (2.53) we note that $W^{k\pm}_{n+1}$ denotes the $(n+1)^{st}$ waiting time in a modified queue, where the k^{th} service time S_k has been replaced by S^+_k and S^-_k, respectively. Hence, for each $k \geq 1$, one can construct two parallel processes $W^{k\pm}_{\cdot}$ whose sample paths coincide with those of the original process up to time k, after time $k+1$ follow a parallel path with that of the original process and once that the "higher path" reaches level 0 the two paths coincide again (the two processes couple). A graphical representation of the two parallel processes $\{W^{k\pm}_n : n \geq 1\}$ can be seen in Figure 2.2.

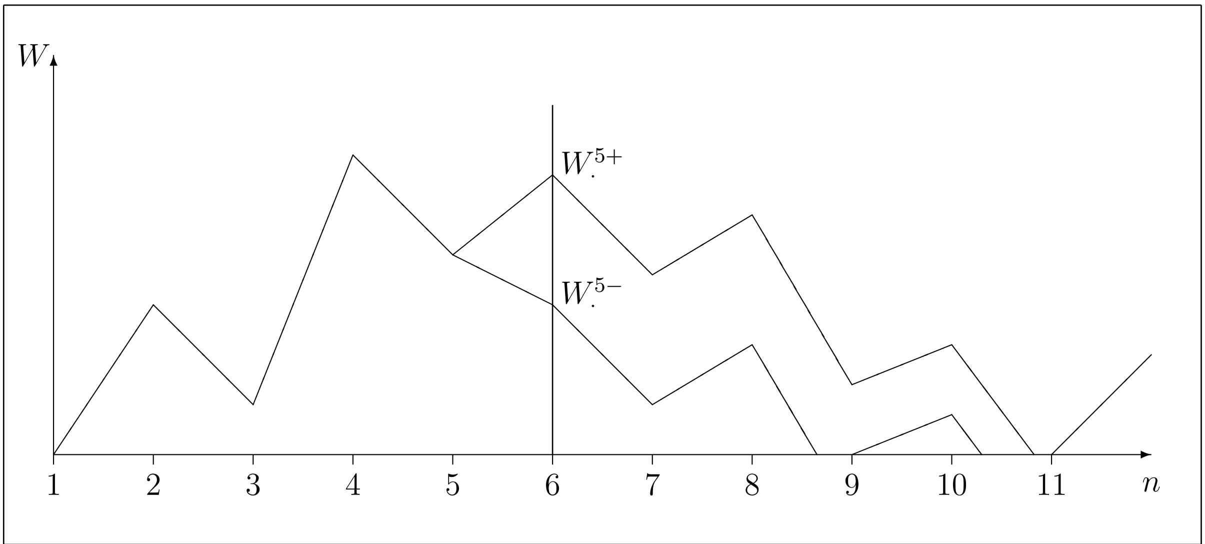

Fig. 2.2: A sample path of the parallel processes $\{W^{k\pm}_n\}_{n\geq 1}$, for $k=5$. They are obtained by replacing the 5^{th} service time in the original queue by S^+_5 and S^-_5, respectively.

In particular, Theorem 2.7 shows that the expression

$$\hat{\partial}_\theta := c_\theta \sum_{k=1}^{n} \left(f(W^{k+}_{n+1}) - f(W^{k-}_{n+1}) \right),$$

with $W^{k\pm}_{n+1}$ given by (2.53), provides an unbiased estimator for the gradient $\frac{d}{d\theta}\mathbb{E}_\theta[f(W_{n+1})]$. More specifically, provided that one has the means to simulate $f(W_{n+1})$, then by parallel simulations one can also simulate the stochastic gradient of $f(W_{n+1})$. While the joint distribution of the pair (S^+_k, S^-_k) plays no role in Theorem 2.7, it becomes crucial when simulating the two parallel processes. For a better performance it is recommended that the correlation of the two random variables to be maximal; see [48]. For more details on the relation between weak derivatives and unbiased estimators, we refer to [31].

2.6 Concluding Remarks

Throughout this chapter much work has been put into formalizing and studying a few relevant types of measure-valued differentiation. Concepts such as weak and strong differentiation have already been treated in the literature; see, e.g., [27], [32], [48] and recall that strong differentiation is a particular case of Fréchet differentiation. The concept of regular differentiation, however, is rather new and is meant to ensure a "smooth" extension of some properties related to classical weak convergence of measures to general signed measures. It turns out that, for some applications, e.g., set-wise differentiation, weak differentiability, which is a minimal differentiability condition, is not sufficient while strong differentiability is a too strong condition as it is not enjoyed by an important class of distributions; e.g., truncated distributions. Therefore, since regular differentiation is a property enjoyed by most of the common weakly differentiable distributions it makes sense to consider and study such a concept.

An important aspect of the theory of measure-valued differentiation is the "triple" representation of the derivatives of probability measures which makes possible to represent the derivatives of an expected value as the re-scaled difference between two expected values; see (2.19), or, when dealing with product measures, as a linear combination of expected values; see Corollary 2.3. This fact is important in simulations as it allows for unbiased (resp. asymptotically unbiased) gradient estimations with reduced variance for the transient (resp. steady-state) performance measures of complex stochastic systems, compared to other parallel methods such as infinitesimal perturbation analysis and score functions method; see, e.g., [25], [33], [39], [47], [48]. However, establishing the accuracy of the estimates is subject for future research.

Most of the results put forward in this chapter are new and are based on classical theory of weak convergence of probability measures and the link between measure theory and functional analysis. Out of these results I would like to point out Theorem 2.3, which is crucial for establishing weak differentiability of product measures and makes this theory fruitful for applications. It also provides the means to represent the weak derivative of the product measure which leads to gradient estimations for complex systems; see Section 2.5. I would also like to mention Theorem 2.2, which establishes sufficient conditions for strong differentiability, for which I am grateful to Prof.Dr. A. Hordijk for his contribution in establishing this result. The definition of the new concept of regular differentiability is motivated by the results in Section 2.4 which lead to gradient estimations for non-continuous performance measures in Section 2.5.1.

The theory of weak differentiation has been successfully applied to discrete-time stochastic processes, e.g., random walks or, more generally, Markov chains; see [22], [26], [27], [31], [32] and [33]. An interesting topic for future research is to extend these techniques to continuous-time processes, e.g., diffusions, Lévy processes, and to see to what extent the resulting theory overlaps with the well known *Malliavin Calculus*.

Eventually, an important topic for future research is to develop applications in the area of stochastic optimization and risk theory, based on weak differentiation theory.

3. STRONG BOUNDS ON PERTURBATIONS BASED ON LIPSCHITZ CONSTANTS

It is known that, in classical analysis one can easily establish bounds on the variations of a differentiable function by using the Mean Value Theorem. More specifically, differentiability implies local Lipschitz continuity, i.e., the variation of a differentiable function can be bounded by means of Lipschitz constants, provided that the derivative is bounded on a given domain. This chapter is intended to extend the classical results to measure-valued mappings in order to establish bounds on perturbations for the performance measures of parameter-dependent stochastic models.

3.1 Introduction

In classical analysis, a function $f : \mathbb{S} \to \mathbb{T}$, where $(\mathbb{S}, d_{\mathbb{S}})$ and $(\mathbb{T}, d_{\mathbb{T}})$ are metric spaces, is called *Lipschitz continuous on* $A \subset \mathbb{S}$ if there exists some constant $L > 0$ such that

$$\forall s_1, s_2 \in A : \; d_{\mathbb{T}}(f(s_1), f(s_2)) \leq L \cdot d_{\mathbb{S}}(s_1, s_2) \tag{3.1}$$

and it is called *locally Lipschitz continuous* if for each $s \in \mathbb{S}$ there exists a neighborhood U of s such that f is Lipschitz continuous on U. Any constant L satisfying (3.1) is called a *Lipschitz constant*. In addition, f is said to be Lipschitz continuous if it is Lipschitz continuous on $\mathbb{S}$.

Obviously, Lipschitz continuity implies local Lipschitz continuity, but the converse is, in general, not true. A standard counterexample is the function $f(x) = 1/x$ on $(0, \infty)$. Further, local Lipschitz continuity implies (uniform) continuity but the converse is, again, not true since any real-valued function which is continuous but nowhere differentiable (e.g., Weierstrass's function) is not locally Lipschitz continuous.

On Banach spaces, most common examples of locally Lipschitz continuous functions are the Fréchet differentiable functions. Moreover, if f is Fréchet differentiable and its derivative is bounded on some domain A, it follows from the Mean Value Theorem that f is Lipschitz continuous on A. In fact, on Euclidian spaces Lipschitz continuity is essentially equivalent to Fréchet differentiability. More specifically, a function $f : A \subset \mathbb{R}^n \to \mathbb{R}^m$ is Lipschitz continuous if and only if it is differentiable almost everywhere and the essential supremum of its derivative is finite (Rademacher's Theorem).

Lipschitz constants play an important role in perturbation/sensitivity analysis as they provide bounds on the variation of functions. Starting from the fact that Theorem 2.1 essentially says that weak differentiability implies strong local Lipschitz continuity we aim, in this chapter, to extend this result to product measures and to establish bounds on perturbations for performance measures of stochastic systems by means of Lipschitz constants which can be easily derived from the expression of weak derivatives.

The setup of this chapter is as follows: let $\mu_{i,\theta}$, for $1 \leq i \leq n$, be a family of probability measures depending on some parameter $\theta \in \Theta$ and set

$$P_g(\theta) := \mathbb{E}_\theta[g(X_1, \dots, X_n)] = \int \dots \int g(s_1, \dots, s_n)\Pi_\theta(ds_1, \dots, ds_n), \tag{3.2}$$

for a cost-function g, where X_i is distributed according to $\mu_{i,\theta}$, for each $1 \leq i \leq n$, and Π_θ denotes the product of the measures $\mu_{i,\theta}$, for $1 \leq i \leq n$; for a formal definition see (2.27). Throughout this chapter we study the following type of bounds on perturbations:

(i) Bounds on $|P_g(\theta_2) - P_g(\theta_1)|$, for some specified cost-function g.

(ii) Uniform (strong) bounds with respect to $[\mathcal{D}]_v$, i.e., for

$$\sup_{\|g\|_v \leq 1} |P_g(\theta_2) - P_g(\theta_1)|,$$

for some $\theta_1, \theta_2 \in \Theta$.

Starting point of the analysis put forward in this chapter is Theorem 2.1 which asserts that weak $[\mathcal{D}]_v$-differentiability of a measure-valued mapping implies strong local Lipschitz continuity, i.e., for each neighborhood V of 0, there exists some constant $M > 0$ such that

$$\forall \xi \in V,\ g \in [\mathcal{D}]_v : \left| \int g(s)\mu_{\theta+\xi}(ds) - \int g(s)\mu_\theta(ds) \right| \leq M|\xi|\|g\|_v. \tag{3.3}$$

A constant M, satisfying (3.3), is called a Lipschitz constant (bound) for μ_θ. Note that any $M' > M$ is a Lipschitz bound for μ_θ, provided that M is. Therefore, one can increase the effectiveness of a bound by decreasing it, when possible. Although the Lipschitz bounds are, in general, not very effective they still play an important role when studying strong stability of stochastic systems. In other words, they are qualitatively important rather than quantitatively. Extending this result to product measures leads to the desired bounds, provided that the input distributions $\mu_{i,\theta}$ are weakly differentiable, for $1 \leq i \leq n$.

The chapter is organized as follows: In Section 3.2 we derive Lipschitz bounds for some standard probabilistic models and in Section 3.3 we extend our analysis to the steady-state waiting time. In particular, we show that the stationary distribution of waiting times in the G/G/1 queue is strongly local Lipschitz continuous, provided that the service-times distribution is weakly differentiable.

3.2 Bounds on Perturbations

Theorem 2.1 establishes strong local Lipschitz continuity of weakly differentiable probability measures. For practical purposes one is interested in calculating an actual Lipschitz bound. Therefore, this section is intended to show how Lipschitz bounds can be derived from evaluating the weak derivative of a probability measure. While the procedure for deriving Lipschitz constants is rather similar to the one in classical analysis, the particular setting we address here imposes, however, some specific formulation and this will be explained in the main result of this section, Theorem 3.1. In Section 3.2.1 we derive bounds on perturbations for product probability measures and in Section 3.2.2 we obtain similar results for homogenous Markov chains and we illustrate the results with an application to the sequence of the waiting times in the G/G/1 queue.

3.2.1 Bounds on Perturbations for Product Measures

The aim of this section is to establish bounds on perturbations for product measures. This is useful when considering performance measures which depend on a finite collection of random variables as in (3.2). We start with a basic result which establishes bounds on perturbations for one-dimensional distributions.

Theorem 3.1. *Let* $\mu_* : \Theta \to \mathcal{M}_v^1$ *be* $[\mathcal{C}]_v$*-differentiable on* Θ*. For* $\theta_1, \theta_2 \in \Theta$*, such that* $\theta_1 < \theta_2$*, let us define*

$$\mathbf{L}_\mu^v(\theta_1, \theta_2) := \sup_{\theta \in [\theta_1, \theta_2]} \|\mu'_\theta\|_v.$$

(i) Then it holds that

$$\|\mu_{\theta_2} - \mu_{\theta_1}\|_v \leq \mathbf{L}_\mu^v(\theta_1, \theta_2)\, (\theta_2 - \theta_1). \tag{3.4}$$

(i) For any $g \in [\mathcal{F}]_v$ *it holds that*

$$|\mathbb{E}_{\theta_2}[g(X)] - \mathbb{E}_{\theta_1}[g(X)]| \leq \mathbf{L}_\mu^v(\theta_1, \theta_2)\, \|g\|_v\, (\theta_2 - \theta_1), \tag{3.5}$$

where, for $\theta \in \Theta$*,* $\mathbb{E}_\theta$ *is an expectation operator consistent with* $X \sim \mu_\theta$*.*

(iii) If $\mu'_\theta = (c_\theta, \mu_\theta^+, \mu_\theta^-)$ *and* $g \geq 0$ *then we can replace* $\mathbf{L}_\mu^v(\theta_1, \theta_2)$ *in (3.5) by*

$$\mathbf{M}_\mu^v(\theta_1, \theta_2) = \sup_{\theta \in [\theta_1, \theta_2]} (c_\theta \cdot \max\{\|\mu_\theta^+\|_v, \|\mu_\theta^-\|_v\}).$$

Proof. (i) Fix $g \in [\mathcal{C}]_v$. Applying the Mean Value Theorem yields

$$\left| \int g(s)\, \mu_{\theta_2}(ds) - \int g(s)\, \mu_{\theta_1}(ds) \right| = (\theta_2 - \theta_1) \left| \int g(s)\, \mu'_{\theta_g}(ds) \right|,$$

for some $\theta_g \in (\theta_1, \theta_2)$, depending on g. On the other hand,

$$\forall \theta \in (\theta_1, \theta_2) : \left| \int g(s)\, \mu'_\theta(ds) \right| \leq \|g\|_v \cdot \|\mu'_\theta\|_v \leq \mathbf{L}_\mu^v(\theta_1, \theta_2)\, \|g\|_v,$$

according to Cauchy-Schwartz Inequality, and we conclude that

$$\forall g \in [\mathcal{C}]_v : \left| \int g(s)\mu_{\theta_2}(ds) - \int g(s)\mu_{\theta_1}(ds) \right| \leq \mathbf{L}_\mu^v(\theta_1, \theta_2)\, \|g\|_v\, (\theta_2 - \theta_1). \tag{3.6}$$

Taking in (3.6) the supremum with respect to $\|g\|_v \leq 1$, concludes (i).

(ii) Applying again the Cauchy-Schwarz Inequality we obtain

$$\forall g \in [\mathcal{F}]_v : \left| \int g(s)\mu_{\theta_2}(ds) - \int g(s)\mu_{\theta_1}(ds) \right| \leq \|g\|_v \|\mu_{\theta_2} - \mu_{\theta_1}\|_v$$

and from (3.4) we conclude (ii).

(iii) Finally, for $g \geq 0$ we have

$$\begin{aligned} \left| \int g(s)\mu'_\theta(ds) \right| &= c_\theta \left| \int g(s)\mu_\theta^+(ds) - \int g(s)\mu_\theta^-(ds) \right| \\ &\leq c_\theta \cdot \max\left\{ \int g(s)\mu_\theta^+(ds), \int g(s)\mu_\theta^-(ds) \right\} \\ &\leq c_\theta \cdot \max\{\|\mu_\theta^+\|_v, \|\mu_\theta^-\|_v\} \|g\|_v \end{aligned}$$

which, together with (3.5), concludes the proof of (iii). □

Lipschitz Bounds for Some Usual Distributions

In applications one is often interested in bounds of moments of certain random variables. The following example illustrates how Theorem 3.1 applies to two usual types of distributions.

Example 3.1. *Let $\mathbb{S} = [0, \infty)$ and $v(s) = s^p$, for each $s \geq 0$ and some $p \geq 0$.*

(i) Let μ_θ denote the exponential distribution with rate θ discussed in Example 2.5. Standard calculations show that

$$\begin{aligned} \|\mu'_\theta\|_v &= \int_0^\infty x^p |1 - \theta x| e^{-\theta x} dx \\ &= \frac{2e^{-1} + p\overline{\gamma}(p+1,1) - p\underline{\gamma}(p+1,1)}{\theta^{p+1}}, \end{aligned}$$

where $\overline{\gamma}$ and $\underline{\gamma}$ denote the superior (resp. inferior) incomplete Gamma functions, which are defined as follows

$$\forall p > 0, x \geq 0 : \ \overline{\gamma}(p,x) = \int_x^\infty s^{p-1} e^{-s} ds, \ \underline{\gamma}(p,x) = \int_0^x s^{p-1} e^{-s} ds.$$

Therefore, the Lipschitz constant $\mathbf{L}^v_\mu(\theta_1, \theta_2)$ in Theorem 3.1 is given by

$$\mathbf{L}^v_\mu(\theta_1,\theta_2) = \frac{2e^{-1} + p\overline{\gamma}(p+1,1) - p\underline{\gamma}(p+1,1)}{\theta_1^{p+1}}$$

and the constant $\mathbf{M}^v_\mu(\theta_1, \theta_2)$ satisfies

$$\begin{aligned} \mathbf{M}^v_\mu(\theta_1,\theta_2) &= \sup_{\theta \in [\theta_1,\theta_2]} \max\left\{ \frac{e^{-1} - p\underline{\gamma}(p+1,1)}{\theta^{p+1}}, \frac{e^{-1} + p\overline{\gamma}(p+1,1)}{\theta^{p+1}} \right\} \\ &= \frac{e^{-1} + p\overline{\gamma}(p+1,1)}{\theta_1^{p+1}}. \end{aligned}$$

In particular, if $p \geq 0$ is an integer it holds that

$$\mathbf{L}^v_\mu(\theta_1,\theta_2) = \frac{2(1 + pp!\sum_{k=0}^p (1/k!)) - pp!e}{\theta_1^{p+1} e}$$

and

$$\mathbf{M}^v_\mu(\theta_1,\theta_2) = \frac{1 + pp!\sum_{k=0}^p (1/k!)}{\theta_1^{p+1} e}.$$

(ii) For the uniform distribution ψ_θ on $[0, \theta)$, in accordance with Example 2.6, we obtain the following Lipschitz constants:

$$\mathbf{L}^v_\psi(\theta_1,\theta_2) = \frac{p+2}{p+1}\theta_2^{p-1}$$

and

$$\mathbf{M}^v_\psi(\theta_1,\theta_2) = \theta_2^{p-1}.$$

Example 3.1 illustrates the fact that the Lipschitz constants very often depend on the values $\theta_1, \theta_2 \in \Theta$. Thus, from this point of view, notations such as $\mathbf{L}^v_\mu(\theta_1, \theta_2)$ and $\mathbf{M}^v_\mu(\theta_1, \theta_2)$ are justified. However, in what follows we omit specifying the values θ_1, θ_2 when not relevant.

Extension to Product Measures

Let us consider now (a) a finite family $\{\mu_{i,\theta} : \theta \in \Theta\} \subset \mathcal{M}^1(\mathbb{S}_i)$ of probability measures (b) a family of non-negative, continuous mappings $v_i \in \mathcal{C}^+(\mathbb{S}_i)$ and (c) recall the definitions of Π_θ, $\vec{v}$ and $P_g(\theta)$ given in (2.27), (2.28) and (3.2), respectively.

Theorem 3.2. *Let $\mu_{i,*} : \Theta \to \mathcal{M}^1_{v_i}$ be $[\mathcal{C}(\mathbb{S}_i)]_{v_i}$-differentiable on Θ, for each $1 \leq i \leq n$, and for arbitrary $\theta_1, \theta_2 \in \Theta$, such that $\theta_1 < \theta_2$ set*

$$\forall 1 \leq i \leq n : \ \mathbf{L}^i = \sup_{\theta \in [\theta_1, \theta_2]} \|\mu'_{\theta,i}\|_{v_i}.$$

(i) Then it holds that

$$\|\Pi_{\theta_2} - \Pi_{\theta_1}\|_{\vec{v}} \leq \mathbf{L}^*(\theta_2 - \theta_1),$$

where

$$\mathbf{L}^* = \sum_{i=1}^{n} \left(\mathbf{L}^i \prod_{j=1}^{i-1} \|\mu_{j,\theta_2}\|_{v_j} \prod_{k=i+1}^{n} \|\mu_{k,\theta_1}\|_{v_k} \right) \tag{3.7}$$

and we agree that a void product, such as $\prod_{j=1}^{0} \|\mu_{j,\theta_2}\|_{v_j}$, equal to 1.

(ii) For each $g \in [\mathcal{F}(\mathbb{S}_1 \times \ldots \times \mathbb{S}_n)]_{\vec{v}}$ it holds that

$$|P_g(\theta_2) - P_g(\theta_1)| \leq \mathbf{L}^* \|g\|_{\vec{v}} (\theta_2 - \theta_1). \tag{3.8}$$

(iii) If $g \geq 0$ and $\mu'_{i,\theta} = (c_{i,\theta}, \mu^+_{i,\theta}, \mu^-_{i,\theta})$, for $1 \leq i \leq n$, then the constant $\mathbf{L}^*$ in (3.8) can be improved by replacing in (3.7) $\mathbf{L}^i$ by

$$\mathbf{M}^i = \sup_{\theta \in [\theta_1, \theta_2]} (c_{i,\theta} \cdot \max\{\|\mu^+_{i,\theta}\|_v, \|\mu^-_{i,\theta}\|_v\}),$$

i.e., $\mathbf{L}^*$ can be replaced by

$$\mathbf{M}^* := \sum_{i=1}^{n} \left(\mathbf{M}^i \prod_{j=1}^{i-1} \|\mu_{j,\theta_2}\|_{v_j} \prod_{k=i+1}^{n} \|\mu_{k,\theta_1}\|_{v_k} \right). \tag{3.9}$$

Proof. For arbitrary $g \in [\mathcal{C}(\mathbb{S}_1 \times \ldots \times \mathbb{S}_n)]_{\vec{v}}$, we have

$$\forall (s_1, \ldots, s_n) : \ |g(s_1, \ldots, s_n)| \leq \|g\|_{\vec{v}} \cdot \vec{v}(s_1, \ldots, s_n) = \|g\|_{\vec{v}} \cdot v(s_1) \cdot \ldots \cdot v(s_n).$$

Therefore, for each $1 \leq i \leq n$, the mapping g_i defined as

$$g_i(s_i) = \int \ldots \int g(s_1, \ldots, s_n) \prod_{j=1}^{i-1} \mu_{j,\theta_2}(ds_j) \prod_{k=i+1}^{n} \mu_{k,\theta_1}(ds_k)$$

is continuous (for a proof, use the Dominated Convergence Theorem) and satisfies (apply Fubini Theorem)

$$\|g_i\|_{v_i} \leq \|g\|_{\vec{v}} \cdot \prod_{j=1}^{i-1} \|\mu_{i,\theta_2}\|_{v_j} \cdot \prod_{k=i+1}^{n} \|\mu_{k,\theta_1}\|_{v_k}. \tag{3.10}$$

Therefore, $g_i \in [\mathcal{C}(\mathbb{S}_i)]_{v_i}$, for each $1 \leq i \leq n$, and since $\mu_{i,\theta}$ is weakly $[\mathcal{C}(\mathbb{S}_i)]_{v_i}$-differentiable, we conclude from Theorem 3.1 (*ii*) that

$$\left| \int g_i(s_i)\mu_{i,\theta_2}(ds_i) - \int g_i(s_i)\mu_{i,\theta_1}(ds_i) \right| \leq \mathbf{L}^i \|g_i\|_{v_i}(\theta_2 - \theta_1). \tag{3.11}$$

On the other hand, simple algebraic calculations show that

$$\int g(s)\Pi_{\theta_2}(ds) - \int g(s)\Pi_{\theta_1}(ds) = \sum_{i=1}^{n} \left(\int g_i(s_i)\mu_{i,\theta_2}(ds_i) - \int g_i(s_i)\mu_{i,\theta_1}(ds_i) \right).$$

Hence, (3.11) together with (3.10) imply that

$$\forall g \in [\mathcal{C}(\mathbb{S}_1 \times \ldots \times \mathbb{S}_n)]_{\vec{v}} : \left| \int g(s)\Pi_{\theta_2}(ds) - \int g(s)\Pi_{\theta_1}(ds) \right| \leq \mathbf{L}^* \|g\|_{\vec{v}}(\theta_2 - \theta_1)$$

holds true, for $\mathbf{L}^*$ defined by (3.7). Taking in the above inequality the supremum with respect to $\|g\|_{\vec{v}} \leq 1$ concludes (*i*). A similar reasoning as in Theorem 3.1 concludes the proofs of (*ii*) and (*iii*). □

Remark 3.1. *If $\mu_{i,\theta}$ is $[\mathcal{C}]_{v_i}$-differentiable, having derivative $\mu'_{i,\theta}$, we conclude from Theorem 2.4 that $P_g(\theta)$ is differentiable with respect to θ, for $g \in \mathcal{C}_{\vec{v}}$. Therefore, one could also derive a Lipschitz bound for $P_g(\theta)$ by bounding the derivative $P'_g(\theta)$ which, according to Theorem 3.1, satisfies*

$$P'_g(\theta) = \sum_{i=1}^{n} \int \ldots \int g(s_1, \ldots, s_n)\mu'_{i,\theta}(ds_i) \prod_{j \neq i} \mu_{j,\theta}(ds_j).$$

Using a similar reasoning as in Theorem 3.2 one would obtain in (3.8) the following Lipschitz bound:

$$\mathbf{L}' = \sum_{i=1}^{n} \sup_{\theta \in [\theta_1, \theta_2]} \left(\|\mu'_{i,\theta}\|_{v_i} \prod_{j \neq i} \|\mu_{j,\theta}\|_{v_j} \right)$$

which, in general, is less accurate (larger) than $\mathbf{L}^*$ defined in (3.7).

Corollary 3.1. *Under the conditions put forward in Theorem 3.2, if for each $1 \leq i \leq n$ $\mu_{i,\theta} = \mu_\theta$ and $v_i = v$ then*

$$\mathbf{L}^* = \mathbf{L}_v^\mu \sum_{k=1}^{n} \|\mu_{\theta_2}\|_v^{k-1} \|\mu_{\theta_1}\|_v^{n-k}, \quad \mathbf{M}^* = \mathbf{M}_v^\mu \sum_{k=1}^{n} \|\mu_{\theta_2}\|_v^{k-1} \|\mu_{\theta_1}\|_v^{n-k}.$$

A Simple Application from Finance

Let us consider the following simple example from finance. Assume that an investor is purchasing $S > 0$ units worth of stock each month for a number $n \geq 1$ months in a row. If we denote by X_k the spot price per share in month k, for $1 \leq k \leq n$, then the amount purchased in month k equals to S/X_k. Hence, the average price per share X_a he or she pays over the n months is obtained by dividing the total amount of wealth spent divided by the total number of shares purchased; in formula:

$$X_a = \frac{n \cdot S}{\frac{S}{X_1} + \ldots + \frac{S}{X_n}} = \frac{n}{\frac{1}{X_1} + \ldots + \frac{1}{X_n}},$$

i.e., the average price is just the harmonic mean of the spot prices $X_1, \ldots, X_n$.

Let us fix $n \geq 1$ and assume that $\{X_i : 1 \leq i \leq n\}$ are i.i.d. random variables with distribution μ_θ depending on some parameter θ. One is interested in studying the sensitivity of the expected average price with respect to θ, i.e., to obtain a bound for the perturbation

$$\Delta_p(\theta_1, \theta_2) := |\mathbb{E}_{\theta_2}[X_a] - \mathbb{E}_{\theta_1}[X_a]|,$$

for some $\theta_1, \theta_2 \in \Theta$, such that $\theta_1 < \theta_2$. To this end, note that the expected average price can be written as

$$\mathbb{E}_\theta[X_a] = P_h(\theta) = \int \ldots \int h(x_1, \ldots, x_n)\mu_\theta(dx_1) \ldots \mu_\theta(dx_n), \tag{3.12}$$

where h is defined as

$$\forall x_1, \ldots, x_n > 0 : \ h(x_1, \ldots, x_n) = \frac{n}{\frac{1}{x_1} + \ldots + \frac{1}{x_n}}.$$

Letting $v \in \mathcal{C}^+((0, \infty))$, $v(x) = \sqrt[n]{x}$, it holds that

$$\forall x_1, \ldots, x_n > 0 : \ h(x_1, \ldots, x_n) \leq \sqrt[n]{x_1 \cdot \ldots \cdot x_n} = \vec{v}(x_1, \ldots, x_n).$$

Therefore, since $h \geq 0$, by Theorem 3.2 (*iii*), one concludes that $P_h(\theta)$ is Lipschitz continuous with respect to θ, provided that μ_θ is $[\mathcal{C}]_v$-differentiable on Θ and by Corollary 3.1, it follows that a Lipschitz bound is given by

$$\mathbf{M}^* = \mathbf{M}_v^\mu \sum_{k=1}^{n} \|\mu_{\theta_2}\|_v^{k-1} \|\mu_{\theta_1}\|_v^{n-k}. \tag{3.13}$$

More specifically, noting that $\|h\|_{\vec{v}} = 1$, we conclude that

$$\Delta_p(\theta_1, \theta_2) \leq \mathbf{M}^*(\theta_2 - \theta_1).$$

Example 3.2. *If, for instance, μ_θ is the exponential distribution with rate θ (introduced in Example 2.5) then we have*

$$\forall \theta \in \Theta : \ \|\mu_\theta\|_v = \int_0^\infty \sqrt[n]{x} \theta e^{-\theta x} dx = \frac{1}{\sqrt[n]{\theta}} \Gamma\left(\frac{n+1}{n}\right),$$

where Γ denotes the usual Gamma function. Therefore, in accordance with Example 3.1, we obtain the following Lipschitz bound in (3.13):

$$\mathbf{M}^* = \frac{\frac{1}{e} + \frac{1}{n}\overline{\gamma}\left(\frac{n+1}{n}, 1\right)}{\sqrt[n]{\theta_1^{n+1}}} \cdot \frac{\theta_2 - \theta_1}{\sqrt[n]{\theta_2} - \sqrt[n]{\theta_1}} \cdot \left(\frac{\Gamma\left(\frac{n+1}{n}\right)}{\sqrt[n]{\theta_1 \theta_2}}\right)^{n-1}.$$

3.2.2 Bounds on Perturbations for Markov Chains

Throughout this section we aim to derive bounds on perturbations for Markov chains. For practical reasons we consider homogenous Markov chains which are generated by transition kernels. Eventually, we illustrate the results with an application to the sequence of waiting times in the G/G/1 queue. To this end, we briefly present the connection between Markov chains and Markov operators and show how the concept of weak differentiation extends to transition kernels, providing the means of deriving Lipschitz bounds.

Markov Chains Generated By Markov Operators

Recall that a *transition kernel* on $\mathbb{S}$ is a mapping $Q : \mathbb{S} \times \mathcal{S} \to \mathbb{R}$ satisfying

(i) $\forall A \in \mathcal{S}$, the mapping $Q(\cdot, A)$ is measurable,

(ii) $\forall s \in \mathbb{S}$, $Q(s, \cdot) \in \mathcal{M}(\mathcal{S})$.

If $Q(s, \cdot) \in \mathcal{M}^1$, for all $s \in \mathbb{S}$, we call Q a *Markov operator*. A transition kernel Q can be identified with a linear operator (denoted also by Q) on the set of measurable mappings on $\mathbb{S}$ defined as

$$\forall s \in \mathbb{S} : \ (Qf)(s) = \int f(x) Q(s, dx),$$

for all measurable f for which the right-hand side integral makes sense. Note that one can recover the transition kernel Q from the operator Q, as follows:

$$\forall s \in \mathbb{S}, A \in \mathcal{S} : \ Q(s, A) = (Q\, \mathbb{I}_A)(s).$$

For $v \in \mathcal{C}^+$ we introduce the v-norm of Q as follows:

$$\|Q\|_v = \sup_{\|f\|_v \le 1} \|Qf\|_v = \sup_{\|f\|_v \le 1} \sup_{s \in \mathbb{S}} \frac{|(Qf)(s)|}{v(s)}, \tag{3.14}$$

where the above supremum is taken with respect to $f \in \mathcal{D}$. Note that, we have

$$\forall f \in [\mathcal{F}]_v : \ \|Qf\|_v \le \|Q\|_v \cdot \|f\|_v. \tag{3.15}$$

In particular, if $\|Q\|_v < \infty$, then Q maps $[\mathcal{F}]_v$ onto itself, i.e.,

$$f \in [\mathcal{F}]_v \Rightarrow Qf \in [\mathcal{F}]_v,$$

and note that, in general, such an implication does not hold true for $\mathcal{C}$.

Remark 3.2. *In general, determining the v-norm $\|Q\|_v$ of a transition kernel Q is not an easy task since a similar method as in the case of measures is not appropriate. For instance, unlike in the measures case, there is no straightforward way to show that the supremum in (3.14) is attained for a particular f. Consequently, the value $\|Q\|_v$ may depend on the choice of $\mathcal{D}$.*

However, if Q defines a positive operator (for instance, Q is a Markov operator), i.e.,

$$\forall s \in \mathbb{S} : \ Q(s, \cdot) \in \mathcal{M}^+(\mathbb{S}),$$

by monotonicity of the integral we obtain

$$\|Q\|_v = \sup_{\|f\|_v \le 1} \left\| \int f(x)Q(\cdot, dx) \right\|_v = \left\| \int v(x)Q(\cdot, dx) \right\|_v = \sup_{s \in \mathbb{S}} \frac{\int v(x)Q(s, dx)}{v(s)}.$$

For general Q we can only show that the v-norm is bounded, i.e.,

$$\|Q\|_v \le \sup_{s \in \mathbb{S}} \frac{\int v(x)|Q(s, dx)|}{v(s)},$$

where we denote by $|Q(s, \cdot)|$ the variation of the measure $Q(s, \cdot)$.

If Q_1, Q_2 are transition kernels on $\mathbb{S}$ we define the composition $Q_2 Q_1$, as follows:

$$\forall s \in \mathbb{S}, \forall A \in \mathcal{S} : \ (Q_2 Q_1)(s, A) = \int Q_2(x, A) Q_1(s, dx).$$

It is immediate that $Q_2 Q_1$ is itself a transition kernel on $\mathbb{S}$, if Q_1, Q_2 are Markov operators so it is their composition $Q_2 Q_1$ and the induced operator $Q_2 Q_1$ is given by

$$(Q_2 Q_1 f)(s) = \int f(y)(Q_2 Q_1)(s, dy) = \int \int f(y) Q_2(x, dy) Q_1(s, dx)$$

One can easily check that $Q_2 Q_1 f = Q_2(Q_1 f)$ and according to (3.15) we have

$$\forall f \in \mathcal{F} : \ \|Q_2 Q_1 f\|_v \le \|Q_2\|_v \cdot \|Q_1 f\|_v \le \|Q_2\|_v \|Q_1\|_v \|f\|_v.$$

Taking in the above inequality the supremum with respect to $\|f\|_v \le 1$ yields

$$\|Q_2 Q_1\|_v \le \|Q_2\|_v \cdot \|Q_1\|_v.$$

Moreover, one can iterate the composition of kernels. By convention, for an arbitrary transition kernel Q we define Q^0 as the identity operator[1] and for $n \ge 1$ we define the "n^{th} power" of Q as $Q^n := Q_n \dots Q_1$, where $Q_i = Q$, for each $1 \le i \le n$. Then it holds that $\|Q^n\|_v \le \|Q\|_v^n$, for each $n \ge 0$.

We say that the Markov chain $\{Z_n : n \ge 0\}$ is generated by the Markov operator Q if for all $n \ge 0$ and all measurable f it holds that

$$\mathbb{E}\left[f(Z_{n+1}) \,|Z_n\right] = (Qf)(Z_n),$$

where the expression on the left-hand side denotes the conditional expectation of $f(Z_{n+1})$ with respect to Z_n. It turns out that, from a probabilistic point of view, the Markov chain $\{Z_n : n \ge 0\}$ is completely determined by the operator Q and the distribution of Z_0, which will be called the *initial distribution* and denoted by χ^0. Indeed, one can show inductively that for all $n \ge 0$ and measurable f we have

$$\mathbb{E}\left[f(Z_n) \,|Z_0\right] = (Q^n f)(Z_0),$$

[1] The identity operator corresponds to Dirac transition kernel $\mathbf{1}(x, A) = \delta_x(A)$. It follows that $\mathbf{1}f = f$, for all measurable f and $\|\mathbf{1}\|_v = 1$, for any $v \in \mathcal{C}^+$.

which, by integration with respect to the initial distribution, yields

$$\mathbb{E}\left[f(Z_n)\right] = \mathbb{E}\left[(Q^n f)(Z_0)\right] = \int (Q^n f)(s)\chi^0(ds). \tag{3.16}$$

Therefore, if for $n \geq 0$ we denote by χ^n the distribution of Z_n, it follows that

$$\forall n \geq 0, A \in \mathcal{S}:\ \mathbb{P}\{Z_n \in A\} = \chi^n(A) = \int (Q^n \mathbb{I}_A)(s)\chi^0(ds). \tag{3.17}$$

Example 3.3. *Recall the G/G/1 queue described in Section 2.5.2. From Lindley's recursion we conclude that, for all measurable f, it holds that*

$$\mathbb{E}\left[f(W_{n+1})|W_n\right] = \mathbb{E}\left[f\left(\max\{W_n + S_n - T_{n+1}, 0\}\right)|W_n\right],$$

i.e., the Markov operator generating the sequence of waiting times satisfies

$$\forall x \geq 0, A \in \mathcal{S}:\ Q(x, A) = \int\int \mathbb{I}_A((x + s - t)_+)\eta(dt)\mu(ds),$$

or in functional operator form

$$\forall x \geq 0:\ (Qf)(x) = \int\int f((x + s - t)_+)\eta(dt)\mu(ds) = \mathbb{E}\left[f\left((x + S - T)_+\right)\right], \tag{3.18}$$

where S and T are independent random variables distributed according to μ and η, respectively. Indeed, one can check using Lemma E.2 (see the Appendix) that

$$\forall n \geq 1:\ \mathbb{E}\left[f(W_{n+1})|W_n\right] = (Qf)(W_n),$$

for each measurable f for which $\mathbb{E}\left[f(W_{n+1})\right]$ exists.

Furthermore, let $v(x) = e^{\alpha x}$, for some $\alpha \geq 0$. Then, we have

$$\|Q\|_v = \sup_{x\geq 0} e^{-\alpha x} \cdot \mathbb{E}\left[e^{\alpha(x+S-T)_+}\right] = \sup_{x\geq 0} \mathbb{E}\left[e^{\alpha[(x+S-T)_+ - x]}\right];$$

see Remark 3.2. Since the mapping $x \mapsto \alpha\left[(x + S - T)_+ - x\right]$ is non-increasing on $[0,\infty)$, a simple application of the Dominated Convergence Theorem yields

$$\|Q\|_v = \mathbb{E}\left[e^{\alpha \sup\limits_{x\geq 0}[(x+S-T)_+ - x]}\right] = \mathbb{E}\left[e^{\alpha(S-T)_+}\right],$$

provided that the right-hand side expectation above is finite.

Let $\{Q_\theta : \theta \in \Theta\}$ be a family of Markov operators on $\mathbb{S}$, for some $\Theta \subset \mathbb{R}$. We say that Q_θ is weakly $[\mathcal{D}]_v$-differentiable if there exist a transition kernel Q'_θ, such that

$$\forall s \in \mathbb{S}, f \in [\mathcal{D}]_v:\ \frac{d}{d\theta}(Q_\theta f)(s) = (Q'_\theta f)(s).$$

Let $\mathbb{E}_\theta$ be an expectation operator under which $\{Z_n : n \geq 0\}$ is a Markov chain generated by the Markov operator Q_θ, i.e., $\{Z_n : n \geq 0\}$ is a Markov chain satisfying $Z_n \sim \chi^n_\theta$, for

$n \geq 0$, where χ_θ^n is defined as in (3.17), for $Q = Q_\theta$. In addition, we assume that the initial distribution χ^0 is independent of θ, i.e., the expression

$$\mathbb{E}_\theta\left[f(Z_0)\right] = \int f(s)\chi^0(ds)$$

is constant[2] with respect to θ for any measurable f.

In the following we address the following problem: we aim to establish bounds for the expression

$$\Delta_{n,f}^{Z_0}(\theta_1,\theta_2) := \left|\mathbb{E}_{\theta_2}\left[f(Z_n)\right] - \mathbb{E}_{\theta_1}\left[f(Z_n)\right]\right|, \tag{3.19}$$

for arbitrary $\theta_1, \theta_2 \in \Theta$, such that $\theta_1 < \theta_2$, $n \geq 1$ and f for which the right-hand side is finite. The following result provides a bound for $\Delta_{n,f}^{Z_0}(\theta_1,\theta_2)$, assuming weak differentiability of Q_θ.

Theorem 3.3. *Let $\{Q_\theta : \theta \in \Theta\}$ be a family of Markov operators on $\mathbb{S}$, for some $\Theta \subset \mathbb{R}$, such that Q_θ is weakly $[\mathcal{D}]_v$-differentiable for each $\theta \in \Theta$. Then, if $\{Z_n : n \geq 0\}$ is a Markov chain generated by the operator Q_θ, it holds that*

$$\forall f \in [\mathcal{D}]_v : \ \Delta_{n,f}^{Z_0}(\theta_1,\theta_2) \leq \mathbf{C}_n \mathbf{L} \|f\|_v (\theta_2 - \theta_1)\mathbb{E}\left[v(Z_0)\right], \tag{3.20}$$

where[3]

$$\mathbf{C}_n = \sum_{k=1}^{n} \|Q_{\theta_2}\|_v^{n-k} \|Q_{\theta_1}\|_v^{k-1}, \quad \mathbf{L} = \sup_{\theta \in [\theta_1,\theta_2]} \|Q'_\theta\|_v.$$

Proof. Taking (3.16) into account we conclude that

$$\Delta_{n,f}^{Z_0}(\theta_1,\theta_2) = \left|\mathbb{E}\left[(Q_{\theta_2}^n f)(Z_0)\right] - \mathbb{E}\left[(Q_{\theta_1}^n f)(Z_0)\right]\right|.$$

Consequently, the expression in (3.19) can be bounded as follows:

$$\begin{aligned}\Delta_{n,f}^{Z_0}(\theta_1,\theta_2) &\leq \mathbb{E}\left[\left|(Q_{\theta_2}^n f)(Z_0) - (Q_{\theta_1}^n f)(Z_0)\right|\right] \\ &\leq \|Q_{\theta_2}^n - Q_{\theta_1}^n\|_v \cdot \|f\|_v \cdot \mathbb{E}\left[v(Z_0)\right]. \end{aligned} \tag{3.21}$$

Elementary algebraic calculations show that

$$Q_{\theta_2}^n - Q_{\theta_1}^n = \sum_{k=1}^{n} Q_{\theta_2}^{n-k}(Q_{\theta_2} - Q_{\theta_1})Q_{\theta_1}^{k-1}.$$

Hence, using standard properties of operator norms, we arrive at

$$\|Q_{\theta_2}^n - Q_{\theta_1}^n\|_v \leq \sum_{k=1}^{n} \|Q_{\theta_2}\|_v^{n-k} \|Q_{\theta_2} - Q_{\theta_1}\|_v \|Q_{\theta_1}\|_v^{k-1}. \tag{3.22}$$

Since Q_θ is $[\mathcal{D}]_v$-differentiable on Θ, one can apply the Mean Value Theorem to the mapping $\theta \mapsto (Q_\theta g)(x)$, which yields

$$\forall x \in \mathbb{S}, g \in [\mathcal{D}]_v : \ (Q_{\theta_2} g)(x) - (Q_{\theta_1} g)(x) = (\theta_2 - \theta_1) \cdot (Q'_\theta g)(x),$$

[2] To illustrate this, we omit the subscript θ, writing $\mathbb{E}\left[f(Z_0)\right]$ instead.
[3] It is not crucial to assume that $\mathbf{L} < \infty$ since (3.20) is obviously satisfied by $\mathbf{L} = \infty$.

for some $\theta \in (\theta_1, \theta_2)$ depending on g and x. Thus, for $\|g\|_v \leq 1$ we have

$$\|(Q_{\theta_2} g) - (Q_{\theta_1} g)\|_v \leq (\theta_2 - \theta_1)\|Q'_\theta\|_v \leq (\theta_2 - \theta_1) \sup_{\theta \in [\theta_1, \theta_2]} \|Q'_\theta\|_v .$$

Therefore, taking the supremum with respect to $\|g\|_v \leq 1$ yields

$$\|Q_{\theta_2} - Q_{\theta_1}\|_v \leq (\theta_2 - \theta_1) \sup_{\theta \in [\theta_1, \theta_2]} \|Q'_\theta\|_v ,$$

which, together with (3.21) and (3.22), concludes the proof. □

Application to the Transient Waiting Time

Let us consider the G/G/1 queue as introduced in Section 2.7 and let us assume that the service time distribution μ_θ depends on some design parameter $\theta \in \Theta$. Recall from Example 3.3 that the corresponding sequence of waiting times is generated by the operator Q_θ, defined as in (3.18), by letting $\mu = \mu_\theta$. More specifically, we let $\mathbb{S} = [0, \infty)$, consider $\{\mu_\theta : \theta \in \Theta\} \subset \mathcal{M}^1(\mathcal{S})$ and denote by Q_θ the Markov operator defined as

$$\forall x \geq 0, f \in \mathcal{F} : \; (Q_\theta f)(x) = \int \int f\left((x + s - t)_+\right) \eta(dt) \mu_\theta(ds). \tag{3.23}$$

for all $\theta \in \Theta$. Let $\theta_1, \theta_2 \in \Theta$ be such that $\theta_1 < \theta_2$. Using Theorem 3.3, we aim to establish bounds for the expression

$$\Delta^x_{n,f}(\theta_1, \theta_2) = \left| \mathbb{E}^x_{\theta_2}\left[f(W_{n+1})\right] - \mathbb{E}^x_{\theta_1}\left[f(W_{n+1})\right] \right|, \tag{3.24}$$

for arbitrary $n \geq 1$, $x \geq 0$ and $f \in [\mathcal{C}]_v$, where $\mathbb{E}^x_\theta$ denotes the expectation operator, when $W_1 = x$, and the service times follow distribution μ_θ. To do so, let us consider the family of Markov operators $\{Q_\theta : \theta \in \Theta\}$ introduced in (3.23) and let $Z_n := W_{n+1}$, for each $n \geq 1$, i.e., $\chi_0 = \delta_x$, in Theorem 3.3.

To apply Theorem 3.3 one has to investigate weak differentiability of Q_θ. A formal differentiation of Q_θ, in (3.23), with respect to θ yields

$$\frac{d}{d\theta}(Q_\theta f)(x) = \int \int f\left((x + s - t)_+\right) \eta(dt) \mu'_\theta(ds). \tag{3.25}$$

It turns out that weak differentiability of Q_θ is related to that of μ_θ. This relation will be established by our next result. Specifically, we present a class of mappings v for which $\mathcal{C}_v$-differentiability of μ_θ implies that of Q_θ.

Lemma 3.1. *Let $(\mathcal{D}, v)$ be a Banach base on $\mathbb{S}$, $\mu_* : \Theta \to \mathcal{M}^1_v$ and let Q_θ be defined as in (3.23). If:*

(i) μ_θ is weakly $[\mathcal{D}]_v$-differentiable,

(ii) for each $x \geq 0$ the mapping v satisfies

$$\sup_{s \geq 0} \frac{\int v\left((x + s - t)_+\right) \eta(dt)}{v(s)} < \infty,$$

then Q_θ is weakly $[\mathcal{D}]_v$-differentiable and (3.25) holds true, i.e.,

$$(Q'_\theta f)(x) = \int\int f\left((x+s-t)_+\right)\eta(dt)\mu'_\theta(ds).$$

Proof. For $s, x \geq 0$ and $f \in \mathcal{F}$ let

$$H_f(s,x) = \int f\left((x+s-t)_+\right)\eta(dt).$$

From (i) we conclude that it suffices to show that $H_f(\cdot, x) \in [\mathcal{D}]_v$, for all $f \in [\mathcal{D}]_v$ and $x \geq 0$. Indeed, differentiating (3.23) with respect to θ yields

$$\forall x \geq 0, f \in \mathcal{C}_v: \; \frac{d}{d\theta}(Q_\theta f)(x) = \frac{d}{d\theta}\int H_f(s,x)\mu_\theta(ds) \stackrel{(i)}{=} \int H_f(s,x)\mu'_\theta(ds),$$

which concludes (3.25).

Condition (ii) essentially says that $\|H_f(\cdot,x)\|_v < \infty$, for all $x \geq 0$, provided that $\|f\|_v < \infty$. It follows that the implication

$$\forall x \geq 0: \; f \in [\mathcal{D}]_v \Rightarrow H_f(\cdot, x) \in [\mathcal{D}]_v \tag{3.26}$$

holds true for $\mathcal{D} = \mathcal{F}$. In order to show that (3.26) holds true for $\mathcal{D} = \mathcal{C}$ as well, one has to check that for each $x \geq 0$ it holds that $H_f(\cdot, x)$ is continuous provided that f is continuous. Indeed, let us assume that f is continuous, let $s \geq 0$ be fixed and $\epsilon > 0$. Since continuity of f implies uniform continuity on each compact set (see, e.g., [53]) it follows that there exists some $\zeta_\epsilon > 0$ such that for each $s_1, s_2 \in [0, x+s+1]$ it holds that

$$|s_2 - s_1| < \zeta_\epsilon \Rightarrow |f(s_2) - f(s_1)| < \epsilon.$$

Therefore, it follows that for any $x \geq 0$ and $|r| < \min\{1, \zeta_\epsilon\}$ we have

$$\begin{aligned} |H_f(s+r,x) - H_f(s,x)| &= \left|\int f\left((x+s+r-t)_+\right) - f\left((x+s-t)_+\right)\eta(dt)\right| \\ &\leq \int \left|f\left((x+s+r-t)_+\right) - f\left((x+s-t)_+\right)\right|\eta(dt) \\ &\leq \int \epsilon\, \mathbb{I}_{[0,x+s+1]}(t)\eta(dt) = \epsilon\, \eta([0, x+s+1]), \end{aligned}$$

where we used the fact that for $t \geq x+s+1$ and $|r| < 1$ we have

$$(x+s+r-t)_+ = (x+s-t)_+ = 0.$$

Since ϵ was arbitrary chosen it follows that $H_f(\cdot, x)$ is continuous which concludes the proof. □

Let $v(x) = e^{\alpha x}$, for some $\alpha \geq 0$. Since for $s, t, x \geq 0$ it holds that

$$e^{\alpha(x+s-t)_+} \leq e^{\alpha(x+s)},$$

we obtain

$$\forall x \geq 0: \sup_{s \geq 0} \frac{\int e^{\alpha(x+s-t)_+} \eta(dt)}{e^{\alpha s}} \leq e^{\alpha x} < \infty.$$

Provided that μ_θ is weakly $\mathcal{C}_v$-differentiable, Lemma 3.1 applies and we conclude that Q_θ is weakly $\mathcal{C}_v$-differentiable, as well. Moreover, provided that $\|f\|_v \leq 1$, it holds that (see Remark 3.2)

$$\begin{aligned} \|Q'_\theta f\|_v &\leq \sup_{x \geq 0} \int \int e^{\alpha[(x+s-t)_+ - x]} |\mu'_\theta|(ds) \eta(dt) \\ &= \int \int e^{\alpha(s-t)_+} |\mu'_\theta|(ds) \eta(dt) \end{aligned}$$

Finally, taking the supremum with respect to $\|f\|_v \leq 1$, yields

$$\|Q'_\theta\|_v \leq c_\theta \mathbb{E}_\theta \left[e^{\alpha(S^+ - T)_+} + e^{\alpha(S^- - T)_+} \right], \tag{3.27}$$

where $\mathbb{E}_\theta$ is an expectation operator consistent with $(S^\pm, T) \sim \mu_\theta^\pm \times \eta$, $\mu'_\theta = (c_\theta, \mu_\theta^+, \mu_\theta^-)$.

Example 3.4. *Let us consider a M/U/1 queue where service times have uniform distribution ψ_θ, on $[0, \theta)$ and inter-arrival times are exponentially distributed with rate λ, i.e., the corresponding Markov operator Q_θ is given by*

$$\forall x \geq 0, f \in \mathcal{F}: (Q_\theta f)(x) = \frac{1}{\theta} \int_0^\theta \int_0^\infty f\left((x+s-t)_+\right) \lambda e^{-\lambda t}\, dt\, ds.$$

Then, according to Example 3.3, for $v(x) = e^{\alpha x}$, for some $\alpha \geq 0$, it holds that

$$\forall \alpha, \lambda, \theta: \|Q_\theta\|_v = \mathbb{E}_\theta \left[e^{\alpha(S-T)_+} \right] = \frac{\lambda^2[e^{\alpha\theta} - 1] + \alpha^2[1 - e^{-\lambda\theta}]}{\alpha\lambda(\alpha+\lambda)\theta}.$$

Similarly, according to (3.27) we conclude that

$$\forall \alpha, \lambda, \theta: \|Q'_\theta\|_v \leq \frac{\lambda^2[(1+\alpha\theta)e^{\alpha\theta} - 1] + \alpha^2[1 - (1+\lambda\theta)e^{-\lambda\theta}]}{\alpha\lambda(\alpha+\lambda)\theta}$$

and a bound for the perturbation in (3.24) can be obtained according to (3.20).

To illustrate the above findings we let $\alpha = \lambda$ and we obtain

$$\forall \lambda, \theta: \|Q_\theta\|_v = \frac{1}{\lambda\theta} \sinh(\lambda\theta),\ \|Q'_\theta\|_v \leq \frac{1+\lambda\theta}{\lambda\theta} \sinh(\lambda\theta),$$

where sinh *denotes the hyperbolic sine function. Consequently, we have*

$$\mathbf{C}_n = \sum_{k=1}^n \left(\frac{\sinh(\lambda\theta_2)}{\lambda\theta_2} \right)^{n-k} \left(\frac{\sinh(\lambda\theta_1)}{\lambda\theta_1} \right)^{k-1}, \quad \mathbf{L} = \frac{1+\lambda\theta_2}{\lambda\theta_2} \sinh(\lambda\theta_2),$$

where we use the fact that the function $x \mapsto \frac{1+x}{x} \sinh(x)$ is non-decreasing. Substituting the above constants in (3.20) yields the following bound for the expression in (3.24):

$$\frac{\Delta^x_{n,f}(\theta_1, \theta_2)}{(\theta_2 - \theta_1)} \leq \|f\|_v (1 + \lambda\theta_2) e^{\lambda x} \sum_{k=1}^n \left(\frac{\sinh(\lambda\theta_2)}{\lambda\theta_2} \right)^k \left(\frac{\sinh(\lambda\theta_1)}{\lambda\theta_1} \right)^{n-k}.$$

3.3 Bounds on Perturbations for the Steady-State Waiting Time

Throughout this section we extend our results on the transient waiting times in the G/G/1 queue to stationary waiting times. More specifically, we show that the stationary distribution in a G/G/1 queue governed by service time distribution μ_θ and inter-arrival times distribution η is strongly Lipschitz continuous with respect to θ, provided that μ_θ is weakly $[\mathcal{C}]_v$-differentiable for a certain class of mappings $v \in \mathcal{C}^+(\mathbb{R})$.

3.3.1 Strong Stability of the Steady-State Waiting Time

A straightforward approach would be the one presented in Theorem 3.3, in Section 3.2.2, by letting $n \to \infty$ in (3.20), provided that the sequence of waiting times $\{W_n : n \geq 1\}$ is weakly $[\mathcal{C}]_v$-convergent to the stationary waiting time W. Unfortunately, such an approach is to no avail since the constant $\mathbf{C}_n$ in (3.20) is unbounded with respect to n. This stems from the fact that

$$\forall \alpha \geq 0, \theta \in \Theta : \ \|Q_\theta\|_v = \mathbb{E}_\theta \left[e^{\alpha (S-T)_+} \right] \geq 1;$$

see Example 3.3. Therefore, a sharper approach is needed.

A first observation is that for $v(x) = e^{\alpha x}$, with $\alpha > 0$, the distribution of the $(n+1)^{st}$ waiting time is $[\mathcal{C}]_v$-differentiable, for all $n \geq 1$, provided that μ_θ is $[\mathcal{C}]_v$-differentiable; see Theorem 2.7. Moreover, as shown by (2.52), the weak derivative can be expressed by summing up differences between n pairs of parallel processes which, under the stability condition, couple in finite time, almost surely; see Figure 2.2. Therefore, intuitively, an early perturbation in the service time distribution counts less after n steps than a late perturbation, provided that the process is stable. In other words, the "magnitude" of the perturbation will decrease with respect to n and eventually will vanish as n tends to infinity. This is formalized in the following result.

Lemma 3.2. *Let $v(x) = e^{\alpha x}$, for some $\alpha \geq 0$, such that μ_θ is $[\mathcal{C}]_v$-differentiable. If $\mu'_\theta = (c_\theta, \mu^+_\theta, \mu^-_\theta)$ then for all $f \in [\mathcal{C}]_v$ and $1 \leq k \leq n$ it holds that*

$$\left| \mathbb{E}_\theta \left[f(W^{k+}_{n+1}) - f(W^{k-}_{n+1}) \right] \right| \leq 2\|f\|_v \mathbb{E}_\theta \left[v(W^{k*}_{n+1}) \cdot \mathbb{I}_{\{W^{k*}_{k+1} > 0, \ldots, W^{k*}_{n+1} > 0\}} \right], \tag{3.28}$$

where $W^{k\pm}_i$ are defined in (2.53) and $\{W^{k*}_i : i \geq 1\}$ denotes the sequence of waiting times in a modified queue, where the k^{th} service time S_k is replaced by $\max\{S^+_k, S^-_k\}$; see Section 2.5.2.

Proof. First, note that the perturbation $f(W^{k+}_{n+1}) - f(W^{k-}_{n+1})$ can only survive until the sequence $\{W^{k*}_i : i \geq k+1\}$ reaches 0. Indeed, it is immediate that $W^{k*}_i = \max\{W^{k+}_i, W^{k-}_i\}$, for all $i \geq k+1$. Consequently, if $W^{k*}_i = 0$ then $W^{k+}_j = W^{k-}_j$, for all $j \geq i$; see Figure 2.2. In formula:

$$\{f(W^{k+}_{n+1}) - f(W^{k-}_{n+1}) \neq 0\} \subset \bigcap_{i=k+1}^{n+1} \{W^{k*}_i > 0\}. \tag{3.29}$$

Furthermore, the fact that v is non-decreasing implies that

$$|f(W^{k+}_{n+1}) - f(W^{k-}_{n+1})| \leq |f(W^{k+}_{n+1})| + |f(W^{k-}_{n+1})| \leq 2\|f\|_v \cdot v(W^{k*}_{n+1}),$$

which, together with (3.29), proves the claim. □

Lemma 3.2 establishes a bound on the effect of the perturbation of the k^{th} service time distribution, at time n. In the following, we show that the right-hand side in (3.28) is bounded by a geometric sequence. To this end, we consider the following operators

$$\begin{aligned}\forall x \geq 0, f \in \mathcal{F}: \ (P_\theta f)(x) &:= \mathbb{E}_\theta\left[f(x+S-T)\cdot \mathbb{I}_{\{x+S>T\}}\right],\\ (P_\theta^* f)(x) &:= \mathbb{E}_\theta\left[f(x+S^*-T)\cdot \mathbb{I}_{\{x+S^*>T\}}\right].\end{aligned}$$

where $S^\pm$ are S-measurable samples (see Remark 2.6) of $\mu_\theta^\pm$, respectively, and we define[4] $S^* := \max\{S^+, S^-\}$. Note that P_θ is different from Q_θ in (3.23). Indeed, while Q_θ in (3.23) denotes the transition kernel generating the sequence of waiting times, P_θ denotes the corresponding *taboo kernel*, i.e., a transition kernel which avoids a certain subset of the state-space (in this case the subset is $\{0\}$). The following result will provide a bound for the effect of the perturbation of the k^{th} service time distribution in terms of $\|P_\theta\|_v$, $\|P_\theta^*\|_v$, $\|f\|_v$ and $\mathbb{E}_\theta[v(W_k)]$.

Lemma 3.3. *Under the conditions put forward in Lemma 3.2 it holds that*

$$\left|\mathbb{E}_\theta\left[f(W_{n+1}^{k+}) - f(W_{n+1}^{k-})\right]\right| \leq 2\|f\|_v \|P_\theta^*\|_v \|P_\theta\|_v^{n-k} \mathbb{E}_\theta\left[v(W_k)\right].$$

Proof. For the ease of notation, for $i \geq k+1$, we set

$$Y_i := v(W_i^{k*}) \cdot \mathbb{I}_{\{W_{k+1}^{k*}>0,\ldots,W_i^{k*}>0\}} = v(W_i^{k*}) \cdot \mathbb{I}_{\{W_{k+1}^{k*}>0\}} \cdot \ldots \cdot \mathbb{I}_{\{W_i^{k*}>0\}}.$$

Using basic properties of conditional expectations (see Section D in the Appendix) one can show that

$$\begin{aligned}\mathbb{E}_\theta\left[Y_{i+1}\middle|W_i^{k*}\right] &= \mathbb{E}_\theta\left[v(W_{i+1}^{k*})\cdot \mathbb{I}_{\{W_{i+1}^{k*}>0\}}\middle|W_i^{k*}\right]\cdot \mathbb{I}_{\{W_{k+1}^{k*}>0,\ldots,W_i^{k*}>0\}}\\ &= (P_\theta v)(W_i^{k*})\cdot \mathbb{I}_{\{W_{k+1}^{k*}>0,\ldots,W_i^{k*}>0\}}\\ &\leq \|P_\theta\|_v \cdot v(W_i^{k*})\cdot \mathbb{I}_{\{W_{k+1}^{k*}>0,\ldots,W_i^{k*}>0\}} = \|P_\theta\|_v Y_i.\end{aligned}$$

Consequently, for $n \geq k$ we have

$$\mathbb{E}_\theta\left[Y_{n+1}\middle|W_k\right] = \mathbb{E}_\theta\left[\mathbb{E}_\theta\left[Y_{n+1}\middle|W_n^{k*}\right]\middle|W_k\right] \leq \|P_\theta\|_v \mathbb{E}_\theta\left[Y_n\middle|W_k\right]$$

and it follows by finite induction that

$$\mathbb{E}_\theta\left[Y_{n+1}\middle|W_k\right] \leq \|P_\theta\|_v^{n-k}\mathbb{E}_\theta\left[Y_{k+1}\middle|W_k\right]. \tag{3.30}$$

Furthermore, we have

$$\mathbb{E}_\theta\left[Y_{k+1}\middle|W_k\right] = (P_\theta^* v)(W_k) \leq \|P_\theta^*\|_v \cdot v(W_k). \tag{3.31}$$

From (3.30) together with (3.31) one concludes that

$$\mathbb{E}_\theta\left[v(W_{n+1}^{k*})\cdot \mathbb{I}_{\{W_{k+1}^{k*}>0,\ldots,W_{n+1}^{k*}>0\}}\middle|W_k\right] \leq \|P_\theta^*\|_v \|P_\theta\|_v^{n-k} v(W_k).$$

Taking now the expectation in the above inequality, yields

$$\mathbb{E}_\theta\left[v(W_{n+1}^{k*})\cdot \mathbb{I}_{\{W_{k+1}^{k*}>0,\ldots,W_{n+1}^{k*}>0\}}\right] \leq \|P_\theta^*\|_v \|P_\theta\|_v^{n-k}\mathbb{E}_\theta\left[v(W_k)\right]. \tag{3.32}$$

Therefore, Lemma 3.2 concludes the proof. □

[4] Note that, the distribution of S^* depends on the joint distribution of the pair (S^+, S^-). While this is not directly relevant here, in some numerical applications this fact should not be overlooked.

Remark 3.3. *Recall that the G/G/1 queue is stable if*

$$\mathbb{E}_\theta[S - T] = \int s\, \mu_\theta(ds) - \int t\, \eta(dt) < 0. \tag{3.33}$$

If the queue is stable, it is known that the sequence $\{W_n : n \geq 1\}$ converges in distribution to the steady-state waiting time W. Moreover, if we denote by $\overline{W}$ the maximal waiting time in the queue, i.e.,

$$\overline{W} = \sup\{W_n : \ n \geq 1\},$$

then $\overline{W}$ is almost surely finite and has the same distribution as W. For more details on stability of queues see, e.g., [43]. In the following, we denote by $\Theta_s \subset \Theta$ the stability subset of Θ, i.e.,

$$\Theta_s := \{\theta \in \Theta : \ \mathbb{E}_\theta[S - T] < 0\}.$$

Let $v_\alpha(x) = e^{\alpha x}$, for some $\alpha \geq 0$. Then we have

$$\|P_\theta\|_{v_\alpha} = \sup_{x \geq 0} \mathbb{E}_\theta \left[e^{\alpha(S-T)} \cdot \mathbb{I}_{\{x+S>T\}} \right]$$

and by the Dominated Convergence Theorem we obtain (see Example 3.3)

$$\|P_\theta\|_{v_\alpha} = \mathbb{E}_\theta \left[e^{\alpha(S-T)} \right],$$

provided that the expectation in the right-hand side is finite.

Remark 3.4. *In general, requiring that $\|P_\theta\|_{v_\alpha}$ is finite is a quite restrictive condition since, in particular, this requires that all moments of S exist. A sufficient condition for $\|P_\theta\|_{v_\alpha} < \infty$ is that μ_θ has sub-exponential tail, for each θ, i.e.*

$$(\mathrm{C}) : \ \exists \gamma, \beta, M > 0 : \ \mathbb{P}_\theta\{S > x\} \leq \gamma e^{-\beta x}, \forall x \geq M.$$

Indeed, let us assume that condition (C) holds true for some $\theta \in \Theta$. Since $e^{\alpha S}$ is a strictly positive random variable it holds that

$$\mathbb{E}_\theta \left[e^{\alpha S} \right] = \int_1^\infty \mathbb{P}_\theta \left\{ e^{\alpha S} > x \right\} dx \leq e^{\alpha M} + \gamma \int_{e^{\alpha M}}^\infty e^{-\frac{\beta}{\alpha} \ln x} dx.$$

Hence, we conclude that

$$\forall \alpha < \beta : \ \|P_\theta\|_{v_\alpha} = \mathbb{E}_\theta \left[e^{\alpha(S-T)} \right] \leq \mathbb{E}_\theta \left[e^{\alpha S} \right] \leq e^{\alpha M} \left(1 + \frac{\gamma \alpha e^{-\beta M}}{\beta - \alpha} \right) < \infty.$$

The key observation is that, under the stability condition, given by (3.33), we have $\|P_\theta\|_{v_\alpha} < 1$, for some $\alpha > 0$, which means that the bound in (3.28) decreases at a geometric rate. More specifically, given $\theta_1, \theta_2 \in \Theta_s$, such that $\theta_1 < \theta_2$, there exist sufficiently small $\alpha > 0$ such that $\|P_\theta\|_{v_\alpha} < 1$, uniformly in $\theta \in [\theta_1, \theta_2]$. The precise statement is as follows.

Lemma 3.4. *For arbitrary $\alpha \geq 0$ let $v_\alpha(x) = e^{\alpha x}$, for all $x \geq 0$, and $\theta_1, \theta_2 \in \Theta_s$ be such that $\theta_1 < \theta_2$. If μ_θ is weakly $[\mathcal{C}]_{v_{\alpha^*}}$-continuous on $[\theta_1, \theta_2]$, for some $\alpha^* > 0$, then there exists $\bar{\alpha} > 0$ such that for each $\alpha \in (0, \bar{\alpha})$ it holds that*

$$\sup_{\theta \in [\theta_1, \theta_2]} \|P_\theta\|_{v_\alpha} < 1. \tag{3.34}$$

Proof. Let $F : [0,\infty) \times [\theta_1, \theta_2] \to \mathbb{R} \cup \{\infty\}$ be defined as

$$\forall \alpha, \theta : \; F(\alpha, \theta) = \|P_\theta\|_{v_\alpha} = \mathbb{E}_\theta \left[e^{\alpha(S-T)} \right]. \tag{3.35}$$

Note that, by hypothesis we have that $F(\alpha,\theta) < \infty$ for all $\alpha \in [0, \alpha^*]$ and $\theta \in [\theta_1, \theta_2]$. Moreover, for $\alpha \in [0, \alpha^*)$ and $\theta \in [\theta_1, \theta_2]$ we have

$$\forall n \geq 0 : \; \sup_{y \in \mathbb{R}} |y|^n e^{(\alpha - \alpha^*)y} = \left[\frac{1}{(\alpha^* - \alpha)e} \right]^n.$$

Therefore, it follows that[5]

$$\forall n \geq 0, y \in \mathbb{R} : \; |y|^n e^{\alpha y} \leq \left[\frac{1}{(\alpha^* - \alpha)e} \right]^n e^{\alpha^* y} \tag{3.36}$$

and by letting $y = S - T$ in (3.36) and taking expected values, we arrive at

$$\mathbb{E}_\theta \left[|S-T|^n e^{\alpha(S-T)} \right] \leq \left[\frac{1}{(\alpha^* - \alpha)e} \right]^n \mathbb{E}_\theta \left[e^{\alpha^*(S-T)} \right] < \infty.$$

On the other hand, for $\theta \in [\theta_1, \theta_2]$ we have $F(0, \theta) = 1$, $\lim_{\alpha \uparrow \infty} F(\alpha, \theta) = \infty$ and

$$\lim_{\alpha \downarrow 0} \frac{d}{d\alpha} F(\alpha, \theta) = \lim_{\alpha \downarrow 0} \mathbb{E}_\theta \left[(S-T) e^{\alpha(S-T)} \right] = \mathbb{E}_\theta [S - T] < 0.$$

Moreover, the second derivative with respect to α satisfies

$$\forall \alpha \in (0, \alpha^*), \theta \in [\theta_1, \theta_2] : \; \frac{d^2}{d\alpha^2} F(\alpha, \theta) = \mathbb{E}_\theta \left[(S-T)^2 e^{\alpha(S-T)} \right] > 0.$$

Hence, we conclude that F is strictly convex in α and consequently for each $\theta \in [\theta_1, \theta_2]$ there exist an unique $\alpha > 0$ satisfying $F(\alpha, \theta) = 1$. If we denote this value by α_θ then

$$\forall \alpha \in (0, \alpha_\theta) : \; F(\alpha, \theta) < 1.$$

Continuity[6] of F in both α and θ implies continuity of the implicit function $\theta \mapsto \alpha_\theta$; see, e.g., [38]. Therefore, we have $\inf\{\alpha_\theta : \; \theta \in [\theta_1, \theta_2]\} > 0$. Letting

$$\bar{\alpha} = \min\{\alpha^*, \inf\{\alpha_\theta : \; \theta \in [\theta_1, \theta_2]\}\}$$

concludes the proof. □

Now we are able to state and prove the main result of this section. The precise statement is as follows.

[5] Note that, if $v(y) = e^{\alpha^* y}$ and $g(y) = |y|^n e^{\alpha y}$, for $y \in \mathbb{R}$, then $g \in [\mathcal{C}]_v$ and $\|g\|_v = [(\alpha^* - \alpha)e]^{-n}$. Consequently, the inequality in (3.36) reads $|g(y)| \leq \|g\|_v \cdot v(y)$.

[6] Note that, if $\alpha < \alpha^*$ then $[\mathcal{C}]_{v_{\alpha^*}}$-continuity implies $[\mathcal{C}]_{v_\alpha}$-continuity; see Remark 1.1.

Theorem 3.4. *Let $v_\alpha(x) = e^{\alpha x}$, for $\alpha \geq 0$, and $\theta_1, \theta_2 \in \Theta$ be such that $\theta_1 < \theta_2$. If μ_θ is $[\mathcal{C}]_{v_{\alpha^*}}$-continuous on Θ, for some $\alpha^* > 0$ then for each $\alpha \in (0, \bar{\alpha})$, i.e., α satisfies (3.34) (see Lemma 3.4), we have:*

(i) For each $\theta \in [\theta_1, \theta_2]$ the distribution of W_n is $[\mathcal{C}]_{v_\alpha}$-convergent to its stationary distribution, i.e.,

$$\forall f \in [\mathcal{C}]_{v_\alpha} : \lim_{n\to\infty} \mathbb{E}_\theta[f(W_n)] = \mathbb{E}_\theta[f(W)].$$

(ii) If, in addition, μ_θ is weakly $[\mathcal{C}]_{v_\alpha}$-differentiable on Θ, for some $\alpha \in (0, \bar{\alpha})$, then the stationary distribution of the sequence $\{W_n : n \geq 1\}$ is strongly v_α-norm Lipschitz continuous, i.e., there exist $\mathbf{K}_\alpha(\theta_1, \theta_2) > 0$ such that

$$\forall f \in [\mathcal{C}]_{v_\alpha} : \frac{|\mathbb{E}_{\theta_2}[f(W)] - \mathbb{E}_{\theta_1}[f(W)]|}{\theta_2 - \theta_1} \leq \|f\|_{v_\alpha} \mathbf{K}_\alpha(\theta_1, \theta_2). \tag{3.37}$$

Moreover, the constant $\mathbf{K}_\alpha(\theta_1, \theta_2)$ can be chosen as

$$\mathbf{K}_\alpha(\theta_1, \theta_2) = 2 \sup_{\theta\in[\theta_1,\theta_2]} \frac{c_\theta \|P_\theta^*\|_{v_\alpha}}{(1 - \|P_\theta\|_{v_\alpha})^2} < \infty. \tag{3.38}$$

Proof. First, we show that for $\alpha \in (0, \bar{\alpha})$ (see Lemma 3.4) and $\theta \in [\theta_1, \theta_2]$ it holds that

$$\sup_{n\geq 0} \mathbb{E}_\theta \left[v_\alpha(W_{n+1})\right] \leq \frac{1}{1 - \|P_\theta\|_{v_\alpha}}. \tag{3.39}$$

Indeed, from Lindley's recursion we have

$$\begin{aligned} \forall n \geq 1 : \mathbb{E}_\theta\left[e^{\alpha W_{n+1}}\right] &= \mathbb{P}_\theta\{W_{n+1} = 0\} + \mathbb{E}_\theta\left[e^{\alpha W_{n+1}} \cdot \mathbb{I}_{\{W_{n+1}>0\}}\right] \\ &\leq 1 + \mathbb{E}_\theta\left[e^{\alpha(W_n+S-T)}\right] \\ &\leq 1 + \|P_\theta\|_{v_\alpha}\mathbb{E}_\theta\left[e^{\alpha W_n}\right] \end{aligned}$$

and from finite induction it follows that

$$\forall n \geq 0 : \mathbb{E}_\theta\left[e^{\alpha W_{n+1}}\right] \leq \sum_{k=0}^{n} \|P_\theta\|_{v_\alpha}^k \leq \frac{1}{1 - \|P_\theta\|_{v_\alpha}}.$$

Taking the supremum with respect to $n \geq 0$ concludes the proof of (3.39).

As explained in Remark 3.3, the distribution of W_n is $\mathcal{C}_B$-convergent to the stationary distribution. Then, according to Theorem 1.1 *(i)*, $[\mathcal{C}]_{v_\alpha}$-convergence follows from the uniform integrability of the sequence $\{v_\alpha(W_n) : n \geq 1\}$. A sufficient condition, according to Lemma 1.1, is the existence of a function ϑ satisfying

$$\sup_{n\geq 1} \mathbb{E}_\theta[\vartheta(v_\alpha(W_n))] < \infty, \ \lim_{x\to\infty} \frac{\vartheta(x)}{x} = \infty.$$

Recall that $v_\alpha(x) = e^{\alpha x}$, for some $\alpha \in (0, \bar{\alpha})$. Choosing some $\epsilon \in (0, \bar{\alpha} - \alpha)$ it follows from (3.39) that the function ϑ defined as

$$\forall x \geq 0 : \vartheta(x) = x^{\frac{\alpha+\epsilon}{\alpha}},$$

i.e., $\vartheta(v_\alpha(W_n)) = e^{(\alpha+\epsilon)W_n}$, satisfies the desired conditions, which concludes part *(i)* of the theorem.

Let $\alpha \in (0, \bar{\alpha})$. For $n \geq 0$ and $f \in [\mathcal{C}]_{v_\alpha}$ the Mean Value Theorem yields

$$\frac{\mathbb{E}_{\theta_2}[f(W_{n+1})] - \mathbb{E}_{\theta_1}[f(W_{n+1})]}{\theta_2 - \theta_1} = c_\theta \sum_{k=1}^{n} \mathbb{E}_\theta \left[f(W_{n+1}^{k+}) - f(W_{n+1}^{k-}) \right],$$

for some $\theta \in [\theta_1, \theta_2]$, depending on f and n and Lemma 3.3 implies that

$$\frac{|\mathbb{E}_{\theta_2}[f(W_{n+1})] - \mathbb{E}_{\theta_1}[f(W_{n+1})]|}{\theta_2 - \theta_1} \leq 2c_\theta \|f\|_{v_\alpha} \|P_\theta^*\|_{v_\alpha} \sum_{k=1}^{n} \|P_\theta\|_{v_\alpha}^{n-k} \mathbb{E}_\theta \left[v_\alpha(W_k) \right].$$

Therefore, taking (3.39) into account we conclude that

$$\begin{aligned} \frac{|\mathbb{E}_{\theta_2}[f(W_{n+1})] - \mathbb{E}_{\theta_1}[f(W_{n+1})]|}{\theta_2 - \theta_1} &\leq \frac{2c_\theta \|f\|_{v_\alpha} \|P_\theta^*\|_{v_\alpha}}{1 - \|P_\theta\|_{v_\alpha}} \sum_{k=1}^{n} \|P_\theta\|_{v_\alpha}^{n-k} \\ &\leq 2\|f\|_{v_\alpha} \frac{c_\theta \|P_\theta^*\|_{v_\alpha}}{(1 - \|P_\theta\|_{v_\alpha})^2}. \end{aligned} \tag{3.40}$$

and taking in (3.40) the supremum with respect to $\theta \in [\theta_1, \theta_2]$ yields

$$\frac{|\mathbb{E}_{\theta_2}[f(W_{n+1})] - \mathbb{E}_{\theta_1}[f(W_{n+1})]|}{\theta_2 - \theta_1} \leq 2\|f\|_{v_\alpha} \sup_{\theta \in [\theta_1, \theta_2]} \frac{c_\theta \|P_\theta^*\|_{v_\alpha}}{(1 - \|P_\theta\|_{v_\alpha})^2}. \tag{3.41}$$

Letting now $n \to \infty$ in (3.41) and taking *(i)* into account concludes *(ii)*, i.e., (3.37) holds true for $\mathbf{K}_\alpha(\theta_1, \theta_2)$ given by (3.38). □

The following result is a direct consequence of Theorem 3.4.

Corollary 3.2. *The stationary distribution of the waiting times in a G/G/1 queue with parameter-dependent service time distribution μ_θ is locally Lipschitz continuous on the stability set Θ_s, provided that the service time distribution is weakly $[\mathcal{C}]_{v_\alpha}$-differentiable on Θ_s, for some $\alpha > 0$.*

Proof. Let us denote by σ_θ, for $\theta \in \Theta_s$, the stationary distribution of the Markov chain $\{W_n : n \geq 1\}$ with respect to the expectation operator $\mathbb{E}_\theta$; see Remark 3.3. Since Θ_s is an open set, for arbitrary $\theta \in \Theta_s$ we choose $\theta_1, \theta_2 \in \Theta_s$ such that $\theta_1 < \theta < \theta_2$ and apply Theorem 3.4. By taking in (3.37) the supremum with respect to $\|f\|_{v_\alpha} \leq 1$ we obtain

$$\|\sigma_{\theta_2} - \sigma_{\theta_1}\|_{v_\alpha} \leq (\theta_2 - \theta_1) \mathbf{K}_\alpha(\theta_1, \theta_2),$$

with $\mathbf{K}_\alpha(\theta_1, \theta_2)$ given by (3.38), which concludes the proof. □

3.3.2 Comments and Bound Improvements

This section is intended to illustrate how the results in Section 3.3.1, in particular Theorem 3.4, can be used in practice and what issues have to be taken into account when doing so. In particular, we show how the bounds obtained in Section 3.3.1 can be slightly improved, in order to derive more accurate bounds.

We start by noting that in Theorem 3.4 $v_\alpha = e^{\alpha\cdot}$, for some $\alpha > 0$ satisfying (3.34); see Lemma 3.4. Examining the proof of Lemma 3.4 it turns out that α has to be small enough to satisfy (3.34). In principle, the largest value $\bar{\alpha}$ will decrease as $\theta_2 - \theta_1$ increases. In words, the larger the perturbation of the parameter, the smaller $\bar{\alpha}$ will be and apparently the less performance measures will be in $[\mathcal{C}]_{v_\alpha}$. But in fact, this is not the real issue since, by construction, we have $\bar{\alpha} > 0$ which means that usual performance measures, such as bounded and continuous mappings and moments belong to $[\mathcal{C}]_{v_\alpha}$, for α satisfying (3.34). There is, however, a trade-off between decreasing $\bar{\alpha}$ and the quality of bounds in (3.37) which is related to the v_α-norm of f. More specifically, the v_α-norm of a typical function will increase as α decreases; for instance, if f is the identity mapping note that

$$\|f\|_{v_\alpha} = \frac{1}{\alpha e}.$$

Therefore, we conclude that, while in principle Theorem 3.4 applies to any typical continuous mapping f, the quality of the bound depends on the v_α-norm of f which, in certain situations, can be prohibitively large. Nevertheless, Theorem 3.4 is still a worthy theoretical result where, depending on the situation, the bounds can be improved by using particular properties of the performance measure under consideration.

In addition, note that α_θ in the proof of Theorem 3.4 is defined as an implicit function in θ which, in practice, makes it quite difficult to calculate exactly. It is, however, worth noting that if $\{\mu_\theta : \theta \in \Theta\}$ is a stochastically monotone family, say increasing, then things become simpler. Indeed, the function F defined by (3.35) is non-decreasing in θ and a simple analysis shows that α_θ is non-increasing with respect to θ, which yields

$$\bar{\alpha} = \min\{\alpha^*, \alpha_{\theta_2}\}.$$

Moreover, if $\mu'_\theta = (c_\theta, \mu^+_\theta, \mu^-_\theta)$ then μ^+_θ is stochastically larger than μ^-_θ and one can choose $S^\pm \sim \mu^\pm_\theta$ such that $S^+ \geq S^-$, a.s.

$$\|P^*_\theta\|_{v_\alpha} = \mathbb{E}_\theta\left[e^{\alpha(S^+ - T)}\right].$$

Recall that $v_\alpha(x) = e^{\alpha x}$, for some $\alpha \geq 0$, and the right-hand side in (3.37) depends on α through v_α. By Remark 1.1 it follows that $[\mathcal{C}]_{v_\alpha}$-differentiability, for some $\alpha > 0$, implies $[\mathcal{C}]_{v_\beta}$-differentiability for any $\beta \in (0, \alpha)$. Therefore, for fixed f, one can obtain a more accurate Lipschitz bound in (3.41) by minimizing the right-hand side in (3.37) with respect to $\beta \in (0, \alpha)$, i.e., to replace $\mathbf{K}_\alpha(\theta_1, \theta_2)$ in (3.38) by

$$\mathbf{L}_\alpha(\theta_1, \theta_2) = 2 \inf_{\beta \in (0,\alpha)} \left(\|f\|_{v_\beta} \sup_{\theta \in [\theta_1, \theta_2]} \frac{c_\theta \|P^*_\theta\|_{v_\beta}}{(1 - \|P_\theta\|_{v_\beta})^2} \right).$$

We conclude this section with two examples which illustrate the above facts.

Example 3.5. *We revisit the M/U/1 queue treated in Example 3.4. Standard computation shows that*

$$\forall \lambda, \theta, \alpha : \; c_\theta = \frac{1}{\theta}, \quad \|P_\theta\|_v = \frac{\lambda(e^{\alpha\theta} - 1)}{\alpha\theta(\alpha+\lambda)}, \quad \|P_\theta^*\|_v = \frac{\lambda e^{\alpha\theta}}{\alpha+\lambda},$$

which leads to the following Lipschitz bound in (3.37):

$$\mathbf{K}_\alpha(\theta_1, \theta_2) = 2 \sup_{\theta \in [\theta_1, \theta_2]} \frac{\alpha^2 \lambda \theta(\alpha+\lambda) e^{\alpha\theta}}{[\alpha\theta(\lambda+\alpha) - \lambda(e^{\alpha\theta} - 1)]^2}.$$

Moreover, for fixed $\lambda > 0$, the stability set is given by $\Theta_s = [0, 2\lambda^{-1})$ and α_θ is the unique solution $\alpha > 0$ of the equation

$$\lambda(e^{\alpha\theta} - 1) = \alpha\theta(\alpha+\lambda).$$

Since μ_θ is stochastically increasing, in this case, it holds that $\bar{\alpha} = \alpha_{\theta_2}$. If, for instance, $\lambda = 1$ and $\theta_2 = 1$ then $\bar{\alpha} = \alpha_1 \approx 1.7934$. If $\theta_2 = 1.8$, i.e., a high traffic rate we have $\bar{\alpha} \approx 0.9984$ whereas for $\theta_2 = 0.1$, i.e., a small traffic rate it turns out that $\bar{\alpha} \approx 1.8768$.

For $f(x) = x$, we have $\|f\|_v = (\alpha e)^{-1}$ and we obtain

$$\mathbf{L} = \frac{2}{e} \cdot \inf_{\alpha \in (0, \bar{\alpha})} \sup_{\theta \in [\theta_1, \theta_2]} \frac{\alpha \lambda \theta(\alpha+\lambda) e^{\alpha\theta}}{[\alpha\theta(\lambda+\alpha) - \lambda(e^{\alpha\theta} - 1)]^2}.$$

Things become somewhat easier when considering the M/M/1 case, as the following example shows.

Example 3.6. *Let us replace in Example 3.4 μ_θ by the exponential distribution with rate θ. Then*

$$\forall \lambda, \theta, \alpha < \theta : \; c_\theta = \frac{1}{\theta e}, \quad \|P_\theta\|_{v_\alpha} = \frac{\lambda\theta}{(\alpha+\lambda)(\theta-\alpha)}, \quad \|P_\theta^*\|_{v_\alpha} = \left(\frac{\theta}{\theta-\alpha}\right)^2 e^{\frac{\alpha}{\theta}},$$

which leads to the following Lipschitz bound in (3.37):

$$\mathbf{K}_\alpha(\theta_1, \theta_2) = 2 \sup_{\theta \in [\theta_1, \theta_2]} \theta \left(\frac{(\alpha+\lambda)}{\alpha(\theta-\lambda-\alpha)}\right)^2 e^{-\frac{\theta-\alpha}{\theta}}.$$

In this situation, the stability set is given by $\Theta_s = (\lambda, \infty)$ and α_θ can be found in explicit form as the unique positive solution of the equation

$$\lambda\theta = (\alpha+\lambda)(\theta-\alpha).$$

It turns out that $\alpha_\theta = \theta - \lambda$ and since μ_θ is stochastically decreasing, in this case, we conclude that $\bar{\alpha} = \alpha_{\theta_1} = \theta_1 - \lambda$.

3.4 Concluding Remarks

This chapter presents an important class of applications of weak differentiation theory. Starting from the observation that the gradient provides relevant information about the local variation of some function we perform a sensitivity analysis for some common mathematical models, among which Markov chains are maybe the most important ones. In this setting we derive bounds on perturbations for transient performance measures in Section 3.2.2 and, moreover, under stability conditions we extend our analysis to steady-state performances in Section 3.3.

Sensitivity analysis, based on weak differentiation, has been investigated in [33] and the theory of weak differentiation was applied for studying stability of stationary Markov chains in [27]. In addition, the stability of steady-state performances of a Markov chain has been investigated in [35]. Here, we present a general (unified) approach which applies to virtually any stochastic system defined by a finite family of independent random variables and for a large class of performance measures. Unfortunately, while bounds on perturbations can be easily established by using representations of weak derivatives, the main pitfall of this method is the poor accuracy of the bounds which stems from the fact that the bound should apply to a highly diversified class of performance measures. Therefore, improving the bounds is conditioned on restricting their range of applicability and it is subject to future research.

Another possible direction of research is to establish results regarding weak differentiability of the stationary distribution of stable stochastic processes, in both discrete and continuous-time, provided that the theory of weak differentiation can be extended to the later ones. For instance, an interesting application would be to investigate weak differentiability of the stationary distribution of one-dimensional diffusions with reflecting barrier(s) with respect to the barrier(s) level.

Eventually, it is worth noting that the methods presented in this chapter can be applied to study sensitivity of non-parametric models. That is, to study the influence of replacing an input distribution, say μ, by another one, say η. This can be achieved by considering the parametric family of mixed distributions $\{\mu_\theta : \theta \in \Theta\}$ defined as follows:

$$\forall \theta \in [0, 1] : \ \mu_\theta := (1 - \theta) \cdot \mu + \theta \cdot \eta. \tag{3.42}$$

Obviously, $\mu_0 = \mu$, $\mu_1 = \eta$ and the parameter θ can be seen as a measure of the deviation from the initial distribution μ. It readily follows that the distribution μ_θ given by (3.42) is $[\mathcal{F}]_v$-differentiable, for any $v \in \mathcal{C}^+ \cap \mathcal{L}^1(\mu, \eta)$ and its weak derivative satisfies

$$\forall \theta \in [0, 1] : \ \mu'_\theta = \eta - \mu.$$

If, for instance, μ is an exponential distribution and η is a non-exponential distribution having the same mean, i.e.,

$$\int s \, \eta(ds) = \int s \, \mu(ds),$$

then one can use Theorem 3.4 to evaluate the steady-state effect of deviations from the M/G/1 regime in a stable queue.

4. MEASURE-VALUED DIFFERENTIAL CALCULUS

Throughout this chapter we aim to further extend the theory of differentiation for product measures, in order to develop a *weak differential calculus*, i.e., higher-order differentiation formulas and results on analyticity (read: Taylor series expansions). A first step into this direction has already been made in Section 2.3 where Theorem 2.3 and its extension to finite products (in Theorem 2.4) establish rules of differentiation for the first-order derivative of a product measure. In this chapter we extend these rules to higher-order differentiation.

4.1 Introduction

Starting point of our analysis will be the resemblance of Theorem 2.3 with classical analysis (differentiation formula for products of functions). Based on this, it is reasonable to expect that a "Leibnitz-Newton" rule for higher-order derivatives of the product of two measures would hold true, as well. Such a result will be established and then extended to finite products of measures, in Section 4.2.1. Like in conventional analysis, the extended Leibnitz-Newton rule, established by Theorem 4.2, will serve as a basis for measure-valued differential calculus. In addition to that, it provides the theoretical background for a formal differential calculus for a particular class of random objects, to be introduced in Chapter 5.

The similarities between classical and measure-valued differential calculus extend further to analyticity, which is a crucial condition for performing Taylor series expansions. This lead us to introduce and study the concept of weak analyticity in Section 4.2.2. It will turn out that, just like in conventional analysis, products of weakly analytic measures are again weakly analytic. This result will be very important in applications as it provides Taylor series approximations for the performance measures of parameter-dependent stochastic models with weakly analytic input distributions.

Although weak analyticity means actually *point-wise with respect to* g in $[\mathcal{D}]_v$, for some Banach base $(\mathcal{D}, v)$, it will turn out that some stronger results hold true. More specifically, we will show that the Taylor series attached to some weakly analytic probability measure converges strongly on some domain, i.e., "weak analyticity implies strong analyticity." This fact leads to the concept of $[\mathcal{D}]_v$*-radius of convergence.*

The chapter is organized as follows: In Section 4.2 we extend Theorem 2.4 to higher-order differentiation and we introduce and study the concept of weak analyticity while in Section 4.3 we illustrate the concept of weak analyticity by evaluating the completion time in a stochastic activity network.

4.2 Leibnitz-Newton Rule and Weak Analyticity

In this section we continue the analysis of product measures and show that, like in conventional analysis, properties such as higher-order differentiation and analyticity are inherited by products of measures. In Section 4.2.1 we present a generalized Leibnitz-Newton rule for weak derivatives and in Section 4.2.2 we will deal with analyticity issues.

4.2.1 Leibnitz-Newton Rule and Extensions

Inspired by Theorem 2.3, we proceed to establish the Leibnitz-Newton product rule which extends Theorem 2.3 to higher-order derivatives. The precise statement is the following.

Theorem 4.1. *Let $(\mathcal{D}(\mathbb{S}), v)$ and $(\mathcal{D}(\mathbb{T}), u)$ be Banach bases on $\mathbb{S}$ and $\mathbb{T}$, respectively. If μ_θ is n-times $[\mathcal{D}(\mathbb{S})]_v$-differentiable and if η_θ is n-times $[\mathcal{D}(\mathbb{T})]_u$-differentiable, then the product measure $\mu_\theta \times \eta_\theta \in \mathcal{M}(\sigma(\mathcal{S} \times \mathcal{T}))$ is n-times $[\mathcal{D}(\mathbb{S}) \otimes \mathcal{D}(\mathbb{T})]_{v\otimes u}$-differentiable and it holds that*

$$(\mu_\theta \times \eta_\theta)^{(n)} = \sum_{j=0}^{n} \binom{n}{j} \left(\mu_\theta^{(j)} \times \eta_\theta^{(n-j)}\right).$$

Proof. We proceed by induction over $n \geq 1$. For $n = 1$ the assertion reduces to Theorem 2.3. Assume now that the conclusion holds true for $n \geq 1$. Then,

$$(\mu_\theta \times \eta_\theta)^{(n+1)} = \left(\sum_{j=0}^{n} \binom{n}{j} \left(\mu_\theta^{(j)} \times \eta_\theta^{(n-j)}\right)\right)' = \sum_{j=0}^{n} \binom{n}{j} \left(\mu_\theta^{(j)} \times \eta_\theta^{(n-j)}\right)'.$$

Applying Theorem 2.3 to the derivatives in the right-hand side, the proof follows from basic algebraic calculations, just like in conventional analysis, by taking into account that weak derivatives satisfy (see Remark 2.3)

$$\forall j \geq 0 : \left(\mu_\theta^{(j)}\right)' = \mu_\theta^{(j+1)}.$$

□

The next result is a generalization of Theorem 4.1 and introduces the general formula of the weak differential calculus. Recall the definitions of Π_θ and $\vec{v}$ given in Section 2.3 by (2.27) and (2.28), respectively!

Theorem 4.2. *For $1 \leq i \leq k$, let $(\mathcal{D}(\mathbb{S}_i), v_i)$ be Banach bases on $\mathbb{S}_i$ such that $\mu_{i,\theta}$ is n-times $[\mathcal{D}(\mathbb{S}_i)]_{v_i}$-differentiable. Then, Π_θ is n-times $[\mathcal{D}(\mathbb{S}_1) \otimes \ldots \otimes \mathcal{D}(\mathbb{S}_k)]_{\vec{v}}$-differentiable and it holds that*

$$\Pi_\theta^{(n)} = \sum_{\jmath \in \mathcal{J}(k,n)} \binom{n}{\jmath_1, \ldots, \jmath_k} \cdot (\mu_{1,\theta})^{(\jmath_1)} \times \ldots \times (\mu_{k,\theta})^{(\jmath_k)}, \tag{4.1}$$

where, for $k, n \geq 1$, we set

$$\mathcal{J}(k,n) := \{\jmath = (\jmath_1, \ldots, \jmath_k) : \ 0 \leq \jmath_i \leq n, \ \jmath_1 + \ldots + \jmath_k = n\}.$$

Proof. The proof follows from Theorem 4.1, via finite induction over k. □

Theorem 4.2 can be seen as the generalized Leibniz-Newton rule for measure-valued differentiation. It provides an expression for the higher-order derivatives of finite product measures, provided that they exist. However, obtaining an instance of the weak derivative of such a product, i.e., a "triplet representation" is not straightforward since we deal with a sum of product signed measures and obtaining the Hahn-Jordan decomposition of $\Pi_\theta^{(n)}$ in (4.1) is quite demanding even in simple cases. Such a triplet representation would be useful in applications, as explained in Section 2.2.2 and in what follows we aim to establish such a result.

An instance of the weak derivative $\Pi_\theta^{(n)}$ in (4.1) can be obtained by inserting the appropriate weak derivatives for the measures $\mu_{i,\theta}^{(\jmath_i)}$ and rearranging terms in (4.1). In order to present the result we introduce the following notations. For $\jmath = (\jmath_1, \dots \jmath_k) \in \mathcal{J}(k,n)$ we denote by $\nu(\jmath)$ the number of non-zero elements of the vector $\jmath$ and by $\mathcal{I}(\jmath)$ the set of vectors $\imath \in \{-1, 0, +1\}^k$ such that $\imath_l \neq 0$ if and only if $\jmath_l \neq 0$ and such that the product of all non-zero elements of $\imath$ equals one, i.e., there is an even number of -1. For $\imath \in \mathcal{I}(\jmath)$, we denote by $\bar{\imath}$ the vector obtained from $\imath$ by changing the sign of the non-zero element at the highest position.

Corollary 4.1. *Under the conditions put forward in Theorem 4.2, let $\mu_{i,\theta}$ have m^{th}-order $[\mathcal{D}]_{v_i}$-derivative*

$$\mu_{i,\theta}^{(m)} = \left(c_{i,\theta}^{(m)}, \mu_{i,\theta}^{(m,+)}, \mu_{i,\theta}^{(m,-)}\right),$$

for $0 \le m \le n$, with $c_{i,\theta}^{(0)} = 1$ and $\mu_{i,\theta}^{(0,0)} = \mu_{i,\theta}$. For $n \ge 1$, an instance

$$\left(C_\theta^{(n)}, \Pi_\theta^{(n,+)}, \Pi_\theta^{(n,-)}\right)$$

of $\Pi_\theta^{(n)}$ is given by

$$\begin{aligned}
C_\theta^{(n)} &= \sum_{\jmath \in \mathcal{J}(k,n)} 2^{\nu(\jmath)-1} \binom{n}{\jmath_1, \dots, \jmath_k} \prod_{i=1}^{k} c_{i,\theta}^{(\jmath_i)}, \qquad (4.2)\\
\Pi_\theta^{(n,+)} &= \sum_{\jmath \in \mathcal{J}(k,n)} \binom{n}{\jmath_1, \dots, \jmath_k} \frac{\prod_{i=1}^{k} c_{i,\theta}^{(\jmath_i)}}{C_\theta^{(n)}} \cdot \sum_{\imath \in \mathcal{I}(\jmath)} \mu_{1,\theta}^{(\jmath_1,\imath_1)} \times \dots \times \mu_{k,\theta}^{(\jmath_k,\imath_k)},\\
\Pi_\theta^{(n,-)} &= \sum_{\jmath \in \mathcal{J}(k,n)} \binom{n}{\jmath_1, \dots, \jmath_k} \frac{\prod_{i=1}^{k} c_{i,\theta}^{(\jmath_i)}}{C_\theta^{(n)}} \cdot \sum_{\imath \in \mathcal{I}(\jmath)} \mu_{1,\theta}^{(\jmath_1,\bar{\imath}_1)} \times \dots \times \mu_{k,\theta}^{(\jmath_k,\bar{\imath}_k)},
\end{aligned}$$

where, for convenience, we identify

$$\forall 1 \le i \le k: \ \mu_{i,\theta}^{(\jmath_i,+1)} = \mu_{i,\theta}^{(\jmath_i,+)}, \ \mu_{i,\theta}^{(\jmath_i,-1)} = \mu_{i,\theta}^{(\jmath_i,-)}, \ \mu_{i,\theta}^{(0,0)} = \mu_{i,\theta}.$$

For practical purposes, the above result becomes more useful when formulated in terms of random variables. The precise statement is as follows.

Corollary 4.2. Random Variable Version of Theorem 4.2: *Under the conditions put forward in Corollary 4.1, if X_i are random variables having distributions $\mu_{i,\theta}$, for $1 \le i \le n$, respectively, then for each $g \in [\mathcal{D}(\mathbb{S}_1) \otimes \ldots \otimes \mathcal{D}(\mathbb{S}_k)]_{\vec{v}}$ we have*

$$\frac{d^n}{d\theta^n} P_g(\theta) = \sum_{\jmath \in \mathcal{J}(k,n)} C_\theta(\jmath) \sum_{\imath \in \mathcal{I}(\jmath)} \mathbb{E}_\theta \left[g\left(X_1^{(\jmath_1,\imath_1)}, \ldots, X_k^{(\jmath_k,\imath_k)} \right) - g\left(X_1^{(\jmath_1,\bar{\imath}_1)}, \ldots, X_k^{(\jmath_k,\bar{\imath}_k)} \right) \right],$$

where $P_g(\theta) = \mathbb{E}_\theta [g(X_1, \ldots, X_k)]$, $\mathbb{E}_\theta$ is an expectation operator consistent with

$$\forall \jmath \in \mathcal{J}(k,n), \imath \in \{-1,0,1\}^k : \left(X_1^{(\jmath_1,\imath_1)}, \ldots, X_k^{(\jmath_k,\imath_k)} \right) \sim \mu_{1,\theta}^{(\jmath_1,\imath_1)} \times \cdots \times \mu_{k,\theta}^{(\jmath_k,\imath_k)}$$

and for $\jmath \in \mathcal{J}(k,n)$ we set

$$C_\theta(\jmath) := \binom{n}{\jmath_1, \ldots, \jmath_k} \prod_{i=1}^{k} c_{i,\theta}^{(\jmath_i)}.$$

4.2.2 Weak Analyticity

In this section we introduce the concept of weak $[\mathcal{D}]_v$-analyticity for probability measures and we provide results regarding the radius of convergence of the Taylor series and weak analyticity of product measures.

Definition 4.1. *Let $(\mathcal{D}, v)$ be a Banach base on $\mathbb{S}$. We call the measure-valued mapping $\mu_* : \Theta \to \mathcal{M}_v$ weakly $[\mathcal{D}]_v$-analytic at θ, or weakly $[\mathcal{D}]_v$-analytic for short, if*

- *all higher-order $[\mathcal{D}]_v$-derivatives of μ_θ exist,*
- *exists a neighborhood V of θ such that for all ξ, satisfying $\theta + \xi \in V$, it holds that*

$$\forall g \in [\mathcal{D}]_v : \int g(s)\mu_{\theta+\xi}(ds) = \sum_{n=0}^{\infty} \frac{\xi^n}{n!} \cdot \int g(s)\mu_\theta^{(n)}(ds). \tag{4.3}$$

The expression $\mathbf{T}_n(\mu, \theta, \xi)$ defined as

$$\forall n \ge 0, \xi \in \mathbb{R} : \mathbf{T}_n(\mu, \theta, \xi) := \sum_{k=0}^{n} \frac{\xi^k}{k!} \cdot \mu_\theta^{(k)} \tag{4.4}$$

will be called the n^{th}-order Taylor polynomial of μ_* in θ. In addition, for fixed $g \in [\mathcal{D}]_v$, the maximal set $D_\theta(g, \mu)$ for which the equality in (4.3) holds true is called the domain of convergence of the Taylor series.

Remark 4.1. *Note that the n^{th}-order Taylor polynomial $\mathbf{T}_n(\mu, \theta, \xi)$ defined by (4.4) is, in fact, an element in $\mathcal{M}_v$ and defines a linear functional on $[\mathcal{D}]_v$. Therefore, (4.3) is equivalent to*

$$\forall \xi;\ \theta + \xi \in V : \mathbf{T}_n(\mu, \theta, \xi) \overset{[\mathcal{D}]_v}{\Longrightarrow} \mu_{\theta+\xi}. \tag{4.5}$$

Moreover, since all higher-order derivatives of μ_θ exist it follows by Theorem 2.1 that for each $n \ge 1$ $\mu_\theta^{(n)}$ is strongly continuous and by Theorem 2.2 (i) we conclude that $\mu_\theta^{(n-1)}$ is strongly differentiable. In particular, it follows that if μ_θ is weakly analytic then it is strongly differentiable of any order $n \ge 1$.

Note that the domain of convergence $D_\theta(g,\mu)$ of the series in (4.3) depends on g. Our next result provides a set $D_\theta^v(\mu) \subset \Theta$ where the Taylor series in (4.3) converges for all $g \in [\mathcal{D}]_v$. The precise statement is as follows.

Theorem 4.3. *Let $(\mathcal{D}, v)$ be a Banach base on $\mathbb{S}$ such that μ_θ is $[\mathcal{D}]_v$-analytic. Then for each $g \in [\mathcal{D}]_v$ the Taylor series in (4.3) converges for all ξ such that $|\xi| < R_\theta^v(\mu)$, where $R_\theta^v(\mu)$ is given by*

$$\frac{1}{R_\theta^v(\mu)} = \limsup_{n\in\mathbb{N}} \left(\frac{\|\mu_\theta^{(n)}\|_v}{n!} \right)^{\frac{1}{n}}. \tag{4.6}$$

In particular, the set $D_\theta^v(\mu) := \Theta \cap (\theta - R_\theta^v(\mu), \theta + R_\theta^v(\mu))$ satisfies

$$\forall g \in [\mathcal{D}]_v : \ D_\theta^v(\mu) \subset D_\theta(g,\mu).$$

Proof. We apply the Cauchy-Hadamard Theorem; see Theorem A.2 in the Appendix. It follows that the radius of convergence $R_\theta(g,\mu)$ of the Taylor series in (4.3) is given by

$$\frac{1}{R_\theta(g,\mu)} = \limsup_{n\in\mathbb{N}} \left(\frac{\left|\int g(s)\mu_\theta^{(n)}(ds)\right|}{n!} \right)^{\frac{1}{n}},$$

i.e., the series converges for $|\xi| < R_\theta(g,\mu)$ and it suffices to show that

$$\forall g \in [\mathcal{D}]_v : \ R_\theta^v(\mu) \le R_\theta(g,\mu). \tag{4.7}$$

This follows from the Cauchy-Schwartz inequality. To see this, note that

$$\left| \int g(s)\mu_\theta^{(n)}(ds) \right|^{\frac{1}{n}} \le \left(\|g\|_v \cdot \|\mu_\theta^{(n)}\|_v \right)^{\frac{1}{n}}$$

which, together with the fact that $\lim_{n\to\infty} \sqrt[n]{\|g\|_v} = 1$, for $g \in [\mathcal{D}]_v$, concludes the proof. □

The non-negative number $R_\theta^v(\mu)$ is called the $[\mathcal{D}]_v$*-radius of convergence* of μ_θ and the set $D_\theta^v(\mu)$ is called *the* $[\mathcal{D}]_v$*-domain of convergence* of μ_θ. Note, however, that in general this is not the maximal set for which the series converges for all $g \in [\mathcal{D}]_v$ since the inequality in (4.7) may be strict.

Example 4.1. *Let μ_θ denote the exponential distribution cf. Example 2.5. We show that the $[\mathcal{F}]_v$-radius of convergence of μ_θ satisfies $R_\theta^v(\mu) = \theta$, for $v(x) = 1 + x$, which shows that the Taylor series converges for $|\xi| < \theta$.*

Recall that an instance of the n^{th}-order derivative $\mu_\theta^{(n)}$ is given by

$$\mu_\theta^{(n)} = \begin{cases} \left(\frac{n!}{\theta^n}, \varepsilon_{n,\theta}, \varepsilon_{n+1,\theta}\right), & \text{if } n \text{ is odd,} \\ \left(\frac{n!}{\theta^n}, \varepsilon_{n+1,\theta}, \varepsilon_{n,\theta}\right), & \text{for } n \text{ even,} \end{cases}$$

where, for $n \ge 1$,

$$\varepsilon_{n,\theta}(dx) = \frac{\theta^n \cdot x^{n-1}}{(n-1)!} e^{-\theta x} dx.$$

Consequently, the v-norm $\|\mu_\theta^{(n)}\|_v$ satisfies

$$\left|\int v(x)\mu_\theta^{(n)}(dx)\right| \le \|\mu_\theta^{(n)}\|_v \le \frac{n!}{\theta^n}\int v(x)\varepsilon_{n+1,\theta}(dx) + \frac{n!}{\theta^n}\int v(x)\varepsilon_{n,\theta}(dx).$$

Elementary computation shows that for $p \ge 1$ we have

$$\int x^p\,\varepsilon_{n,\theta}(dx) = \frac{\theta^n}{(n-1)!}\int x^{n+p-1}e^{-\theta x}dx = \frac{1}{\theta^p}\cdot\frac{(n+p-1)!}{(n-1)!}.$$

Hence, for $v(x) = 1 + x$ we obtain the following inequalities

$$\frac{1}{\theta^{n+1}} \le \frac{\|\mu_\theta^{(n)}\|_v}{n!} \le \frac{2n+2\theta+1}{\theta^{n+1}}.$$

Finally, we obtain

$$\frac{1}{R_\theta^v(\mu)} = \limsup_{n\in\mathbb{N}}\left(\frac{\|\mu_\theta^{(n)}\|_v}{n!}\right)^{\frac{1}{n}} = \frac{1}{\theta}.$$

The same result holds true if one replaces v by any polynomial function.

Remark 4.2. *Theorem 4.3 shows that the Taylor series converges for $|\xi| < R_\theta^v(\mu)$, i.e., $\theta + \xi \in D_\theta^v(\mu)$. However, in general, the convergence of the Taylor series does not imply analyticity. Indeed, it can happen that the Taylor series is convergent but the limit does not coincide with the "true value". A standard example is that of the function*

$$f(x) = \begin{cases} e^{-\frac{1}{x^2}}, & x \ne 0, \\ 0, & x = 0, \end{cases}$$

for which all higher-order derivatives in 0 are equal to 0 but the function has obviously strictly positive values in any neighborhood of 0. Therefore, the maximal neighborhood V for which (4.3) holds true may not be equal to the domain of convergence of the Taylor series in the right-hand side of (4.3).

Nevertheless, since most of the usual functions, for which the Taylor series converge, are analytic we will assume in the following that the Taylor series converges to the "true value" for any ξ such that $\theta + \xi \in D_\theta^v(\mu)$.

The $[\mathcal{D}]_v$-domain of convergence $D_\theta^v(\mu)$ plays an important role in applications. The following result, which is a consequence of Theorem 4.3, will show that the sequence of Taylor polynomials converges strongly for $|\xi| < R_\theta^v(\mu)$.

Theorem 4.4. *Let $(\mathcal{D}, v)$ be a Banach base on $\mathbb{S}$ such that μ_θ is $[\mathcal{D}]_v$-analytic with $[\mathcal{D}]_v$-radius of convergence $R_\theta^v(\mu)$. Then,*

$$\forall\xi;\ |\xi| < R_\theta^v(\mu):\ \lim_{n\to\infty}\|\mathbf{T}_n(\mu,\theta,\xi) - \mu_{\theta+\xi}\|_v = 0.$$

Proof. By hypothesis, we have

$$\|\mathbf{T}_n(\mu,\theta,\xi) - \mu_{\theta+\xi}\|_v = \left\| \sum_{k=n+1}^{\infty} \frac{\xi^k}{k!} \cdot \mu_\theta^{(k)} \right\|_v \leq \sum_{k=n+1}^{\infty} \frac{|\xi|^k}{k!} \left\| \mu_\theta^{(k)} \right\|_v . \tag{4.8}$$

Let ξ be such that $|\xi| < R_\theta^v(\mu)$ and choose $\epsilon > 0$ such that $|\xi| + \epsilon < R_\theta^v(\mu)$. Since

$$\frac{1}{R_\theta^v(\mu) - \epsilon} > \frac{1}{R_\theta^v(\mu)} = \limsup_{n\in\mathbb{N}} \left(\frac{\|\mu_\theta^{(n)}\|_v}{n!} \right)^{\frac{1}{n}}$$

it follows that there exists some $n_\epsilon \geq 1$ such that

$$\forall k \geq n_\epsilon : \left(\frac{\|\mu_\theta^{(k)}\|_v}{k!} \right)^{\frac{1}{k}} < \frac{1}{R_\theta^v(\mu) - \epsilon}.$$

Consequently, we conclude from (4.8) that for each $n \geq n_\epsilon$ it holds that

$$\begin{aligned} \|\mathbf{T}_n(\mu,\theta,\xi) - \mu_{\theta+\xi}\|_v &\leq \sum_{k=n+1}^{\infty} \left(\frac{|\xi|}{R_\theta^v(\mu) - \epsilon} \right)^k \\ &= \frac{R_\theta^v(\mu) - \epsilon}{R_\theta^v(\mu) - \epsilon - |\xi|} \left(\frac{|\xi|}{R_\theta^v(\mu) - \epsilon} \right)^{n+1}, \end{aligned} \tag{4.9}$$

since, by assumption, $|\xi| < R_\theta^v(\mu) - \epsilon$. Therefore, the conclusion follows by letting $n \to \infty$ in (4.9). □

Example 4.2. *Let us consider the Bernoulli distribution*[1] β_θ *introduced in Example 2.4. Since* $\beta'_\theta = \delta_{x_2} - \delta_{x_1}$ *and higher-order derivatives* $\beta_\theta^{(n)}$*, for* $n \geq 2$*, are not significant it follows that* β_θ *is weakly analytic and the radius of convergence is* ∞ *(note that the Taylor series is finite). Indeed, we have*

$$\begin{aligned} \forall \theta, \xi \in \mathbb{R} : \ \beta_{\theta+\xi} &= (1 - \theta - \xi) \cdot \delta_{x_1} + (\theta + \xi) \cdot \delta_{x_2} \\ &= (1 - \theta) \cdot \delta_{x_1} + \theta \cdot \delta_{x_2} + \xi \cdot (\delta_{x_2} - \delta_{x_1}) \\ &= \beta_\theta + \xi \cdot \beta'_\theta. \end{aligned}$$

Example 4.3. *Let us revisit Example 4.1. We aim to show that the exponential distribution* μ_θ *is* $[\mathcal{F}]_v$*-analytic for any polynomial* v*, i.e., we show that (4.3) holds true for* $|\xi| < \theta$*,* $\mathcal{D} = \mathcal{F}$ *and polynomial* v*. To this end, note that the density* $f(x,\theta)$ *of* μ_θ *is analytic (in classical sense) in* θ*, i.e.,*

$$\forall x > 0, \forall \xi \in \mathbb{R} : \ f(x, \theta + \xi) = \sum_{k=0}^{\infty} \frac{\xi^k}{k!} \frac{d^k}{d\theta^k} f(x,\theta).$$

[1] Note that, for $\theta \in [0,1]$, β_θ is a probability distribution while, for general $\theta \in \mathbb{R}$, β_θ is a (signed) measure having total mass 1.

Hence, (4.3) is equivalent to

$$\forall g \in [\mathcal{F}]_v : \sum_{k=0}^{\infty} \frac{\xi^k}{k!} \int g(x) \frac{d^k}{d\theta^k} f(x,\theta) dx = \int g(x) \sum_{k=0}^{\infty} \frac{\xi^k}{k!} \frac{d^k}{d\theta^k} f(x,\theta) dx.$$

Fix $g \in [\mathcal{F}]_v$. In order to apply the Dominated Convergence Theorem it suffices to show that for each ξ such that $|\xi| < \theta$ the function

$$F_\theta(\xi, x) := \sum_{k=0}^{\infty} \left| g(x) \frac{\xi^k}{k!} \frac{d^k}{d\theta^k} f(x,\theta) \right|$$

is integrable with respect to x. Computing the derivatives of $f(x,\theta)$; see Example 2.5, we arrive at the following inequality

$$F_\theta(\xi, x) \le |g(x)| \sum_{k=0}^{\infty} \frac{|\xi|^k}{k!} (\theta x^k + k x^{k-1}) e^{-\theta x} \le \|g\|_v (\theta + |\xi|) v(x) e^{-(\theta - |\xi|)x}.$$

Since the right-hand side above is obviously integrable for $|\xi| < \theta$ we conclude that for $\theta > 0$ the exponential distribution μ_θ is weakly $[\mathcal{F}]_v$-analytic, for any polynomial v, and the corresponding Taylor series converges for $|\xi| < \theta$; compare to Example 4.1.

In classical analysis it is well known that the product of two analytic functions is again analytic. The following theorem establishes the counterpart of this fact for weak analyticity of measures. Namely, if μ_θ and η_θ are weakly analytic measures then the product $(\mu \times \eta)_\theta$ is again weakly analytic, where

$$\forall \theta \in \Theta : \ (\mu \times \eta)_\theta := \mu_\theta \times \eta_\theta.$$

The precise statement is as follows.

Theorem 4.5. *Let $(\mathcal{D}(\mathbb{S}), v)$ and $(\mathcal{D}(\mathbb{T}), u)$ be Banach bases on $\mathbb{S}$ and $\mathbb{T}$, respectively. Let μ_θ be $[\mathcal{D}(\mathbb{S})]_v$-analytic and η_θ be $[\mathcal{D}(\mathbb{T})]_u$-analytic with domains of convergence $D_\theta^v(\mu)$ and $D_\theta^u(\eta)$, respectively. Then the product measure $\mu_\theta \times \eta_\theta$ is $[\mathcal{D}(\mathbb{S}) \otimes \mathcal{D}(\mathbb{T})]_{v \otimes u}$-analytic and its domain of convergence $D_\theta^{v\otimes u}(\mu \times \eta)$ satisfies*

$$D_\theta^v(\mu) \cap D_\theta^u(\eta) \subset D_\theta^{v\otimes u}(\mu \times \eta). \tag{4.10}$$

More specifically, if $\theta + \xi \in D_\theta^v(\mu) \cap D_\theta^u(\eta)$ and $g \in [\mathcal{D}(\mathbb{S}) \otimes \mathcal{D}(\mathbb{T})]_{v\otimes u}$ it holds that

$$\int g(s,t) (\mu \times \eta)_{\theta+\xi}(ds, dt) = \sum_{k=0}^{\infty} \frac{\xi^k}{k!} \int g(s,t) (\mu \times \eta)_\theta^{(k)}(ds, dt). \tag{4.11}$$

Proof. Recall that, by definition, we have

$$D_\theta^v(\mu) = \Theta \cap (\theta - R_\theta^v(\mu), \theta + R_\theta^v(\mu)), \ D_\theta^u(\eta) = \Theta \cap (\theta - R_\theta^u(\eta), \theta + R_\theta^u(\eta)).$$

Hence, if we set $R_\theta := \min\{R_\theta^v(\mu), R_\theta^u(\eta)\}$ it follows that

$$D_\theta^v(\mu) \cap D_\theta^u(\eta) = \Theta \cap (\theta - R_\theta, \theta + R_\theta).$$

Next, we show that (4.11) holds true for any ξ such that $|\xi| < R_\theta$ and $g \in [\mathcal{D}(\mathbb{S})\otimes\mathcal{D}(\mathbb{T})]_{v\otimes u}$. To this end, note that according to Theorem 4.1 all higher-order derivatives of $(\mu \times \eta)_\theta$ exist. In addition, the right-hand side of (4.11) can be re-written as

$$\lim_{k\to\infty} \sum_{0\le j+l\le k} \frac{\xi^{j+l}}{j!l!} \int\int g(s,t)\mu_\theta^{(j)}(ds)\eta_\theta^{(l)}(dt). \tag{4.12}$$

Let us consider the Taylor polynomials[2] $\mathbf{T}_n : [\mathcal{D}(\mathbb{S})]_v \to \mathbb{R}$ defined as

$$\forall n \ge 0: \; \mathbf{T}_n(f) := \sum_{j=0}^{n} \frac{\xi^j}{j!} \int f(s)\mu_\theta^{(j)}(ds),$$

for $f \in [\mathcal{D}]_v$. First, note that according to (1.30) it holds that

$$\|g(\cdot,t)\|_v \le \|g\|_{v\otimes u} u(t).$$

Therefore, $g(\cdot,t) \in [\mathcal{D}(\mathbb{S})]_v$, for each $t \in \mathbb{T}$, and by hypothesis we conclude from (4.5) that

$$\forall t \in \mathbb{T}: \; \int g(s,t)\mu_{\theta+\xi}(ds) = \lim_{n\to\infty} \mathbf{T}_n(g(\cdot,t)).$$

In addition, an application of the Cauchy-Schwarz inequality yields

$$\forall t \in \mathbb{T}: \; |\mathbf{T}_n(g(\cdot,t))| \le \|\mathbf{T}_n\|_v \|g(\cdot,t)\|_v \le \|g\|_{v\otimes u} u(t) \sup_{n\ge 0} \|\mathbf{T}_n\|_v. \tag{4.13}$$

Next, we show that the Dominated Convergence Theorem applies to the sequence of mappings $\{t \mapsto \mathbf{T}_n(g(\cdot,t))\}_{n\ge 1}$, when integrated withe respect to η_θ, for each $\theta \in \Theta$. Indeed, we note that weak analyticity of μ_θ implies that $\{\mathbf{T}_n(f) : n \in \mathbb{N}\}$ is bounded for each $f \in [\mathcal{D}(\mathbb{S})]_v$. Applying the Banach-Steinhaus Theorem; see Lemma 1.4, we conclude that $\sup_n \|\mathbf{T}_n\|_v < \infty$. Therefore, if for $n \ge 0$ we set

$$\forall t \in \mathbb{T}: \; H_n(t) = \mathbf{T}_n(g(\cdot,t)),$$

it follows from (4.13) that $H_n \in [\mathcal{D}(\mathbb{T})]_u$ and

$$\|H_n\|_u \le \|g\|_{v\otimes u} \sup_{n\ge 0} \|\mathbf{T}_n\|_v.$$

Since, by hypothesis, $u \in \mathcal{L}^1(\eta_\theta : \theta \in \Theta)$, the Dominated Convergence Theorem applies to the sequence $\{H_n\}_n$ and yields

$$\int g(s,t)(\mu \times \eta)_{\theta+\xi}(ds,dt) = \lim_{n\to\infty} \int \mathbf{T}_n(g(\cdot,t))\eta_{\theta+\xi}(dt). \tag{4.14}$$

[2] For ease of notation we replace $\mathbf{T}_n(\mu,\theta,\xi)$ by $\mathbf{T}_n$. Recall that the Taylor polynomials $\mathbf{T}_n$, for $n \ge 0$, are linear functionals on $[\mathcal{D}]_v$ (see Remark 4.1) and by Theorem 4.3 weak analyticity of μ_θ implies that for each ξ, satisfying $|\xi| < R_\theta^v(\mu)$, $\mu_{\theta+\xi}$ is the $[\mathcal{D}]_v$-limit of the sequence $\mathbf{T}_n$; see (4.5).

Moreover, from $[\mathcal{D}(\mathbb{T})]_u$-analyticity of η_θ we conclude that the right-hand side in (4.14) equals to

$$\lim_{n\to\infty}\lim_{m\to\infty}\sum_{l=0}^{m}\frac{\xi^l}{l!}\int \mathbf{T}_n(g(\cdot,t))\eta_\theta^{(l)}(dt).$$

Therefore, we conclude that the left-hand side of (4.11) equals to

$$\lim_{n\to\infty}\lim_{m\to\infty}\sum_{l=0}^{m}\sum_{j=0}^{n}\frac{\xi^{j+l}}{j!l!}\int\int g(s,t)\mu_\theta^{(j)}(ds)\eta_\theta^{(l)}(dt). \tag{4.15}$$

The power series in (4.15) is convergent for $|\xi| < R_\theta$. Hence it is absolutely convergent, so its limit is not affected by re-shuffling terms and from the Rearrangements Theorem (see Theorem A.1 in the Appendix) it follows that the limits in (4.15) and (4.12) coincide, i.e., (4.11) holds true for $|\xi| < R_\theta$. Therefore, it follows that $(\mu\times\eta)_\theta$ is $[\mathcal{D}(\mathbb{S})\otimes\mathcal{D}(\mathbb{T})]_{v\otimes u}$-analytic and the inclusion in (4.10) holds true. □

Just like in conventional analysis, Theorem 4.5 can be extended to finite products of measures.

Corollary 4.3. *For $1 \le i \le k$, let $(\mathcal{D}(\mathbb{S}_i), v_i)$ be a Banach base on $\mathbb{S}_i$ such that $\mu_{i,\theta}$ is weakly $[\mathcal{D}(\mathbb{S}_i)]_{v_i}$-analytic having domain of convergence $D_\theta^{v_i}(\mu_i)$, respectively. Then, Π_θ is $[\mathcal{D}(\mathbb{S}_1)\otimes\ldots\otimes\mathcal{D}(\mathbb{S}_k)]_{\vec{v}}$-analytic and for each ξ such that $\theta+\xi \in D_\theta^{v_i}(\mu_i)$, for each $1 \le i \le k$, it holds that*

$$\forall g \in [\mathcal{D}(\mathbb{S}_1)\otimes\ldots\otimes\mathcal{D}(\mathbb{S}_k)]_{\vec{v}}: \int g(s)\Pi_{\theta+\xi}(ds) = \sum_{n=0}^{\infty}\frac{\xi^n}{n!}\int g(s)\Pi_\theta^{(n)}(ds).$$

Proof. This follows by finite induction from Theorem 4.5. □

4.3 Application: Stochastic Activity Networks (SAN)

Stochastic Activity Networks (SAN) such as those arising in Project Evaluation Review Technique (PERT) form an important class of models for systems and control engineering. Roughly, a SAN is a collection of activities, each with some (deterministic or random) duration, along with a set of precedence constraints, which specify that activities begin only when certain others have finished. Such a network can be modeled as a directed acyclic weighted graph $(\mathcal{V}, \mathcal{E} \subset \mathcal{V}\times\mathcal{V})$ with one source, one sink node and additive[3] weight-function $\tau : \mathcal{E} \to \mathbb{R}$.

A simple example is provided in Figure 4.1 below. The network has 5 nodes, labeled from 1 (source) to 5 (sink) and the edges denote the activities under consideration. The weights X_i, $1 \le i \le 7$, denote the durations of the corresponding activities. For instance, activity 6 can only begin when both activities 2 and 3 have finished. For a more detailed overview of stochastic activity networks we refer to [49].

Let $\mathcal{P}$ denote the set of all paths from the source to the sink node. Should (some) durations be random variables, we assume them mutually independent. However, note

[3] The weight of any path is given by the sum of the weights of the subsequent edges.

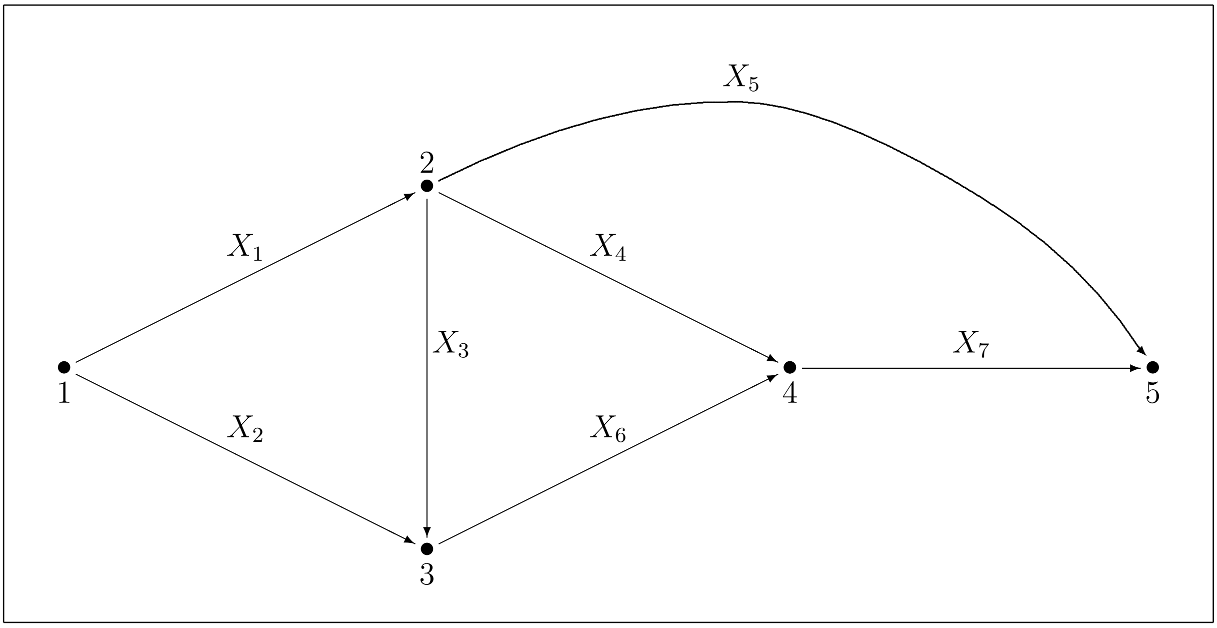

Fig. 4.1: A Stochastic Activity Network with source node 1 and sink node 5.

that in general the path weights are not independent. The completion time, denoted by T, is defined as the weight of the maximal path, i.e.,

$$T = \max\{\tau(\pi) : \pi \in \mathcal{P}\}.$$

For instance, in the above example, the set of paths from source node 1 to sink node 5, is

$$\mathcal{P} = \{(1,2,5); (1,2,4,5); (1,2,3,4,5); (1,3,4,5)\}.$$

Thus, the completion time in this case can be expressed as

$$T = \max\{X_1 + X_5; X_1 + X_4 + X_7; X_1 + X_3 + X_6 + X_7; X_2 + X_6 + X_7\}.$$

One of the most challenging problems in this area is to compute *the expected completion time* $\mathbb{E}[T]$. Distribution free bounds for $\mathbb{E}[T]$ are provided in [18]. In the following we aim to establish a functional dependence between a particular parameter, e.g., the expected duration of some particular task(s), and the expected completion time of the system. Here, we propose a Taylor series approximation for a SAN with exponentially distributed activity times, where the computation of higher-order derivatives relies on the weak differential calculus presented in this chapter.

We start by considering $\mathbb{S} = [0, \infty)$ with the usual metric and $v : \mathbb{S} \to \mathbb{R}$ defined as $v(x) = 1 + x$. Next, we define $g^T : \mathbb{S}^7 \to \mathbb{R}$,

$$g^T(x_1, \ldots, x_7) := \max\{x_1 + x_5; x_1 + x_4 + x_7; x_1 + x_3 + x_6 + x_7; x_2 + x_6 + x_7\},$$

i.e., $T = g^T(X_1, \ldots, X_7)$ and

$$\mathbb{E}[T] = \int \ldots \int g^T(x_1, \ldots, x_7)\mu_1(dx_1) \ldots \mu_7(dx_7),$$

where we denote by μ_i the distribution of X_i, for $1 \leq i \leq 7$. In accordance with Theorem 4.2 it holds that if μ_i is weakly differentiable with respect to some parameter θ, for all $1 \leq i \leq 7$, then the distribution of T is weakly differentiable with respect to θ, as well. Roughly speaking, that means that "the distribution of T is differentiable with respect to each μ_i."[4]

Assume for instance that the random variables X_i, for $1 \leq i \leq 7$, are independent and exponentially distributed with rates λ_i, respectively. We let $\lambda_1 = \lambda_3 = \theta$ be variable and let the other rates be fixed, i.e., deterministic and not a function of θ. By Example 4.3, the exponential distribution is weakly $[\mathcal{F}]_v$-analytic, for $v(x) = 1 + x$, and the domain of convergence is given by $|\xi| < \theta$. Since the distributions which are independent of θ are trivially weakly analytic, we conclude from Theorem 4.5 that the joint distribution of the vector $(X_1, \ldots, X_7)$ is weakly $[\mathcal{F}(\mathbb{S}^7)]_{v\otimes\ldots\otimes v}$-analytic. Moreover, the domain of convergence of the corresponding Taylor series includes the set $\{\xi : |\xi| < \theta\}$. Finally, we note that

$$|g^T(x_1, \ldots, x_7)| \leq \prod_{i=1}^{7}(1 + x_i) = (v \otimes \ldots \otimes v)(x_1, \ldots, x_7),$$

i.e., g^T belongs to $[\mathcal{F}(\mathbb{S}^7)]_{v\otimes\ldots\otimes v}$, the 7-fold product of the Banach base $(\mathcal{F}, v)$.

Next we proceed to the computation of derivatives in accordance with Corollary 4.2. Since only the derivatives of $\mu_{1,\theta}$ and $\mu_{3,\theta}$ are significant we consider, for $j, k \geq 0$, a modified network where X_1 is replaced by the sum of j independent samples from an exponentially distributed random variable with rate θ and X_3 is replaced by the sum of k independent samples from the same distribution whereas all other durations remain unchanged, i.e., we replace the exponential distributions of X_1 and X_3 by the Erlang $\varepsilon_{j,\theta}$ and $\varepsilon_{k,\theta}$ distributions, respectively; see Example 2.5.

More specifically, let $\{X_{1,l} : l \geq 1\}$ and $\{X_{3,l} : l \geq 1\}$ be two sequences of i.i.d. random variables having exponential distribution with rate θ and let $T_{j,k}$ denote the completion time of the modified SAN, i.e.,

$$T_{j,k} = g^T(\tilde{X}_1, \ldots, \tilde{X}_7),$$

where we define

$$\forall 1 \leq i \leq 7 : \ \tilde{X}_i := \begin{cases} \sum_{l=1}^{j} X_{1,l}, & i = 1; \\ \sum_{l=1}^{k} X_{3,l}, & i = 3; \\ X_i, & i \notin \{1, 3\}. \end{cases}$$

We have $T_{1,1} = T$ and we agree that $T_{j,k} = 0$ if either $j = 0$ or $k = 0$. With this notation Corollary 4.2 yields

$$\forall n \geq 0 : \ \frac{d^n}{d\theta^n}\mathbb{E}_\theta[T] = (-1)^n \frac{n!}{\theta^n} \sum_{j+k=n} \mathbb{E}_\theta[T_{j+1,k+1} - T_{j+1,k} - T_{j,k+1} + T_{j,k}] \tag{4.16}$$

and for each $n \geq 1$ we obtain by

$$\mathbf{T}_n(\theta, \xi) := \sum_{m=0}^{n}(-1)^m \left(\frac{\xi}{\theta}\right)^m \sum_{j+k=m} \mathbb{E}_\theta[T_{j+1,k+1} - T_{j+1,k} - T_{j,k+1} + T_{j,k}] \tag{4.17}$$

[4] Note that for a deterministic system, i.e., all the weights are deterministic, the completion time is, in general, not everywhere differentiable w.r.t. the weights. That is because the Dirac distribution δ_θ is not weakly differentiable w.r.t. θ.

the n^{th} *order Taylor polynomial* for $\mathbb{E}_{\theta+\xi}[T]$ at θ, where, for $\theta \in \Theta$, $\mathbb{E}_\theta$ denotes an expectation operator consistent with $X_i \sim \mu_i$, for $i \notin \{1,3\}$, $X_{1,l} \sim \mu_{1,\theta}$ and $X_{3,l} \sim \mu_{3,\theta}$, for all $l \geq 1$. Therefore, the coefficients of Taylor polynomials are completely determined by the values $\mathbb{E}_\theta[T_{j,k}]$, for $j,k \geq 0$.

Moreover, using a monotonicity argument, one can easily check that

$$\forall j,k \geq 0: \ |\mathbb{E}_\theta[T_{j+1,k+1} - T_{j+1,k} - T_{j,k+1} + T_{j,k}]| \leq 2\mathbb{E}_\theta[X_{3,k+1}] = \frac{2}{\theta}. \tag{4.18}$$

Hence, a bound for the error of the n^{th} order Taylor polynomial is given by

$$\begin{aligned} \forall |\xi| < \theta: \ |\mathbb{E}_{\theta+\xi}[T] - \mathbf{T}_n(\theta,\xi)| &\leq \frac{2}{\theta} \sum_{k=n+1}^{\infty} (k+1) \left(\frac{|\xi|}{\theta}\right)^k \\ &= \frac{2}{\theta} \frac{(n+2) - (n+1)\frac{|\xi|}{\theta}}{\left(1 - \frac{|\xi|}{\theta}\right)^2} \left(\frac{|\xi|}{\theta}\right)^{n+1} \\ &= \frac{2}{\theta} \frac{1 + (n+1)(1-\rho)}{(1-\rho)^2} \rho^{n+1}, \end{aligned} \tag{4.19}$$

where, for simplicity, we set $\rho := |\xi|/\theta$.

Example 4.4. *In order to perform a numerical experiment, we consider the following values:*

$$\lambda_1 = \lambda_3 = \theta, \ \lambda_6 = 1, \ \lambda_2 = \lambda_4 = \frac{1}{2}, \ \lambda_5 = \frac{1}{5}, \ \lambda_7 = \frac{1}{3}.$$

Computing the coefficients of the Taylor polynomial is quite demanding and it is worth noting that the coefficients can alternatively be evaluated by simulation. Figure 4.2 shows the Taylor polynomial $\mathbf{T}_3(1,\xi)$ *of order* 3 *compared to the interpolation polynomial, with seven equidistant nodes, corresponding to* $\mathbb{E}_{1+\xi}[T]$*, in the range* $|\xi| \leq 0.6$*. As the Figure 4.2 shows, the difference between the two estimates is quite insignificant in the range* $|\xi| \leq 0.4$*. On the other hand, the relative error for the Taylor polynomial, according to (4.19), is below* 3.4% *in this range.*

4.4 Concluding Remarks

In this chapter we have extended the theory of weak differentiation to higher-order derivatives in order to construct a measure-valued differential calculus. This allows for studying analyticity related issues which, in turn, lead to Taylor series approximations for performance measures of parameter-dependent stochastic systems. Similar issues have been addressed in [7], [12], [24], [55].

The main result of this chapter, Theorem 5.5, which shows that products of weakly analytic measures are again weakly analytic, is the main theoretical tool for performing Taylor series approximations based on weak differentiation. As illustrated by Example 4.4, in practice, the exact calculation of the Taylor series coefficients is quite demanding and this seems to be the main pitfall of this method. Therefore, simulation of the Taylor series coefficients plays a key role in applying this method.

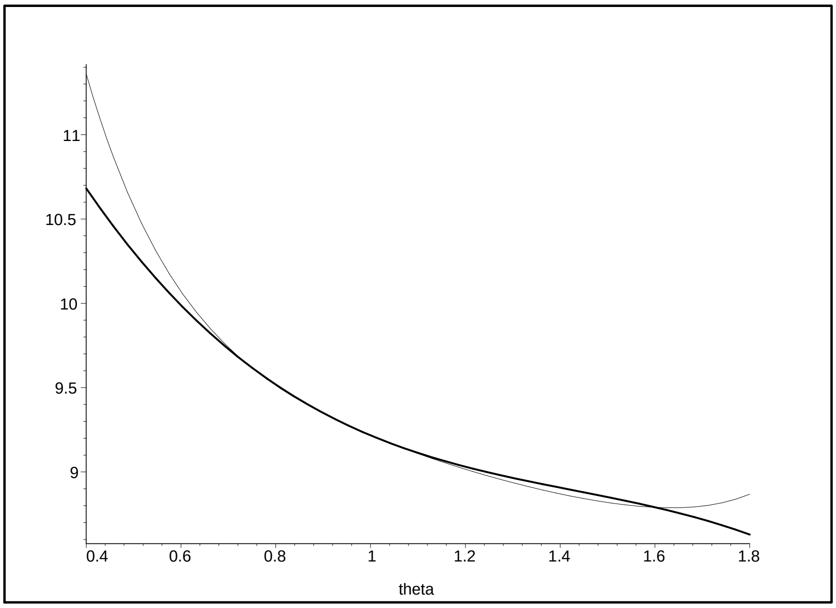

Fig. 4.2: The Taylor polynomial $\mathbf{T}_3(1, \xi)$ (thick line) compared to the interpolation polynomial, with seven equidistant nodes, corresponding to $\mathbb{E}_{1+\xi}[T]$ (thin line), in Example 4.4.

The gain of the method put forward in this chapter, however, comes from the fact that it is suitable for evaluating (it provides asymptotically unbiased estimators) the functional dependence between a performance measure of a certain stochastic system and some intrinsic parameter rather than approximating the value of the performance measure under consideration for some particular parameter θ, which can be easily achieved by using classical simulation.

As the numerical experiments have revealed, in many situations the Taylor series obtained by using weak differential calculus provide a quite good approximation for the true value (seen as a function in θ). In addition, we strongly believe that weak Taylor series are more efficient than interpolation-based methods, i.e., one simulates the performance of the system under consideration for some particular values of the parameter θ, in a given interval and then use interpolation and continuity properties of the corresponding interpolation operator to estimate the true functional dependence. The advantage of the method presented here comes from the fact that, while the complexity (the number of simulations) of the methods is comparable, the resulting estimates, when using Taylor series approximations, are prone to lower variance, i.e., faster convergence, and a very likely reason is that, unlike interpolation polynomials, Taylor polynomials do not involve, in general, divisions by small numbers.

While the above facts rely rather on intuition sustained by some unsystematic experiments, establishing accurate error bounds for the estimates, which lead to determining the convergence rates of the simulation process, and minimizing the errors by choosing convenient representations for the weak derivatives are topics for future research.

5. A CLASS OF NON-CONVENTIONAL ALGEBRAS WITH APPLICATIONS IN OR

In this chapter we apply the measure-valued differential calculus presented in Chapter 4 to distributions of random matrices in some special class of non-conventional algebras in order to construct Taylor series approximations for performance measures of stochastic dynamic systems whose dynamics can be modeled by a general matrix-vector multiplication.

5.1 Introduction

Throughout this chapter we consider stochastic systems whose time-evolution can be modeled as follows:

$$\forall k \geq 0: \; V(k+1) = X(k+1) \odot V(k), \tag{5.1}$$

where $\odot$ denotes a general matrix-vector multiplication, $V(k)$, for $k \geq 0$, is a finite dimensional vector denoting the k^{th} state of the system and $X(k)$, for $k \geq 1$, is a matrix of appropriate size describing the transition of the k^{th} to the $(k+1)^{st}$ state of the system. It follows that the k^{th} state of such a system can be expressed as

$$\forall k \geq 1: \; V(k) = X(k) \odot \ldots \odot X(1) \odot V(0), \tag{5.2}$$

provided that the matrix-vector multiplication $\odot$ is associative. That is, the evolution of the system is completely determined by the initial state $V(0)$ and the sequence of transitions $\{X(k) : k \geq 1\}$. This general model arises when dealing with Discrete Event Systems (DES), e.g., queueing networks, stochastic activity networks and stochastic Petri-nets, where the state dynamic can be modeled through a matrix-vector multiplication in either conventional, max-plus or min-plus algebra. For instance, the optimal cost problem in transportation networks leads to min-plus models whereas synchronization models lead to max-plus algebra. More concrete examples with concrete interpretations can also be found in [4], [13], [15], [29] and [45]. For time-homogenous, deterministic max-plus-linear systems, i.e., $X(k) = X$, for all $k \geq 0$, powerful tools exist for evaluating the system; see, e.g., [4], [29].

Assuming that the distributions of the input variables depend on some design parameter θ, this chapter deals with the problem of computing the expected value of the state vector $\mathbb{E}_\theta[V(k)]$ or, more generally, $\mathbb{E}_\theta[g(V(k))]$, for some cost-function g and a fixed horizon $k \geq 1$, as a function of θ. This problem is known to be notoriously difficult as exact formulae exist only for some special cases.

In the steady-state case, remarkable results have been obtained in [7] for the stationary waiting time in max-plus-linear queueing networks with Poisson arrival stream, using light-traffic approximation. These results have been extended to polynomially bounded performance measures in [1], [5], and explicit expressions for the moments, Laplace transforms and tail probabilities of waiting times are given in [2], [3]. Taylor series approximations have been successfully applied to control of max-plus-linear DES. Applications based on the concept of variability expansion can be found in [20], [59].

Here, we propose Taylor series approximations based on the measure-valued differential calculus developed in Chapter 4 and for ease of implementation we introduce a analogous differential calculus for random matrices which in practice is easier to work with.

The chapter is organized as follows. In Section 5.2 we define the concept of *topological algebra of matrices* for which we introduce the concept of weak differentiability, in Section 5.3, and construct a formal weak differential calculus in Section 5.4. Eventually, we illustrate the results by two examples, in Section 5.5.

5.2 Topological Algebras of Matrices

In this section we consider a separable, locally compact metric space $(\mathbb{S}, d)$ endowed with two binary associative operators, denoted by $\diamond$ and $*$, such that $(\mathbb{S}, \diamond)$ and $(\mathbb{S}, *)$ are monoids with unit elements $1_\diamond$ and 1_*, respectively. Assume further that $\diamond$ is commutative and $1_\diamond$ is *absorbing* for $*$, i.e.,

$$\forall s \in \mathbb{S} : \; 1_\diamond \, * \, s = s \, * \, 1_\diamond = 1_\diamond.$$

For integers $m, n \geq 1$ denote by $\mathfrak{M}_{m,n}(\mathbb{S})$ the set of m, n matrices with elements from $\mathbb{S}$. The generalized product of matrices $X \in \mathfrak{M}_{m,k}(\mathbb{S})$ and $Y \in \mathfrak{M}_{k,n}(\mathbb{S})$, denoted by $X \odot_{(\diamond \cdot *)} Y$ or simply $X \odot Y$, when no confusion occurs, is defined as follows:

$$[X \odot_{(\diamond \cdot *)} Y]_{ij} := (X_{i1} * Y_{1j}) \diamond (X_{i2} * Y_{2j}) \diamond \cdots \diamond (X_{ik} * Y_{kj}), \tag{5.3}$$

for each pair (i, j) with $1 \leq i \leq m, 1 \leq j \leq k$. Note that a "zero" element $\mathbf{0}_{(\diamond \cdot *)}$ can be introduced on $\mathfrak{M}_{m,n}$ by considering a matrix with all entries equal to $1_\diamond$. Moreover, if $m = n = k$ then $\odot$ defines an internal operation on $\mathfrak{M}_{n,n}$ and admits a neutral element, denoted by $\mathbf{I}_{(\diamond \cdot *)}$, which can be constructed just like in conventional algebra by setting all the entries of the matrix to $1_\diamond$ except from those on the main diagonal which are set to 1_*. We omit the subscript $(\diamond \cdot *)$ if no confusion occurs.

For each $m, n \geq 1$ the set $\mathfrak{M}_{m,n}(\mathbb{S})$ becomes a separable metric space when endowed with the metric $\wp_{m,n}$ given by

$$\forall X, Y : \; \wp_{m,n}(X, Y) := \max\{d(X_{ij}, Y_{ij}) : \; 1 \leq i \leq m, 1 \leq j \leq n\}. \tag{5.4}$$

In the sequel, we use the notations $\mathfrak{M}_n(\mathbb{S})$ for $\mathfrak{M}_{n,n}(\mathbb{S})$, $\wp_n$ for $\wp_{n,n}$, and omit specifying the underlying space $\mathbb{S}$, when no confusion occurs, by writing $\mathfrak{M}_{m,n}$ instead of $\mathfrak{M}_{m,n}(\mathbb{S})$.

Assume now that the mappings $\diamond$ and $*$ are bi-continuous with respect to d. It follows that for all $m, n, k \geq 1$ the mapping

$$\odot : \mathfrak{M}_{m,n} \times \mathfrak{M}_{n,k} \to \mathfrak{M}_{m,k}$$

is continuous if one endows $\mathfrak{M}_{m,n}$ with the metric $\wp_{m,n}$, for all $m, n \geq 1$. In particular, if $m = n = k$, then $\odot$ denotes an internal associative binary operation, i.e., $(\mathfrak{M}_n, \odot)$ is a monoid, which acts as a bi-continuous mapping with respect to the corresponding metric $\wp_n$ on $\mathfrak{M}_n$. In addition, we have $(\mathfrak{M}_1, \odot) = (\mathbb{S}, *)$ and, in general, $(\mathfrak{M}_{m,n}, \wp_{m,n})$, with $\wp_{m,n}$ defined by (5.4), for $m, n \geq 1$, is a metric space which inherits most of the topological properties of $(\mathbb{S}, d)$, such as separability and local compactness.

We synthesize the above construction into the following definition.

Definition 5.1. *We call the pair* $\mathcal{A} := (\{(\mathfrak{M}_{m,n}, \wp_{m,n}) : m, n \geq 1\}, \odot)$ *a topological algebra of matrices over the space* $(\mathbb{S}, d, \diamond, *)$ *if*

(i) $(\mathbb{S}, d)$ *is a separable, locally compact metric space,*

(ii) $(\mathbb{S}, \diamond)$ *and* $(\mathbb{S}, *)$ *are monoids with unit elements* $1_\diamond$ *and* 1_**, respectively,*

(iii) $\diamond$ *and* $*$ *are bi-continuous mappings with respect to* d*,*

(iv) $1_\diamond$ *is absorbing for* $*$*, i.e.,* $1_\diamond$ *acts as "zero" element,*

(v) $\diamond$ *is commutative and* $*$ *distributes over* $\diamond$*,*

(vi) for $m, n \geq 1$*,* $\mathfrak{M}_{m,n}$ *denotes the set of* m, n *matrices with entries in* $\mathbb{S}$*,*

(vii) for $m, n \geq 1$*,* $\wp_{m,n}$ *is defined as in (5.4),*

(viii) for $m, n, k \geq 1$*,* $\odot : \mathfrak{M}_{m,n} \times \mathfrak{M}_{n,k} \to \mathfrak{M}_{m,k}$ *is defined as in (5.3).*

By an *upper-bound* on a metric space we mean a real-valued, continuous, non-negative mapping. Let $\mathcal{A} = (\{(\mathfrak{M}_{m,n}, \wp_{m,n}) : m, n \geq 1\}, \odot)$ be a topological algebra of matrices over the space $(\mathbb{S}, d, \diamond, *)$.

We call the family $\|\cdot\| := \{\|\cdot\|_{m,n} : m, n \geq 1\}$ a *pseudo-norm* on $\mathcal{A}$ if

(i) for all $m, n \geq 1$, $\|\cdot\|_{m,n}$ is an upper-bound on $\mathfrak{M}_{m,n}$,

(ii) the family $\|\cdot\|$ satisfies either: for each $m, n, k \geq 1$,

$$\forall X \in \mathfrak{M}_{m,n}, Y \in \mathfrak{M}_{n,k} : \ \|X \odot Y\|_{m,k} \leq \|X\|_{m,n} + \|Y\|_{n,k}$$

or for each $m, n, k \geq 1$,

$$\exists \gamma > 0 : \ \forall X \in \mathfrak{M}_{m,n}, Y \in \mathfrak{M}_{n,k} : \ \|X \odot Y\|_{m,k} \leq \gamma \cdot \|X\|_{m,n} \cdot \|Y\|_{n,k}.$$

The pseudo-norm will be called *additive* (resp. *multiplicative*) according to the operation in the right-hand side and we say that $(\mathcal{A}, \|\cdot\|)$ is a *pseudo-normed topological algebra of matrices*. For simplicity we use the notation $\|\cdot\|$ for $\|\cdot\|_{m,n}$. Note that, by definition, the mapping $\|\cdot\|_{m,n}$ is continuous with respect to $\wp_{m,n}$, for all $m, n \geq 1$, and if $m = n$ then $\|\cdot\|_n$ satisfies

$$\forall X, Y \in \mathfrak{M}_n : \ \|X \odot Y\|_n \leq \|X\|_n + \|Y\|_n,$$

if $\|\cdot\|_n$ is additive or

$$\forall X, Y \in \mathfrak{M}_n : \ \|X \odot Y\|_n \leq \gamma \cdot \|X\|_n \cdot \|Y\|_n,$$

if $\|\cdot\|$ is multiplicative. We call the pair $(\mathfrak{M}_n, \wp_n, \odot, \|\cdot\|)$ a *pseudo-normed topological monoid*. In addition, note that $\|\cdot\|_n$ is by no means a norm on $\mathfrak{M}_n$ since, in general, $\mathfrak{M}_n$ can not be organized as a linear space. Moreover, $\|X\|_n = 0$ does not imply in general $X = \mathbf{0}$.

Example 5.1. *We enumerate here some classical examples of such structures arising in modeling theory.*

(i) $\diamond = +$, $* = \times$ *with* $1_\diamond = 0$ *and* $1_* = 1$. *This is noticeably the conventional algebra setting and we choose* $\mathbb{S} = \mathbb{R}$ *endowed with the usual metric. The mapping* $\|\cdot\| : \mathfrak{M}_{m,n} \to \mathbb{R}$ *defined as*

$$\|X\| = \max_{(i,j)} |X_{ij}|$$

is an upper-bound on $\mathfrak{M}_{m,n}$. *In addition, for* $X \in \mathfrak{M}_{m,n}$ *and* $Y \in \mathfrak{M}_{n,k}$ *it holds that*

$$\forall (i,j) : \ |[X \odot Y]_{ij}| \le \sum_{l=1}^{n} |X_{il}| \cdot |Y_{lj}| \le n \cdot \|X\| \cdot \|Y\|.$$

Taking the maximum with respect to (i,j) *in the left-hand side, we obtain*

$$\forall X, Y : \ \|X \odot Y\| \le n \cdot \|X\| \cdot \|Y\|.$$

Therefore, $\|\cdot\|$ *is a multiplicative pseudo-norm for the conventional algebra of matrices.*

(ii) $\diamond = \max$, $* = +$ *with* $1_\diamond = -\infty$ *and* $1_* = 0$, *i.e., we are dealing with the so called max-plus algebra. We take* $\mathbb{S} = \mathbb{R} \cup \{-\infty\}$. *An appropriate metric on* $\mathbb{S}$ *is given by*

$$d(x,y) = \begin{cases} \frac{|x-y|}{1+|x-y|}, & x, y \in \mathbb{R}, \\ 1, & x \in \mathbb{R}, y = -\infty \text{ or } y \in \mathbb{R}, x = -\infty, \\ 0, & x = y = -\infty. \end{cases} \tag{5.5}$$

Obviously, $d(x,y) = d(y,x) > 0$ *and* $d(x,x) = 0$, *for all* $x, y \in \mathbb{S}$. *To see that* d *is a metric on* $\mathbb{S}$, *one has to show that* d *satisfies the triangle inequality, i.e.,*

$$\forall x, y, z \in \mathbb{S} : \ d(x,y) \le d(x,z) + d(z,y).$$

This is not straightforward and it can be proved by considering several cases. Here we sketch the proof by considering the two non-trivial cases.
(a) If $x, y, z \in \mathbb{R}$ *then taking into account that* $(x,y) \mapsto |x-y|$ *defines a metric on* $\mathbb{R}$, *by the triangle inequality we have*

$$|x-y| \le |x-z| + |z-y|.$$

In addition, the mapping $f(t) = \frac{t}{1+t}$, *for* $t \ge 0$, *is nondecreasing and simple algebra shows that it satisfies*

$$\forall t_1, t_2 \ge 0 : \ f(t_1 + t_2) \le f(t_1) + f(t_2).$$

Then, it follows that for each $x, y, z \in \mathbb{R}$ it holds that

$$d(x,y) = f(|x-y|) \le f(|x-z|+|z-y|) \le f(|x-z|)+f(|z-y|) = d(x,z)+d(z,y).$$

(b) If $x, y \in \mathbb{R}$ and $z = -\infty$ then we have

$$d(x,y) = \frac{|x-y|}{1+|x-y|} < 1 < 2 = d(x,z) + d(z,y).$$

Therefore, d in (5.5) defines a metric on $\mathbb{S}$. For $X \in \mathfrak{M}_{m,n}$ set

$$\|X\| = \begin{cases} \max\limits_{(i,j)} |X_{ij}|, & \text{if } \exists\,(i,j) \text{ s.t. } X_{ij} \neq -\infty, \\ 0, & \text{otherwise.} \end{cases}$$

Since the trace of d on $\mathbb{R}$ is equivalent[1] *to the usual metric on $\mathbb{R}$, it follows that $\|\cdot\|$ is an upper-bound on $\mathfrak{M}_{m,n}$. Moreover, for $X \in \mathfrak{M}_{m,n}$ and $Y \in \mathfrak{M}_{n,k}$ it satisfies*

$$\forall (i,j):\ |[X \odot Y]_{ij}| = \left| \max_{1\le l\le n} (X_{il} + Y_{lj}) \right| \le \|X\| + \|Y\|. \tag{5.6}$$

Taking the maximum over all (i,j) in the left-hand side of (5.6), yields

$$\forall X, Y:\ \|X \odot Y\| \le \|X\| + \|Y\|,$$

which shows that $\|\cdot\|$ is an additive pseudo-norm for the max-plus algebra.

(iii) $\diamond = \min$, $* = +$ *with* $1_\diamond = \infty$ *and* $1_* = 0$*, i.e., we obtain the min-plus algebra of matrices. Set $\mathbb{S} = \mathbb{R} \cup \{\infty\}$ and define the metric d on $\mathbb{S}$, as follows:*

$$d(x,y) = \begin{cases} \frac{|x-y|}{1+|x-y|}, & x, y \in \mathbb{R}, \\ 1, & x \in \mathbb{R}, y = \infty, \\ 1, & y \in \mathbb{R}, x = \infty. \end{cases}$$

In addition, for $X \in \mathfrak{M}_{m,n}$ let us define

$$\|X\| = \begin{cases} \max\limits_{(i,j)} |X_{ij}|, & \text{if } \exists i, j \text{ s.t. } X_{ij} \neq \infty, \\ 0, & \text{otherwise.} \end{cases}$$

Following the same line of argument as in the above example, we conclude that d is a metric on $\mathbb{S}$ and $\|\cdot\|$ is an additive pseudo-norm for the topological algebra of matrices induced by $(\mathbb{S}, d, \min, +)$.

(iv) $\diamond = \max$, $* = \times$ *with* $1_\diamond = -\infty$ *and* $1_* = 1$*. Set $\mathbb{S} = \mathbb{R} \cup \{-\infty\}$, where we agree that*

$$\forall x \in \mathbb{R}:\ -\infty \times x = x \times -\infty = -\infty,$$

[1] i.e., both metrics generate the same topology.

i.e., $-\infty$ is absorbing for $\times$. We choose d and $\|\cdot\|$ just like in (ii) above. In order to show that $\|\cdot\|$ is a pseudo-norm for this algebra of matrices we note that for $X \in \mathfrak{M}_{m,n}$ and $Y \in \mathfrak{M}_{n,k}$ it holds that

$$\forall (i,j): \; |[X \odot Y]_{ij}| = \left| \max_{1 \leq l \leq n} (X_{il} \cdot Y_{lj}) \right| \leq \|X\| \cdot \|Y\|.$$

This leads to $\|X \odot Y\| \leq \|X\| \cdot \|Y\|$, for each X, Y for which the matrix multiplication makes sense, i.e., $\|\cdot\|$ is a multiplicative pseudo-norm.

(v) $\diamond = \min$, $* = \times$ *with* $1_\diamond = \infty$ *and* $1_* = 0$. *We choose* $\mathbb{S} = \mathbb{R} \cup \{\infty\}$ *and agree that*

$$\forall x \in \mathbb{R}: \; \infty \times x = x \times \infty = \infty.$$

We also choose d and $\|\cdot\|$ exactly as in (iii). Following the same arguments as in the above example we obtain that $\|\cdot\|$ is a multiplicative pseudo-norm.

5.3 $\mathcal{D}_p$-Differentiability

In many mathematical models, which can be described by one of the settings enumerated in Example 5.1, one is interested to assess the behavior of the integrals of the moments $\|X\|^p$, for $p \geq 1$. An efficient way to do this is to consider a particular set of test-functions $\mathcal{D}_p$, i.e., the class of polynomially bounded mappings, to be introduced in Section 5.3.1. In Section 5.3.2 we discuss the concept of $\mathcal{D}_p$-differentiability of random matrices.

5.3.1 $\mathcal{D}_p$-spaces

Let $(\mathbb{X}, d)$ be a metric space with upper-bound $\|\cdot\|$ and for $p \geq 1$ let us denote by v_p the mapping defined as

$$\forall x \in \mathbb{X}: \; v_p(x) = \max\{1, \|x\|^p\}. \tag{5.7}$$

Note that $v_p \in \mathcal{C}^+(\mathbb{X})$. In addition, if $(\mathcal{D}(\mathbb{X}), v_p)$ is a Banach base on $\mathbb{X}$, we define $\mathcal{D}_p(\mathbb{X})$ or $\mathcal{D}_p$, when no confusion occurs, as follows:

$$\mathcal{D}_p(\mathbb{X}) := [\mathcal{D}(\mathbb{X})]_{v_p} = \left\{ g \in \mathcal{D}(\mathbb{X}) : \sup_{x \in \mathbb{X}} \frac{|g(x)|}{v_p(x)} < \infty \right\}. \tag{5.8}$$

Note that the spaces $\mathcal{D}_p$, for $p \geq 0$, are Banach spaces and enjoy the property that $q < p$ implies $\mathcal{D}_q \subset \mathcal{D}_p$. Indeed, note first that for $q < p$ it holds that

$$\forall x \in \mathbb{X}: \; v_q(x) = \max\{1, \|x\|^q\} \leq \max\{1, \|x\|^p\} = v_p(x) \tag{5.9}$$

and consequently, for $g \in \mathcal{D}_q$, it follows that

$$\forall x \in \mathbb{X}: \; |g(x)| \leq \|g\|_{v_q} v_q(x) \leq \|g\|_{v_q} v_p(x),$$

i.e., for $q < p$ we have $\|g\|_{v_p} \leq \|g\|_{v_q}$, for all $g \in \mathcal{D}(\mathbb{X})$. The next result provides the main technical tool for dealing with $\mathcal{D}_p$-spaces.

Lemma 5.1. *Let* $(\mathbb{X}, d_{\mathbb{X}})$, $(\mathbb{Y}, d_{\mathbb{Y}})$ *and* $(\mathbb{Z}, d_{\mathbb{Z}})$, *be metric spaces equipped with upper-bounds* $\|\cdot\|_{\mathbb{X}}$, $\|\cdot\|_{\mathbb{Y}}$ *and* $\|\cdot\|_{\mathbb{Z}}$, *respectively. Let* $h : \mathbb{X} \times \mathbb{Y} \to \mathbb{Z}$ *be continuous and define* $w : \mathbb{X} \times \mathbb{Y} \to \mathbb{R}$ *as*

$$\forall x \in \mathbb{X}, y \in \mathbb{Y} : \; w(x, y) = v_p(x) v_p(y);$$

see (5.7) for a definition of v_p. *If any of the following conditions holds:*

(α) there exist constants $C_{\mathbb{X}}, C_{\mathbb{Y}} > 0$ *such that*

$$\forall x \in \mathbb{X}, y \in \mathbb{Y} : \; \|h(x, y)\|_{\mathbb{Z}} \leq C_{\mathbb{X}} \|x\|_{\mathbb{X}} + C_{\mathbb{Y}} \|y\|_{\mathbb{Y}},$$

(β) there exist $C > 0$ *such that*

$$\forall x \in \mathbb{X}, y \in \mathbb{Y} : \; \|h(x, y)\|_{\mathbb{Z}} \leq C \|x\|_{\mathbb{X}} \|y\|_{\mathbb{Y}},$$

then $\|g \circ h\|_w < \infty$ *for any* $g \in \mathcal{D}_p(\mathbb{Z})$.

Proof. First, note that the conclusion reduces to

$$\sup_{(x,y)} \frac{\|h(x, y)\|_{\mathbb{Z}}^p}{v_p(x) v_p(y)} < \infty,$$

since, by hypothesis, for $z \in \mathbb{Z}$ we have $|g(z)| \leq \|g\|_{v_p} v_p(z)$.

If (α) holds true then for each $x \in \mathbb{X}$ and $y \in \mathbb{Y}$ it holds that

$$\|h(x, y)\|_{\mathbb{Z}}^p \leq \sum_{i=0}^{p} \binom{p}{i} C_{\mathbb{X}}^i \|x\|_{\mathbb{X}}^i C_{\mathbb{Y}}^{p-i} \|y\|_{\mathbb{Y}}^{p-i} \leq \sum_{i=0}^{p} \binom{p}{i} C_{\mathbb{X}}^i C_{\mathbb{Y}}^{p-i} v_p(x) v_p(y), \tag{5.10}$$

where the inequality in (5.10) follows from (5.9). Therefore, from (5.10) we conclude that

$$\forall x \in \mathbb{X}, y \in \mathbb{Y} : \; \frac{\|h(x, y)\|_{\mathbb{Z}}^p}{v_p(x) v_p(y)} \leq (C_{\mathbb{X}} + C_{\mathbb{Y}})^p.$$

Assume now that (β) holds true. Then, for $x \in \mathbb{X}$, $y \in \mathbb{Y}$ it holds that

$$\|h(x, y)\|_{\mathbb{Z}}^p \leq C \|x\|_{\mathbb{X}}^p \|y\|_{\mathbb{Y}}^p \leq C v_p(x) v_p(y).$$

Consequently, we have

$$\forall x \in \mathbb{X}, y \in \mathbb{Y} : \; \frac{\|h(x, y)\|_{\mathbb{Z}}^p}{v_p(x) v_p(y)} \leq C.$$

Therefore, if (α) holds true then $\|g \circ h\|_w \leq (C_{\mathbb{X}} + C_{\mathbb{Y}})^p \|g\|_{v_p}$ whereas if (β) holds true we have $\|g \circ h\|_w \leq C \|g\|_{v_p}$, which concludes the proof. □

In the following, we let $\mathbb{X} = \mathfrak{M}_{m,n}$, i.e., $\mathbb{X}$ consists of (m, n) matrices in some pseudo-normed topological algebra. Recall that $\mathfrak{M}_{m,n}$ becomes a separable, locally compact metric space, when endowed with the metric $\wp_{m,n}$, so that the theory of weak differentiation can be easily adapted to this setting. The next result, which is an immediate consequence of Lemma 5.1, will put forward a remarkable property of $\mathcal{D}_p$-spaces which will be crucial for introducing a weak differential calculus for random matrices on pseudo-normed topological algebras.

Corollary 5.1. *Let* $(\mathcal{A}, \|\cdot\|) = (\{(\mathfrak{M}_{m,n}, \wp_{m,n}) : m, n \geq 1\}, \odot, \|\cdot\|)$ *be a pseudo-normed topological algebra of matrices and for each* $n, m \geq 1$ *and* $p \geq 0$ *let* v_p *and* $\mathcal{D}_p(\mathfrak{M}_{m,n})$ *be defined as in (5.7) and (5.8), respectively. Then, for each* $m, n, k \geq 1$ *and* $g \in \mathcal{D}_p(\mathfrak{M}_{m,k})$ *the mapping*

$$\forall X \in \mathfrak{M}_{m,n}, Y \in \mathfrak{M}_{n,k} : \ (X, Y) \mapsto g(X \odot Y)$$

belongs to $[\mathcal{D}(\mathfrak{M}_{m,n} \times \mathfrak{M}_{n,k})]_w$ *where* $w : \mathfrak{M}_{m,n} \times \mathfrak{M}_{n,k} \to \mathbb{R}$ *is defined as*

$$\forall X \in \mathfrak{M}_{m,n}, Y \in \mathfrak{M}_{n,k} : \ w(X, Y) := v_p(X) \cdot v_p(Y).$$

Proof. The proof follows from Lemma 5.1. Indeed, if we let $\mathbb{X} = \mathfrak{M}_{m,n}$, $\mathbb{Y} = \mathfrak{M}_{n,k}$ and $\mathbb{Z} = \mathfrak{M}_{m,k}$ with the usual $\wp$ metric and upper-bound and $h = \odot$ and note that h is continuous then Lemma 5.1 concludes the proof. □

5.3.2 $\mathcal{D}_p$-Differentiability for Random Matrices

Let $X \in \mathfrak{M}_{m,n}$ be a random matrix defined on some probability space $(\Omega, \mathcal{K})$ having distribution μ_θ, for $\theta \in \Theta$, i.e., $X : \Omega \to \mathfrak{M}_{m,n}$ is measurable and for each $\theta \in \Theta$ there exists some probability measure $\mathbb{P}_\theta$ on $\mathcal{K}$ such that

$$\forall \theta \in \Theta : \ \mathbb{P}_\theta(X \in \mathfrak{N}) = \mu_\theta(\mathfrak{N}),$$

for each Borel subset $\mathfrak{N}$ of $\mathfrak{M}_{m,n}$. Recall that if μ_θ is weakly $\mathcal{D}_p$-differentiable, with derivative $(c_\theta, \mu_\theta^+, \mu_\theta^-)$, it follows that

$$\forall g \in \mathcal{D}_p : \ \frac{d}{d\theta}\mathbb{E}_\theta\left[g(X)\right] = c_\theta \, \mathbb{E}_\theta\left[g(X^+) - g(X^-)\right], \tag{5.11}$$

where $\mathbb{E}_\theta$ is an expectation operator consistent with $X \sim \mu_\theta$ and $X^\pm \sim \mu_\theta^\pm$.

Assume now that X and Y are stochastically independent random matrices such that their distributions are $\mathcal{D}_p$-differentiable. In order to study differentiability properties of the distribution of their product $X \odot Y$ one can apply Theorem 2.3 to the distributions of X and Y and obtain the following result.

Theorem 5.1. *Let* $X \in \mathfrak{M}_{m,n}$ *and* $Y \in \mathfrak{M}_{n,k}$ *be stochastically independent random matrices with distributions* μ_θ *and* η_θ*, respectively. Assume further that the distributions* μ_θ *and* η_θ *are* $\mathcal{D}_p$*-differentiable, having weak derivatives* $(c_\theta^\mu, \mu_\theta^+, \mu_\theta^-)$ *and* $(c_\theta^\eta, \eta_\theta^+, \eta_\theta^-)$*, respectively. Then the distribution of the product* $X \odot Y$ *is again* $\mathcal{D}_p$*-differentiable and for each* $g \in \mathcal{D}_p(\mathfrak{M}_{m,k})$ *it holds that*

$$\frac{d}{d\theta}\mathbb{E}_\theta[g(X \odot Y)] = c_\theta^\mu \mathbb{E}\left[g(X^+ \odot Y) - g(X^- \odot Y)\right] + c_\theta^\eta \mathbb{E}_\theta\left[g(X \odot Y^+) - g(X \odot Y^-)\right],$$

Proof. It follows from Theorem 2.3 that $\mu_\theta \times \eta_\theta$ is $[\mathcal{D}(\mathfrak{M}_{m,n} \times \mathfrak{M}_{n,k})]_w$-differentiable, where

$$\forall X \in \mathfrak{M}_{m,n}, Y \in \mathfrak{M}_{n,k} : \ w(X, Y) := v_p(X) \cdot v_p(Y)$$

and (5.12) holds true for all g chosen such that the mapping

$$\forall X \in \mathfrak{M}_{m,n}, Y \in \mathfrak{M}_{n,k} : \ (X, Y) \mapsto g(X \odot Y)$$

belongs to $[\mathcal{D}(\mathfrak{M}_{m,n} \times \mathfrak{M}_{n,k})]_w$. Therefore, from Corollary 5.1 we conclude that the distribution of $X \odot Y$ is $\mathcal{D}_p(\mathfrak{M}_{m,k})$-differentiable and (5.12) holds true for $g \in \mathcal{D}_p(\mathfrak{M}_{m,k})$. □

Remark 5.1. *Since, throughout this chapter, our focus is on random matrices rather than on their distributions we will use in the remainder of this chapter the notation $\mathbb{E}[g(X_\theta)]$ instead of $\mathbb{E}_\theta[g(X)]$, which would be consistent with the theory in Chapter 2, in order to emphasize the dependence on θ. In this notation (5.11) will be re-written as*

$$\forall g \in \mathcal{D}_p : \ \frac{d}{d\theta}\mathbb{E}\left[g(X_\theta)\right] = c_\theta \, \mathbb{E}\left[g(X_\theta^+) - g(X_\theta^-)\right], \tag{5.12}$$

and a weak derivative of X_θ will be formally denoted by $X'_\theta = (c_\theta, X_\theta^+, X_\theta^-)$.

We say that the random matrix X_θ is $\mathcal{D}_p$-differentiable with respect to θ if its distribution is $\mathcal{D}_p$-differentiable. Consequently, we call the triple $(c_\theta, X_\theta^+, X_\theta^-)$ a weak $\mathcal{D}_p$-derivative of X_θ. In the same vein we define higher-order differentiation for random matrices. More specifically, we say that X_θ is n times $\mathcal{D}_p$-differentiable, for $n \geq 1$, if its distribution is n times $\mathcal{D}_p$-differentiable. It follows that there exists $c_\theta^{(n)} > 0$ and random variables $X_\theta^{(n\pm)}$ such that

$$\forall g \in \mathcal{D}_p : \ \frac{d^n}{d\theta^n}\mathbb{E}[g(X_\theta)] = c_\theta^{(n)}\mathbb{E}\left[g\left(X_\theta^{(n+)}\right) - g\left(X_\theta^{(n-)}\right)\right]. \tag{5.13}$$

Therefore, we call the triple

$$X_\theta^{(n)} := \left(c_\theta^{(n)}, X_\theta^{(n+)}, X_\theta^{(n-)}\right) \tag{5.14}$$

a n^{th}-order weak derivative of X_θ and we set

$$X_\theta^{(n)} := (1, X_\theta, X_\theta),$$

if the n^{th}-order derivative is not significant.

Example 5.2. *We revisit Example 2.4. Assume that X_θ is a Bernoulli distributed random variable, with point masses $A, B \in \mathfrak{M}_{m,n}$ and parameter $\theta \in [0,1]$, such that $X_0 = A$. If we interpret A as a random variable having Dirac distribution δ_A then an instance of a weak-derivative of X_θ is given by $X'_\theta = (1, B, A)$. Therefore, it holds that*

$$\forall g : \ \frac{d}{d\theta}\mathbb{E}[g(X_\theta)] = \mathbb{E}[g(B) - g(A)] = g(B) - g(A).$$

Furthermore, by Example 2.4 it follows that higher-order derivatives $X_\theta^{(n)}$, for $n \geq 2$, are not significant.

Note that the representation of the n^{th}-order derivative $X_\theta^{(n)}$ in (5.14) is not unique. However, by definition, any triplet representation of $X_\theta^{(n)}$ should satisfy (5.13). Moreover, Theorem 5.1 can be re-phrased as: "If X_θ and Y_θ are stochastically independent, $\mathcal{D}_p$-differentiable random matrices then the product $X_\theta \odot Y_\theta$ is $\mathcal{D}_p$-differentiable as well." In addition, the derivative $\frac{d}{d\theta}\mathbb{E}[g(X_\theta \odot Y_\theta)]$ can be evaluated according to (5.12).

On the other hand, provided that the product $X_\theta \odot Y_\theta$ is $\mathcal{D}_p$-differentiable, it would be desirable to have a formula such as

$$(X_\theta \odot Y_\theta)' = X'_\theta \odot Y_\theta + X_\theta \odot Y'_\theta. \tag{5.15}$$

Unfortunately, an equation like (5.15) does not make sense since the weak derivative of a random matrix X_θ is not a matrix anymore and has an algebraic meaning when identified with a triple $(c_\theta, X_\theta^+, X_\theta^-)$, where $c_\theta > 0$ and $X_\theta^\pm$ are again random matrices. Consequently, the expression in the right-hand side in (5.15) has no meaning. Therefore, in order to develop a differential calculus similar to the classical analysis, i.e., to establish a connection between (5.12) and (5.15), we need to embed the algebra of matrices into a richer one, where $\odot$ multiplication between random matrices and their derivatives makes sense. Moreover, the extended algebra should be consistent with the original one, i.e., the extended $\odot$ multiplication should coincide with the original $\odot$ multiplication, when restricted to simple matrices.

Motivated by the above remarks, we will introduce in Section 5.4 an extended algebra where the definition of the derivatives, as given by (5.14), is correct and where equalities such as (5.15) have a precise interpretation.

5.4 A Formal Differential Calculus for Random Matrices

Throughout this section we consider a pseudo-normed topological algebra of matrices

$$(\mathcal{A}, \|\cdot\|) = (\{(\mathfrak{M}_{m,n}, \wp_{m,n}) : m, n \geq 1\}, \odot, \|\cdot\|)$$

and for $m, n \geq 1$ and $p \geq 0$ we consider the mapping v_p and the space $\mathcal{D}_p = [\mathcal{D}]_{v_p}$ as defined by (5.7) and (5.8), respectively. Since in applications working with random variables is often more natural than working with their distributions (measures), we develop in the following a weak $\mathcal{D}_p$-differential calculus for random matrices in the algebra $(\mathcal{A}, \|\cdot\|)$. Starting point of our analysis will be Theorem 5.1 which asserts that $\mathcal{D}_p$-differentiability of random elements X_θ and Y_θ is inherited by the $\odot$ product $X_\theta \odot Y_\theta$ in a pseudo-normed topological algebra of matrices. In Section 5.4.1 we construct an extension $\mathcal{A}_*$ of the algebra $\mathcal{A}$ and in Section 5.4.2 we show that a weak differential calculus, similar to the classical one, holds true on the extended algebra $\mathcal{A}_*$.

5.4.1 The Extended Algebra of Matrices

Let $m, n \geq 1$ and denote by $\overline{\mathfrak{M}}_{m,n}$ the set of all finite sequences of triples (c, A, B), with $c \in \mathbb{R}_+$ and $A, B \in \mathfrak{M}_{m,n}$. A generic element of $\overline{\mathfrak{M}}_{m,n}$ is thus given by

$$\tau = \big((c_1, A_1, B_1), (c_2, A_2, B_2), \ldots, (c_n, A_n, B_n)\big),$$

where $n = n_\tau < \infty$ is called the *length* of τ. If τ is of length one, i.e., $n_\tau = 1$, we call it *elementary.* Note that the weak derivative of a random matrix is elementary in $\overline{\mathfrak{M}}_{m,n}$.

On $\overline{\mathfrak{M}}_{m,n}$ we introduce the addition, denoted by $+$, as the concatenation of strings. For example, let $\tau \in \overline{\mathfrak{M}}_{m,n}$ be given by $\tau = (\tau_i : 1 \leq i \leq n)$, with τ_i elementary, for each $1 \leq i \leq n$. Then we write $\tau = \sum_{i=i}^n \tau_i$. More generally, for $\sigma, \tau \in \overline{\mathfrak{M}}_{m,n}$, the application of the $+$ operator yields

$$\sigma + \tau = (\sigma_1, \ldots, \sigma_{n_\sigma}, \tau_1, \ldots, \tau_{n_\tau}) = \sum_{i=1}^{n_\sigma} \sigma_i + \sum_{j=1}^{n_\tau} \tau_j.$$

For an elementary $\tau = (c, A, B) \in \overline{\mathfrak{M}}_{m,n}$ we define the conjugate

$$\bar{\tau} := (c, B, A)$$

and extend it to general $\tau = (\tau_1, \cdots, \tau_n)$ as follows: $\bar{\tau} := (\bar{\tau}_1, \cdots, \bar{\tau}_n)$.

On $\overline{\mathfrak{M}}_{m,n}$ we introduce a scalar multiplication as follows: for elementary $\tau = (c, A, B)$ and a real number r we set $r \cdot \tau = (r \cdot c, A, B)$ and extend it to general τ such that it distributes over $+$, i.e., $r \cdot \tau = \sum_{i=1}^{n_\tau} r \cdot \tau_i$. Next, we introduce multiplication, denoted also by $\odot$, as follows[2]:

$$\sigma \odot \tau := c^\sigma c^\tau \cdot ((1, A^\sigma \odot A^\tau, \mathbf{0}), (1, B^\sigma \odot B^\tau, \mathbf{0}), (1, \mathbf{0}, A^\sigma \odot B^\tau), (1, \mathbf{0}, B^\sigma \odot A^\tau)),$$

for elementary $\sigma = (c^\sigma, A^\sigma, B^\sigma) \in \overline{\mathfrak{M}}_{m,n}$ and $\tau = (c^\tau, A^\tau, B^\tau) \in \overline{\mathfrak{M}}_{n,k}$ and we extend this operation to general elements via additivity. Specifically, if $\sigma = (\sigma_1, \cdots, \sigma_{n_\sigma}) \in \overline{\mathfrak{M}}_{m,n}$ and $\tau = (\tau_1, \cdots, \tau_{n_\tau}) \in \overline{\mathfrak{M}}_{n,k}$ then we set

$$\sigma \odot \tau = \sum_{i=1}^{n_\sigma} \sum_{j=1}^{n_\tau} \sigma_i \odot \tau_j.$$

Finally, we embed $\mathfrak{M}_{m,n}$ into $\overline{\mathfrak{M}}_{m,n}$ via a monomorphism ι given by

$$X^\iota = \iota(X) = (1, X, \mathbf{0})$$

and we define the *ι-extension* g^ι of a function $g : \mathfrak{M}_{m,n} \to \mathbb{R}$ in the following way: for

$$\sigma = ((c_1, A_1, B_1), (c_2, A_2, B_2), \ldots, (c_{n_\sigma}, A_{n_\sigma}, B_{n_\sigma})) \in \overline{\mathfrak{M}}_{m,n}$$

we set

$$g^\iota(\sigma) = \sum_{i=1}^{n} c_i \left(g(A_i) - g(B_i) \right).$$

Simple manipulations on the above introduced operations show that g^ι is linear with respect to addition and homogenous with respect to scalar multiplication, i.e., for any $g : \mathfrak{M}_{m,n} \to \mathbb{R}$, $\sigma, \tau \in \overline{\mathfrak{M}}_{m,n}$, and $c_\sigma, c_\tau \in \mathbb{R}$ it holds that

$$g^\iota(c_\sigma \cdot \sigma + c_\tau \cdot \tau) = c_\sigma \cdot g^\iota(\sigma) + c_\tau \cdot g^\iota(\tau).$$

In addition, using the properties of the morphism ι we deduce that

$$\sigma \odot \tau = c^\sigma c^\tau \left[(A^\sigma \odot A^\tau)^\iota + (B^\sigma \odot B^\tau)^\iota + \overline{(A^\sigma \odot B^\tau)^\iota} + \overline{(B^\sigma \odot A^\tau)^\iota} \right].$$

The set $\overline{\mathfrak{M}}_{m,n}$ can be embedded into the product space

$$\Sigma_{m,n} := (\mathbb{R} \times \mathfrak{M}_{m,n} \times \mathfrak{M}_{m,n})^{\mathbb{N}},$$

i.e., the space of all (infinite) sequences of triples, via the morphism

$$\forall \tau := (\tau_1, \ldots, \tau_{n_\tau}) \in \overline{\mathfrak{M}}_{m,n} : \ \tau \mapsto (\tau_1, \ldots, \tau_{n_\tau}, \mathbf{0}^\iota, \ldots) = \tau + \mathbf{0}^\iota + \ldots.$$

[2] Recall that $\mathbf{0}$ denotes the "zero" element on $\mathfrak{M}_{m,n}$

Then, $\overline{\mathfrak{M}}_{m,n}$ is isomorphic to the subset of $\Sigma_{m,n}$ consisting of all finite sequences and, for convenience, we identify any element τ with its image.

On the space $\Sigma_{m,n}$ one can introduce a metric $\Lambda_{m,n}$, in the following way: for elementary $\sigma = (c^\sigma, A^\sigma, B^\sigma)$ and $\tau = (c^\tau, A^\tau, B^\tau)$ we set

$$\Lambda_{m,n}(\sigma, \tau) := \max\{|c^\sigma - c^\tau|, \wp_{m,n}(A^\sigma, A^\tau), \wp_{m,n}(B^\sigma, B^\tau)\},$$

and extend $\Lambda_{m,n}$ to general elements by letting

$$\Lambda_{m,n}(\sigma, \tau) := \sum_{n=i}^{\infty} \frac{1}{2^i} \frac{\Lambda_{m,n}(\sigma_i, \tau_i)}{1 + \Lambda_{m,n}(\sigma_i, \tau_i)},$$

for $\sigma = (\sigma_i : i \geq 1)$ and $\tau = (\tau_i : i \geq 1)$. It turns out that $\Lambda_{m,n}$ is indeed a metric on $\Sigma_{m,n}$, so that $(\Sigma_{m,n}, \Lambda_{m,n})$ is a metric space. Therefore, if we identify $\overline{\mathfrak{M}}_{m,n}$ with a subset of $\Sigma_{m,n}$ then $(\overline{\mathfrak{M}}_{m,n}, \Lambda_{m,n})$ becomes a metric space such that $g : \mathfrak{M}_{m,n} \to \mathbb{R}$ is continuous (resp. Borel measurable). It follows that the ι-extension g^ι of g is continuous (resp. Borel measurable) on $\overline{\mathfrak{M}}_{m,n}$.

The structure

$$\mathcal{A}_* := (\{(\overline{\mathfrak{M}}_{m,n}, \Lambda_{m,n}) : m, n \geq 1\}, \odot)$$

will be called the $*$-extension of $\mathcal{A} = (\{(\mathfrak{M}_{m,n}, \wp_{m,n}) : m, n \geq 1\}, \odot)$, or the extended algebra, for short. Unfortunately, note that $\overline{\mathfrak{M}}_{m,n}$ has very poor algebraic properties. For example, the addition fails to be commutative and, moreover, does not admit a neutral (zero) element. Consequently, the $\odot$ multiplication on the extended algebra $\mathcal{A}_*$ is not a proper extension of the original $\odot$ multiplication on $\mathcal{A}$. To see this, note that for $X \in \overline{\mathfrak{M}}_{m,n}$ and $Y \in \overline{\mathfrak{M}}_{n,k}$ we have

$$\begin{aligned}
(X^\iota \odot Y^\iota) &= (1, X, \mathbf{0}) \odot (1, Y, \mathbf{0}) \\
&= (1, X \odot Y, \mathbf{0}) + (1, \mathbf{0} \odot \mathbf{0}, \mathbf{0}) + (1, \mathbf{0}, X \odot \mathbf{0}) + (1, \mathbf{0}, \mathbf{0} \odot Y) \\
&= (1, X \odot Y, \mathbf{0}) + (1, \mathbf{0}, \mathbf{0}) + (1, \mathbf{0}, \mathbf{0}) + (1, \mathbf{0}, \mathbf{0}) \\
&= (X \odot Y)^\iota + \mathbf{0}^\iota + \mathbf{0}^\iota + \mathbf{0}^\iota \neq (X \odot Y)^\iota. \qquad (5.16)
\end{aligned}$$

However, one can avoid this inconveniences by dealing with *weak equalities* on $\overline{\mathfrak{M}}_{m,n}$ as we are about to explain. On $\overline{\mathfrak{M}}_{m,n}$ the equality $\sigma = \tau$ means that σ equals componentwise to τ. Since our aim is to study expressions such as $\mathbb{E}[g(X)]$ for random matrices $X \in \mathfrak{M}_{m,n}$ and a certain class of functions g, we introduce the concept of weak equality on $\overline{\mathfrak{M}}_{m,n}$. Precisely, let $\mathcal{D}$ be a set of mappings of $\mathfrak{M}_{m,n}$ to $\mathbb{R}$. We say that random elements $\sigma, \tau \in \overline{\mathfrak{M}}_{m,n}$ are *weakly equal with respect to* $\mathcal{D}$, and we write $\sigma \equiv_{\mathcal{D}} \tau$, or $\sigma \equiv \tau$, when no confusion occurs, if

$$\forall g \in \mathcal{D} : \ \mathbb{E}[g^\iota(\sigma)] = \mathbb{E}[g^\iota(\tau)].$$

Obviously, if σ and τ are non-random then the above condition becomes

$$\forall g \in \mathcal{D} : \ g^\iota(\sigma) = g^\iota(\tau).$$

Remark 5.2. *By using weak equalities on the extended algebra $\mathcal{A}_*$ one can derive some interesting facts. In the following we enumerate a few of them.*

(i) The addition on $\overline{\mathfrak{M}}_{m,n}$ *is commutative, in a weak sense, with respect to any class of functions. Moreover,* $\mathbf{0}^{\iota} = (1, \mathbf{0}, \mathbf{0})$ *is a zero element. Note, however, that this is not unique. Indeed, any finite sum of elements of type* (c, X, X)*, with* $c > 0$ *and* $X \in \mathfrak{M}_{m,n}$ *acts as a neutral element for the addition.*

(ii) The extended $\odot$ *multiplication is a proper extension of the original* $\odot$ *multiplication of matrices, i.e.,*

$$\forall X \in \mathfrak{M}_{m,n}, Y \in \mathfrak{M}_{n,k} :\ (X \odot Y)^{\iota} \equiv_{\mathcal{D}} X^{\iota} \odot Y^{\iota},$$

where the above weak equality holds true with respect to any class of functions $\mathcal{D}$*. Indeed, this follows from (5.16) and (i) above.*

(iii) If $X_\theta \in \mathfrak{M}_{m,n}$ *is* $\mathcal{D}_p$*-differentiable, having derivative* $(c_\theta, X_\theta^+, X_\theta^-)$*, then it follows from (5.12) that*

$$\forall g \in \mathcal{D}_p(\mathfrak{M}_{m,n}) :\ \frac{d}{d\theta}\mathbb{E}\left[g(X_\theta)\right] = \mathbb{E}\left[g^{\iota}\left(c_\theta, X_\theta^+, X_\theta^-\right)\right].$$

Hence, by setting $X'_\theta := (c_\theta, X_\theta^+, X_\theta^-) \in \overline{\mathfrak{M}}_{m,n}$ *it follows that*

$$\forall g \in \mathcal{D}_p(\mathfrak{M}_{m,n}) :\ \frac{d}{d\theta}\mathbb{E}\left[g(X_\theta)\right] = \mathbb{E}\left[g^{\iota}\left(X'_\theta\right)\right]. \tag{5.17}$$

In particular, note that if $\tilde{X}'_\theta = (\tilde{c}_\theta, \tilde{X}_\theta^+, \tilde{X}_\theta^-)$ *is another representation of the derivative of* X_θ *then it holds that* $X'_\theta \equiv_{\mathcal{D}_p} \tilde{X}'_\theta$*. The same fact holds true for higher-order derivatives, which means that the definition of the derivatives of a random matrix, as given by (5.13), is correct.*

5.4.2 $\mathcal{D}_p$-Differential Calculus

In practice, checking $\mathcal{D}_p$-differentiability of a random matrix is not straightforward. In many applications, however, the distribution of the random matrix X_θ depends on θ through the distribution of some of its entries $[X_\theta]_{ij}$, for some pair of indices (i, j). It is natural that one would expect that $\mathcal{D}_p$-differentiability of X_θ is related to that of its entries. In the following we give a precise meaning to the above statement. To this end recall the notations in Section 4.2.1. Specifically, for $k, n \geq 1$ set

$$\mathcal{J}(k,n) := \{\jmath = (\jmath_1, \ldots, \jmath_k) :\ 0 \leq \jmath_l \leq n,\ \jmath_1 + \ldots + \jmath_k = n\}.$$

For $\jmath = (\jmath_1, \ldots \jmath_k) \in \mathcal{J}(k,n)$ we denote by $\nu(\jmath)$ the number of non-zero elements of the vector $\jmath$ and by $\mathcal{I}(\jmath)$ the set of vectors $\imath \in \{-1, 0, +1\}^k$ such that $\imath_l \neq 0$ if and only if $\jmath_l \neq 0$ and such that the product of all non-zero elements of $\imath$ equals one, i.e., there is an even number of -1. For $\imath \in \mathcal{I}(\jmath)$, we denote by $\bar{\imath}$ the vector obtained from $\imath$ by changing the sign of the non-zero element at the highest position.

Lemma 5.2. *Let* $\{U_{l,\theta} :\ 1 \leq l \leq k\} \subset \mathfrak{M}_1$ *be a collection of* n*-times* $\mathcal{D}_p$*-differentiable, independent random variables with* $\mathcal{D}_p$*-derivatives given by*

$$\forall 1 \leq l \leq k, 1 \leq m \leq n :\ \left(c_{l,\theta}^{(m)}, U_{l,\theta}^{(m,+)}, U_{l,\theta}^{(m,-)}\right).$$

If for each $\theta \in \Theta$ the entries of matrix X_θ satisfy

$$\forall (i,j):\ [X_\theta]_{ij} = X_{ij}(U_{1,\theta}, \ldots, U_{k,\theta}),$$

for some measurable mappings X_{ij}, then X_θ is n times $\mathcal{D}_p$-differentiable, provided that some positive constants $d_1, \ldots, d_k$ exist, such that

$$\forall u_1, \ldots, u_k:\ \|X(u_1, \ldots, u_k)\| \leq d_1\|u_1\| + \ldots + d_k\|u_k\|,$$

where $X(u_1, \ldots, u_k)$ denotes the matrix with entries $\{X_{ij}(u_1, \ldots, u_k) : (i,j)\}$. In addition, the n^{th}-order derivative $X_\theta^{(n)}$ can be represented in the extended algebra as follows:

$$\sum_{\jmath \in \mathcal{J}(k,n)} C_\theta(\jmath) \sum_{\imath \in \mathcal{I}(\jmath)} \left(1, X\left(U_{1,\theta}^{(\jmath_1,\imath_1)}, \ldots, U_{k,\theta}^{(\jmath_k,\imath_k)}\right), X\left(U_{1,\theta}^{(\jmath_1,\bar{\imath}_1)}, \ldots, U_{k,\theta}^{(\jmath_k,\bar{\imath}_k)}\right)\right),$$

where, for $\jmath = (\jmath_1, \ldots, \jmath_k) \in \mathcal{J}(k,n)$ we set

$$C_\theta(\jmath) := \binom{n}{\jmath_1, \ldots, \jmath_k} \prod_{l=1}^{k} c_{l,\theta}^{(\jmath_l)}$$

and, for convenience, we identify

$$\forall 1 \leq l \leq k:\ U_{l,\theta}^{(\jmath_l,+1)} = U_{l,\theta}^{(\jmath_l,+)},\ U_{l,\theta}^{(\jmath_l,-1)} = U_{l,\theta}^{(\jmath_l,-)},\ U_{l,\theta}^{(0,0)} = U_{l,\theta}.$$

Proof. Let us define $h : \mathfrak{M}_1 \times \ldots \times \mathfrak{M}_1 \to \mathfrak{M}_{m,n}$ as follows:

$$\forall u_1, \ldots, u_k:\ h(u_1, \ldots, u_k) := X(u_1, \ldots, u_k).$$

A successive application of Lemma 5.1 concludes the first part of the proof. The second part follows by applying Corollary 4.2 to the random elements $\{U_{l,\theta} : 1 \leq l \leq k\}$ and taking Remark 5.2 (*iii*) into account. □

The basis of our $\mathcal{D}_p$-differential calculus for random matrices is the following result which follows directly from Theorem 5.1 by re-writing (5.15) as an weak equality in the extended algebra $\mathcal{A}_*$.

Theorem 5.2. *Let $X_\theta \in \mathfrak{M}_{m,n}$, $Y_\theta \in \mathfrak{M}_{n,k}$ be stochastically independent, $\mathcal{D}_p$-differentiable random matrices with $\mathcal{D}_p$-derivatives X'_θ and Y'_θ, respectively. Then the generalized product $X_\theta \odot Y_\theta \in \mathfrak{M}_{m,k}$ is $\mathcal{D}_p$-differentiable and we have*

$$(X_\theta \odot Y_\theta)' \equiv_{\mathcal{D}_p} X'_\theta \odot Y^\iota_\theta + X^\iota_\theta \odot Y'_\theta.$$

Proof. From (5.12) in Theorem 5.1 we conclude that

$$\forall g \in \mathcal{D}_p:\ \frac{d}{d\theta}\mathbb{E}[g(X_\theta \odot Y_\theta)] = \mathbb{E}\left[g^\iota(X'_\theta \odot Y^\iota_\theta + X^\iota_\theta \odot Y'_\theta)\right]. \tag{5.18}$$

On the other hand, since $X_\theta \odot Y_\theta$ is $\mathcal{D}_p$-differentiable, it follows from (5.17) that

$$\forall g \in \mathcal{D}_p:\ \frac{d}{d\theta}\mathbb{E}[g(X_\theta \odot Y_\theta)] = \mathbb{E}\left[g^\iota\left((X_\theta \odot Y_\theta)'\right)\right],$$

which together with (5.18) imply that

$$\forall g \in \mathcal{D}_p:\ \mathbb{E}\left[g^\iota\left((X_\theta \odot Y_\theta)'\right)\right] = \mathbb{E}\left[g^\iota(X'_\theta \odot Y^\iota_\theta + X^\iota_\theta \odot Y'_\theta)\right].$$

This concludes the proof. □

The following result is the counterpart of the generalized Leibniz-Newton differentiation rule for random matrices.

Theorem 5.3. *Let $X_\theta(i)$, for $1 \leq i \leq k$, be a sequence of mutually independent, n-times $\mathcal{D}_p$-differentiable random matrices such that the generalized product*

$$X_\theta := X_\theta(k) \odot \ldots \odot X_\theta(1)$$

is well defined. Then X_θ is $\mathcal{D}_p$-differentiable and if we denote by $[X_\theta(i)]^{(m)}$ the m^{th}-order derivative of $X_\theta(i)$, for all $1 \leq i \leq k$, $1 \leq m \leq n$, then it holds that

$$X_\theta^{(n)} \equiv_{\mathcal{D}_p} \sum_{\jmath \in \mathcal{J}(k,n)} \binom{n}{\jmath_1, \ldots, \jmath_k} \cdot [X_\theta(k)]^{(\jmath_k)} \odot \ldots \odot [X_\theta(1)]^{(\jmath_1)},$$

where, for $1 \leq i \leq k$, we agree that $[X_\theta(i)]^{(0)} = [X_\theta(i)]^\iota$.

Proof. For a proof, note first that the function

$$h(x_k, \ldots, x_1) = x_k \odot \ldots \odot x_1$$

satisfies the conditions of Lemma 5.1 and then apply Corollary 4.2 to random variables $\{X_\theta(i) : 1 \leq i \leq k\}$. □

We conclude this section with discussing the concept of $\mathcal{D}_p$-analyticity of random matrices. We say that the random matrix X_θ is $\mathcal{D}_p$-analytic if its distribution is $\mathcal{D}_p$-analytic. Therefore, in accordance with Definition 4.1, it turns out that the random matrix X_θ is $\mathcal{D}_p$-analytic if the following two conditions are satisfied:

- all higher-order derivatives $X_\theta^{(n)}$, for $n \geq 1$, exist,
- there exist some neighborhood V of θ such that

$$\forall \xi;\ \theta + \xi \in V :\ X_{\theta+\xi} \equiv_{\mathcal{D}_p} \sum_{n=0}^{\infty} \frac{\xi^n}{n!} \cdot X_\theta^{(n)}.$$

Example 5.3. *Let us revisit Example 5.2. If X_θ is Bernoulli distributed with point masses A, B and parameter $\theta \in [0,1]$ it follows that all higher order derivatives of X_θ exist and one can easily check that for any $p \geq 1$ it holds that*

$$\forall \theta \in [0,1] :\ X_\theta \equiv_{\mathcal{D}_p} X_0 + \theta \cdot X_0' \equiv_{\mathcal{D}_p} \sum_{n=0}^{\infty} \frac{\theta^n}{n!} \cdot X_0^{(n)},$$

since, for $n \geq 2$, the derivatives $X_0^{(n)}$ are not significant. It follows that X_0 is weakly $\mathcal{D}_p$-analytic, for any $p \geq 1$; see Example 4.2.

Consequently, we extend concepts such as Taylor polynomials, radius and domain of convergence to analytic random matrices by means of their distribution.

Theorem 5.4. *Let* $\{U_{l,\theta} : 1 \leq l \leq k\} \subset \mathfrak{M}_1$ *be a collection of* $\mathcal{D}_p$*-analytic, independent random variables with corresponding domains of convergence* $D^p_\theta(U_l)$*, for* $1 \leq l \leq k$*, respectively. If for each* $\theta \in \Theta$ *the entries of matrix* X_θ *satisfy*

$$\forall (i,j) : [X_\theta]_{ij} = X_{ij}(U_{1,\theta}, \ldots, U_{k,\theta}),$$

for some measurable mappings X_{ij}*, then* X_θ *is* $\mathcal{D}_p$*-analytic, provided that some positive constants* $d_1, \ldots, d_k$ *exist, such that*

$$\forall u_1, \ldots, u_k : \|X(u_1, \ldots, u_k)\| \leq d_1\|u_1\| + \ldots + d_k\|u_k\|,$$

where $X(u_1, \ldots, u_k)$ *denotes the matrix with entries* $\{X_{ij}(u_1, \ldots, u_k) : (i,j)\}$*. More specifically, for each* ξ *such that* $\theta + \xi \in D^p_\theta(U_l)$*, for any* $1 \leq l \leq k$*, it holds that*

$$X_{\theta+\xi} \equiv_{\mathcal{D}_p} \sum_{n=0}^{\infty} \frac{\xi^n}{n!} X^{(n)}_\theta.$$

Proof. The existence of the derivatives $X^{(n)}_\theta$ follows from Lemma 5.2. Now apply Corollary 4.3 to the distributions $\mu_{l,\theta}$ of $U_{l,\theta}$ and use Lemma 5.1. □

Theorem 5.4 relates weak analyticity of a random matrix to that of its entries. In applications this will be an important technical tool, used to prove weak analyticity of a random matrix. Since in many models the state of a stochastic system is described by a finite product of random matrices, our next result will show that products of weakly analytical random matrices are again weakly analytical, in a $\mathcal{D}_p$-sense.

Theorem 5.5. *Let* $X_\theta(i)$*, for* $1 \leq i \leq k$*, be a sequence of stochastically independent,* $\mathcal{D}_p$*-analytic random matrices, having domains of convergence* $D^p_\theta(X(i))$*, respectively, such that the generalized product*

$$X_\theta := X_\theta(k) \odot \ldots \odot X_\theta(1)$$

is well defined. Then X_θ *is* $\mathcal{D}_p$*-analytic. Specifically, for any* ξ *such that* $\theta + \xi \in D^p_\theta(X(i))$*, for each* $1 \leq i \leq k$*, it holds that*

$$X_{\theta+\xi} \equiv_{\mathcal{D}_p} \sum_{n=0}^{\infty} \frac{\xi^n}{n!} \cdot X^{(n)}_\theta.$$

Proof. The existence of the derivatives $X^{(n)}_\theta$ follows from Theorem 5.3. To conclude the proof, apply Corollary 4.3 to the distributions $\mu_{i,\theta}$ of $X_\theta(i)$. □

We have constructed a weak $\mathcal{D}_p$-differential calculus for random matrices by "translating" the results in Chapter 4 in terms of random objects. Apart from being more handy working with random objects rather than working with probability distributions, this differential calculus has also the advantage that is based on a single class of cost-functions on each space, namely $\mathcal{D}_p$. Therefore, $\mathcal{D}_p$ can be seen as a "universal" class of cost-functions. The trade-off, however, is that we restrict our analysis to pseudo-normed algebras of matrices, i.e., we impose some restrictions on the upper-bounds under consideration.

5.5 Taylor Series Approximations for Stochastic Max-Plus Systems

In this section we illustrate our theory with two parameter-dependent max-plus dynamic systems. The first one, treated in Section 5.5.1 is inspired by F. Baccelli & D. Hong (see [6]) and describes a cyclic, multi-server station whereas in Section 5.5.2 we show how one can model a stochastic activity network as a max-plus dynamic system. In both situations we perform Taylor series approximations.

5.5.1 A Multi-Server Network with Delays/Breakdowns

Let us consider a cyclic network with two stations, where the first station has one server and the second one has two servers. The network has three customers two of which initially are beginning their service whereas the third one is in the buffer of the multi-sever station, just about to enter in the server. Assume that the service time at the single server station is σ time units, the service time at the multi-server station is τ time units and assume that each customer, after finishing its service at one station, instantly moves to the other station where he/she either waits in the buffer if the station is busy or enters the available server and begins its service. This system is called the *default system.* In the following we consider two variations of the default system: the *delayed system* and the *breakdown* system. The delayed system differs from the default in that the service time at the multi-server station is increased by an amount δ. In the breakdown system one server is removed from the multi-server station modeling a breakdown of the particular server. The three systems are illustrated in Figure 5.1.

The above three systems can be modeled as $(\max,+)$-linear systems. Indeed, if we choose as the state-variable a 4-dimensional vector $V(k)$ such that $V^1(k)$ denotes the k^{th} arrival epoch at the single-server station, $V^2(k)$ denotes the k^{th} departure epoch from the single-server station, $V^3(k)$ denotes the k^{th} arrival epoch at the multi-server station and $V^4(k)$ denotes the k^{th} departure epoch from the multi-server station, where $V^i(k)$, for $1 \le i \le 4$, denote the components of the vector $V(k)$, then the dynamics of each of the three systems is given by

$$\forall k \ge 0: \; V(k+1) = X \odot V(k),$$

where $\odot$ denotes the $(\max,+)$ matrix-vector multiplication and if we set

$$D := \begin{pmatrix} \sigma & \varepsilon & \tau & \varepsilon \\ \sigma & \varepsilon & \varepsilon & \varepsilon \\ \varepsilon & 0 & \varepsilon & 0 \\ \varepsilon & \varepsilon & \tau & \varepsilon \end{pmatrix}, \; P_d := \begin{pmatrix} \sigma & \varepsilon & \tau+\delta & \varepsilon \\ \sigma & \varepsilon & \varepsilon & \varepsilon \\ \varepsilon & 0 & \varepsilon & 0 \\ \varepsilon & \varepsilon & \tau+\delta & \varepsilon \end{pmatrix}, P_b := \begin{pmatrix} \sigma & \varepsilon & \tau & \varepsilon \\ \sigma & \varepsilon & \varepsilon & \varepsilon \\ \varepsilon & 0 & \tau & \varepsilon \\ \varepsilon & \varepsilon & \tau & \varepsilon \end{pmatrix},$$

then we have $X = D$ for the default system, $X = P_d$ for the delayed system and $X = P_b$ for the system with breakdowns; see [6] for a proof.

One can construct two hybrid stochastic systems out of these three, as follows: First, we consider a system with delays, i.e., each transition takes place according to the default matrix D with a certain probability $1-\theta$ and according to the delayed matrix P_d, with probability $\theta \in [0,1)$. The dynamic of such a system is thus given by

$$\forall k \ge 0: \; V_\theta(k+1) = X_\theta(k+1) \odot V_\theta(k),$$

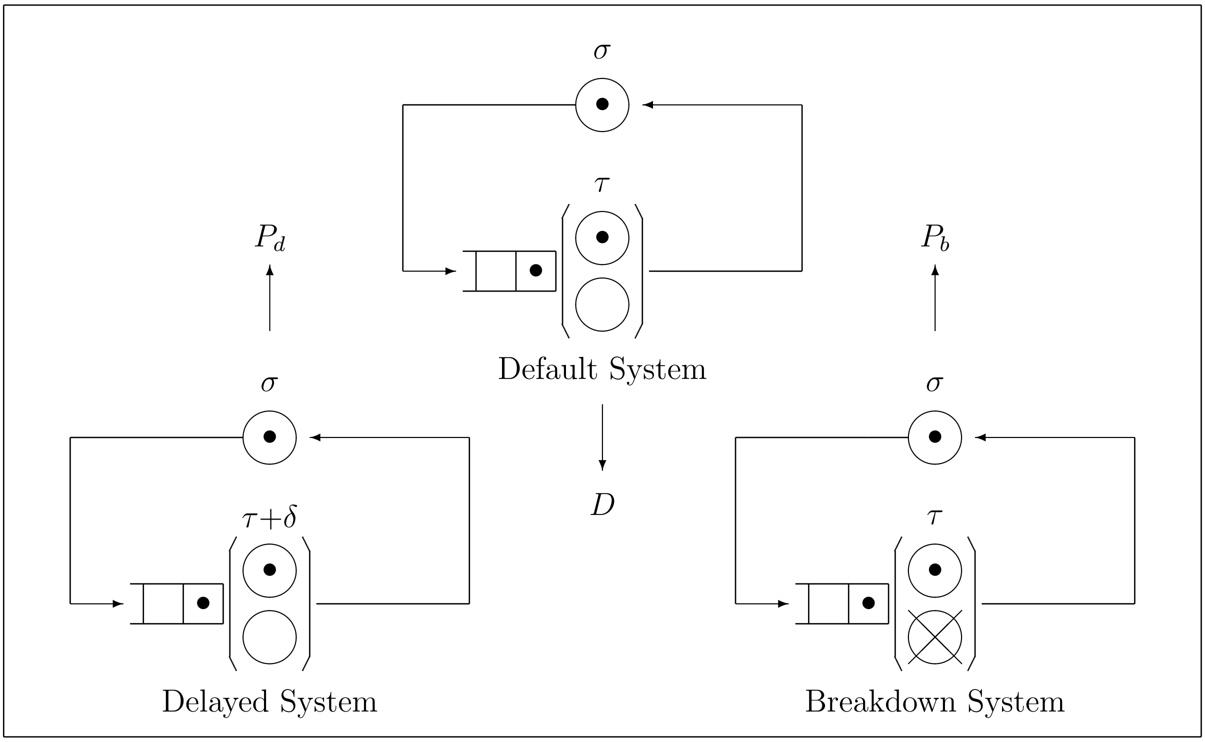

Fig. 5.1: A multi-server, cyclic network with perturbations (delay/breakdown).

where $V_\theta(0) = V(0) = \mathbf{0}$ and $\{A_\theta(k) : k \geq 1\}$ is a sequence of i.i.d. random matrices having their common distribution given by

$$\forall \theta \in [0,1] : \ \mu_\theta = (1-\theta) \cdot \delta_D + \theta \cdot \delta_{P_d}.$$

Secondly, we consider a system with random breakdowns which is actually defined in the same way as the first one, but one replaces the perturbation matrix P_d by P_b. This yields for the common distribution of $\{X(k) : k \geq 1\}$

$$\forall \theta \in [0,1] : \ \eta_\theta = (1-\theta) \cdot \delta_D + \theta \cdot \delta_{P_b}.$$

Therefore, in both situations the k^{th} state vector $V_\theta(k)$ is given by

$$\forall k \geq 1 : \ V_\theta(k) = X_\theta(k) \odot \ldots \odot X_\theta(1) \odot V(0). \tag{5.19}$$

Since $X_0(i)$ is $\mathcal{D}_p$-analytic, for any for any $1 \leq i \leq k$ and $p \geq 1$ (see Example 5.3), by Theorem 5.5 it follows that the product

$$X_0(k) \odot \ldots \odot X_0(1)$$

is $\mathcal{D}_p$-analytic, for any $p \geq 1$, and by Theorem 5.3 it holds that

$$X_\theta(k) \odot \ldots \odot X_\theta(1) \equiv_{\mathcal{D}_p} \sum_{n \geq 0} \sum_{\jmath \in \mathcal{J}(k,n)} \frac{\theta^n}{\jmath_1! \cdots \jmath_k!} \cdot [X_0(k)]^{(\jmath_k)} \odot \ldots \odot [X_0(1)]^{(\jmath_1)},$$

for any $\theta \in [0,1]$, where, for $1 \leq i \leq k$ and $j \geq 0$ we have

$$[X_\theta(i)]^{(j)} = \begin{cases} (1, X_\theta(i), \mathbf{0}), & j = 0, \\ (1, P_d, D), & j = 1, \\ (1, \mathbf{0}, \mathbf{0}) & j \geq 2; \end{cases}$$

see Example 5.2. It follows that for $n > k$ the n^{th}-order derivatives of the product $X_0(k) \odot \ldots \odot X_0(1)$ are not significant. Hence, the Taylor series is finite.

Fix now a finite horizon $k \geq 1$. In the following we illustrate how weak analyticity of the product $X_0(k) \odot \ldots \odot X_0(1)$ can be used to compute the expected values of the components $V_\theta^i(k)$, for $1 \leq i \leq 4$, of the vector $V_\theta(k)$. To this end, define, for $1 \leq i \leq 4$, the mappings $g_i : \mathfrak{M}_4 \to \mathbb{R}$ as follows:

$$\forall X \in \mathfrak{M}_4 : \; g_i(X) = (X \odot V(0))^i$$

and note that $g_i \in \mathcal{D}_p$ for each $p \geq 1$ and $1 \leq i \leq 4$ and from (5.19) we conclude that

$$\forall 1 \leq i \leq 4 : \; V_\theta^i(k) = g_i(X_\theta(k) \odot \ldots \odot X_\theta(1)).$$

Therefore, it follows that

$$\begin{aligned} \mathbb{E}\left[V_\theta^i(k)\right] &= \mathbb{E}\left[g_i(X_\theta(k) \odot \ldots \odot X_\theta(1))\right] \\ &= \sum_{n=0}^{k} \sum_{\jmath \in \mathcal{J}(k,n)} \frac{\theta^n}{\jmath_1! \cdots \jmath_k!} \mathbb{E}\left[g_i^\iota\left([X_0(k)]^{(\jmath_k)} \odot \ldots \odot [X_0(1)]^{(\jmath_1)}\right)\right] \\ &= \sum_{n=0}^{k} u_k(n)\theta^n, \end{aligned} \tag{5.20}$$

where, for $0 \leq n \leq k$, we set

$$\begin{aligned} u_k(n) &:= \sum_{\jmath \in \mathcal{J}(k,n)} \frac{1}{\jmath_1! \cdots \jmath_k!} \mathbb{E}\left[g_i^\iota\left([X_0(k)]^{(\jmath_k)} \odot \ldots \odot [X_0(1)]^{(\jmath_1)}\right)\right] \\ &= \sum_{\jmath \in \mathcal{J}(k,n)} \frac{1}{\jmath_1! \cdots \jmath_k!} g_i^\iota\left([X_0(k)]^{(\jmath_k)} \odot \ldots \odot [X_0(1)]^{(\jmath_1)}\right), \end{aligned} \tag{5.21}$$

since the product $[X_0(k)]^{(\jmath_k)} \odot \ldots \odot [X_0(1)]^{(\jmath_1)}$ is deterministic.

To evaluate the coefficients $\{u_k(n) : 0 \leq n \leq k\}$ let us introduce now the following notations: $[k] := \{1, 2, \ldots, k\}$ and for $I \subset [k]$ denote by $|I|$ its cardinal number and set

$$\Pi_I := B_k \odot \ldots \odot B_1,$$

where

$$B_i = \begin{cases} P_d, & i \in I, \\ D, & i \notin I. \end{cases}$$

For instance, we have $\Pi_\emptyset = D^k$,

$$\forall 1 \le i \le k : \; \Pi_{\{i\}} = D^{i-1} \odot P_d \odot D^{k-i}$$

$$1 \le i < j \le k : \; \Pi_{\{i,j\}} = D^{i-1} \odot P_d \odot D^{j-i-1} \odot P_d \odot D^{k-j}$$

and $\Pi_{[k]} = P_d^k$. Since $[X_0(i)]^{(\jmath_i)}$ is non significant for $\jmath_i \ge 2$, from (5.21) it follows that

$$\forall 0 \le n \le k : \; u_k(n) = \sum_{\jmath} g_i^{\iota} \left([X_0(k)]^{(\jmath_k)} \odot \ldots \odot [X_0(1)]^{(\jmath_1)} \right),$$

where the above sum is taken with respect to all $\jmath = (\jmath_1, \ldots, \jmath_k) \in \mathcal{J}(k,n)$ for which all components satisfy $\jmath_1, \ldots, \jmath_k \in \{0,1\}$. Moreover, if we set

$$I_\jmath := \{ i \in [k] : \jmath_i = 1 \}$$

it follows that

$$\forall 0 \le n \le k : \; u_k(n) = \sum_{|I_\jmath| = n} g_i^{\iota} \left([X_0(k)]^{(\jmath_k)} \odot \ldots \odot [X_0(1)]^{(\jmath_1)} \right).$$

Introducing now the sums

$$\forall 1 \le i \le 4, 0 \le n \le k : \; \sigma_i(n) = \sum_{|I|=n} g_i(\Pi_I),$$

we conclude from (5.21) that the coefficients of the Taylor series satisfy

$$u_k(n) = \sum_{l=0}^{n} (-1)^{n-l} \binom{k-l}{n-l} \sigma_i(l)$$

and from (5.20) we conclude that for any $1 \le i \le 4$ it holds that

$$\forall \theta \in [0,1] : \; \mathbb{E}[g(V_\theta^i(k))] = \sum_{n=0}^{k} \left[\sum_{l=0}^{n} (-1)^{n-l} \binom{k-l}{n-l} \sigma_i(l) \right] \theta^n. \tag{5.22}$$

Remark 5.3. *One could also arrive to (5.22) by using the equality*

$$\mathbb{E}[g(V_\theta^i(k))] = \sum_{I \subset [k]} g_i(\Pi_I) \theta^{|I|} (1-\theta)^{k-|I|} = \sum_{n=0}^{k} \sigma_i(n) \theta^n (1-\theta)^{k-n}.$$

and by calculating the coefficients of θ^m, for $1 \le m \le k$, in the right-hand side above.

Example 5.4. *For a numerical example set: $k = 10$, $\sigma = 14$, $\tau = 24$ and $\delta = 7$. A graphic representation of the Taylor polynomials of degree 1, 2 and 3, respectively, along with the true expected value of the first component of the state-vector $V_\theta(10)$ for the system with delays can be seen in Figure 5.2. For the system with breakdowns one can use a similar reasoning by replacing P_d by P_b. The corresponding Taylor polynomials of degree 1, 2 and 3, along with the true expected value of the first component of the state-vector $V_\theta(10)$ are represented in Figure 5.3. In both pictures the thick line represents the true value.*

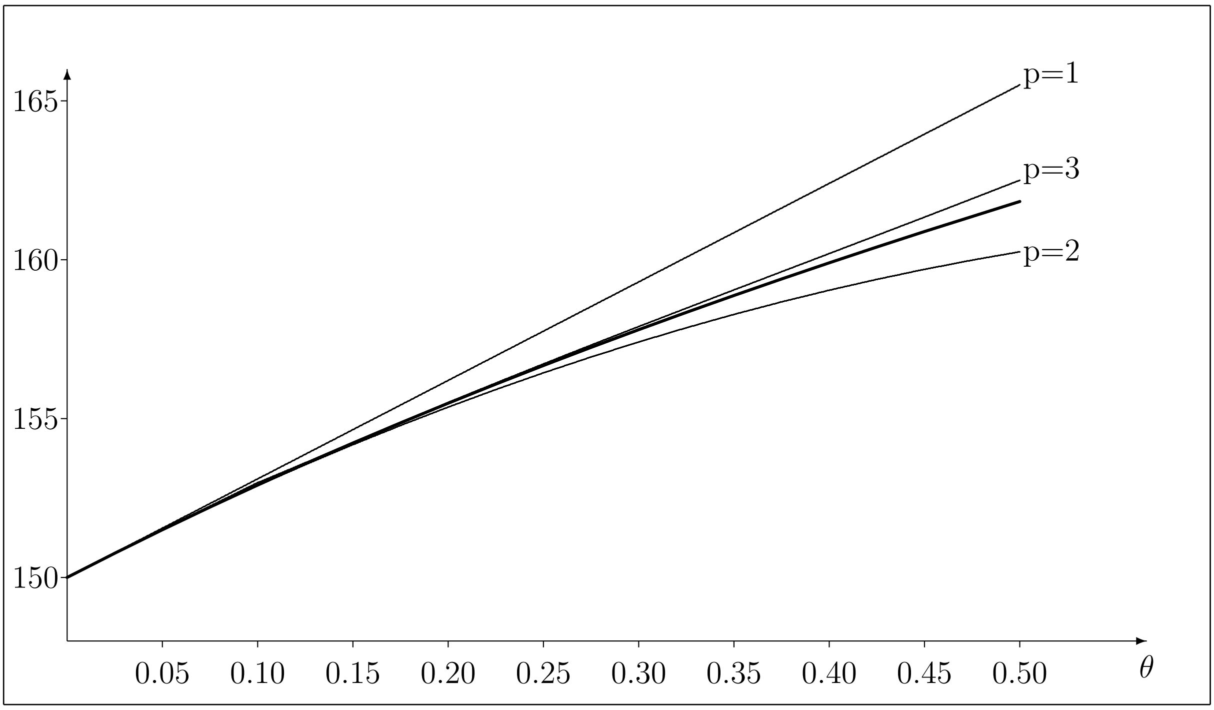

Fig. 5.2: Taylor approximations of orders 1, 2 and 3 along with the true value of $\mathbb{E}\left[V_\theta^1(10)\right]$ (thick line), for the system with delays.

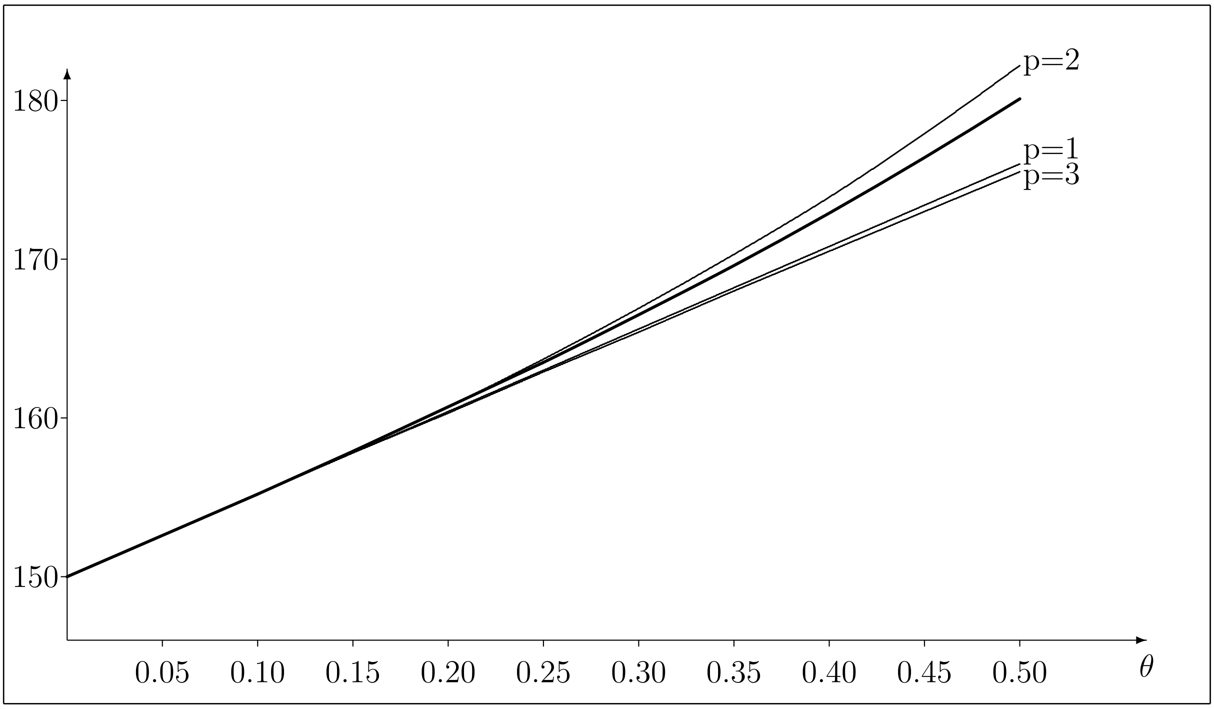

Fig. 5.3: Taylor approximations of orders 1, 2 and 3 along with the true value of $\mathbb{E}\left[V_\theta^1(10)\right]$ (thick line), for the system with breakdowns.

5.5.2 SAN Modeled as Max-Plus-Linear Systems

Let us re-visit the case of stochastic-activity-network models described in Section 4.3. Recall that an SAN can be described as a directed, acyclic, graph $(\mathcal{V}, \mathcal{E} \subset \mathcal{V} \times \mathcal{V})$, with one source and one sink node and additive weight mapping $\tau : \mathcal{E} \to \mathbb{R}$. For convenience $\mathcal{V} = \{1, 2, \ldots, n\}$ and, since we deal with a directed, acyclic graph the nodes can be ordered such that whenever i is connected to j, i.e., $(i, j) \in \mathcal{E}$, it holds that $i < j$.

Recall that the completion time of a SAN is given by the weight of a "critical" path (where critical can be maximal or minimal, according to the situation). For $i \in \mathcal{V}$ let us denote by t_i the completion time of the i^{th} node, i.e., the completion time of the SAN obtained by removing the nodes $\{k : k \geq i+1\}$ and the adjacent edges form the original graph. In general, computing the completion times of large SAN might be very demanding. Classical exhaustive "walk-through graph" methods are hard to implement. That is because the set of paths from source to sink node may become very large (it may have up to 2^n elements).

Alternatively, one can model a SAN as a dynamic max-plus system and compute the vector of completion times $\mathbf{t} := (t_1, \ldots, t_n)^t$, where for a matrix A we denote by A^t its transposition, using the following scheme.

Algorithm 1. *The following algorithm yields the vector of completion times in a SAN:*

1. *Construct the incidence $n \times n$ matrix A of the given graph.*
2. *For $i = 1$ up to n, consider the matrix $A(i)$, obtained from the identity matrix $\mathbf{I}$ by replacing the i^{th} row with the i^{th} row of A.*
3. *Denote by $\boldsymbol{e}^1$ the first unit vector $(0, \varepsilon, \ldots, \varepsilon)^t$ and set*
$$\boldsymbol{t} := A(n) \odot A(n-1) \odot \cdots \odot A(1) \odot \boldsymbol{e}^1.$$
4. *Recover the completion time of the i^{th} node of the SAN from the i^{th} component of the vector $\boldsymbol{t}$.*

Remark 5.4. *The incidence matrix A is sub-diagonal and it follows that $A(i)$ differs from the identity matrix by at most $(i-1)$ entries. Moreover, since $A(1) = I$, the identity matrix, i.e., $\mathbf{t}_1 = \mathbf{e}$, it can be omitted. Finally, provided that the weights $\{\tau(e) : e \in \mathcal{E}\}$ are mutually independent, the matrices $A(i)$, for $1 \leq i \leq n$ are stochastically independent.*

For instance, let us consider the SAN example studied in Section 4.3 where X_i, for $1 \leq i \leq 7$ denote the weights (durations) of the subsequent activities (see Figure 4.1) and recall that $\varepsilon := -\infty$. Then the vector of completion times for this SAN can be obtained by considering the following matrices:

$$A(2) = \begin{pmatrix} 0 & \varepsilon & \varepsilon & \varepsilon & \varepsilon \\ X_1 & 0 & \varepsilon & \varepsilon & \varepsilon \\ \varepsilon & \varepsilon & 0 & \varepsilon & \varepsilon \\ \varepsilon & \varepsilon & \varepsilon & 0 & \varepsilon \\ \varepsilon & \varepsilon & \varepsilon & \varepsilon & 0 \end{pmatrix}, \quad A(3) = \begin{pmatrix} 0 & \varepsilon & \varepsilon & \varepsilon & \varepsilon \\ \varepsilon & 0 & \varepsilon & \varepsilon & \varepsilon \\ X_2 & X_3 & 0 & \varepsilon & \varepsilon \\ \varepsilon & \varepsilon & \varepsilon & 0 & \varepsilon \\ \varepsilon & \varepsilon & \varepsilon & \varepsilon & 0 \end{pmatrix},$$

$$A(4) = \begin{pmatrix} 0 & \varepsilon & \varepsilon & \varepsilon & \varepsilon \\ \varepsilon & 0 & \varepsilon & \varepsilon & \varepsilon \\ \varepsilon & \varepsilon & 0 & \varepsilon & \varepsilon \\ \varepsilon & X_4 & X_6 & 0 & \varepsilon \\ \varepsilon & \varepsilon & \varepsilon & \varepsilon & 0 \end{pmatrix}, \quad A(5) = \begin{pmatrix} 0 & \varepsilon & \varepsilon & \varepsilon & \varepsilon \\ \varepsilon & 0 & \varepsilon & \varepsilon & \varepsilon \\ \varepsilon & \varepsilon & 0 & \varepsilon & \varepsilon \\ \varepsilon & \varepsilon & \varepsilon & 0 & \varepsilon \\ \varepsilon & X_5 & \varepsilon & X_7 & 0 \end{pmatrix}.$$

It is easy to check that the following matrix-vector product in max-plus algebra

$$\mathbf{t} := A(5) \odot A(4) \odot A(3) \odot A(2) \odot \mathbf{e}$$

yields the vector of completion times for the SAN under consideration. More specifically, $\mathbf{t} = (t_1, t_2, t_3, t_4, t_5) \in [0, \infty)^5$ has the property that t_i equals to the completion time at the i^{th} node. In particular, the completion time T of the full network is given by t_5, i.e., $T = t_5$. It follows that the expected completion time T of the SAN can be written as

$$\mathbb{E}[T] = \mathbb{E}\left[(A(5) \odot A(4) \odot A(3) \odot A(2) \odot \mathbf{e})_5\right]. \tag{5.23}$$

Recall that in Section 4.3 the weights X_i, for $1 \leq i \leq 7$, were independent exponentially distributed random variables with rates λ_i, respectively. We have assumed further that $\lambda_1 = \lambda_3 = \theta$ while the other rates are independent of θ. In the following we formalize the reasoning put forward in Section 4.3, i.e., performing Taylor series approximations of the expected completion time T_θ of the SAN with respect to parameter θ, in terms of $\mathcal{D}_p$-differential calculus for random matrices.

To start with, note that for each $1 \leq i \leq 5$ the mapping

$$A \in \mathfrak{M}_5 : \ A \mapsto [A \odot \mathbf{e}]_i$$

belongs to any $\mathcal{D}_p$-space, for $p \geq 0$. Therefore, by Theorem 5.5 it follows that analyticity of $\mathbb{E}[T_\theta]$ follows from $\mathcal{D}_p$-analyticity of the product

$$A_\theta := A_\theta(5) \odot A_\theta(4) \odot A_\theta(3) \odot A_\theta(2), \tag{5.24}$$

where we use the notation $A_\theta(i)$ instead of $A(i)$ in order to illustrate the dependence of their distributions on θ. We agree that $A_\theta(i)$ is constant if its distribution does not depend on θ. Note that, in this case, only $A_\theta(2)$ and $A_\theta(3)$ depend on θ.

Since the exponential distribution is weakly $[\mathcal{D}]_v$-analytic, for any polynomial v (see Example 4.3), it follows by Theorem 5.4 that the matrices $A_\theta(i)$, for $2 \leq i \leq 5$, are $\mathcal{D}_p$-analytic and by applying Theorem 5.5 we conclude that the product A_θ in (5.24) is $\mathcal{D}_p$-analytic. In addition, for each ξ such that $|\xi| < \theta$ it holds that

$$A_{\theta+\xi} \equiv_{\mathcal{D}_p} \sum_{n=0}^{\infty} \frac{\xi^n}{n!} \cdot A_\theta^{(n)}. \tag{5.25}$$

To compute the derivatives $A_\theta^{(n)}$ one can use Lemma 5.2. To this end, let us consider two sequences $\{X_{1,l} : l \geq 1\}$ and $\{X_{3,l} : l \geq 1\}$ of i.i.d. random variables having exponential distribution with rate θ and let $T_{j,k}$ denote the completion time of the modified SAN, i.e., if one replaces X_1 by $\sum_{l=1}^{j} X_{1,l}$ and X_3 by $\sum_{l=1}^{k} X_{3,l}$; see Section 4.3. For instance, if we set

$$\forall n \geq 0: \; S_3^n := \sum_{l=1}^{n} X_{3,l},$$

the derivatives $A_\theta^{(n)}(3)$ of $A_\theta(3)$ can be expressed as $(c_\theta^{(n)}(3), A_\theta^{(n+)}(3), A_\theta^{(n-)}(3))$ where, for each $n \geq 1$, $c_\theta^n(2) = \frac{n!}{\theta^n}$ and

$$A_\theta^{(n,+)}(3) = \begin{pmatrix} 0 & \varepsilon & \varepsilon & \varepsilon & \varepsilon \\ \varepsilon & 0 & \varepsilon & \varepsilon & \varepsilon \\ X_2 & S_3^n & 0 & \varepsilon & \varepsilon \\ \varepsilon & \varepsilon & \varepsilon & 0 & \varepsilon \\ \varepsilon & \varepsilon & \varepsilon & \varepsilon & 0 \end{pmatrix}, \; A_\theta^{(n,-)}(3) = \begin{pmatrix} 0 & \varepsilon & \varepsilon & \varepsilon & \varepsilon \\ \varepsilon & 0 & \varepsilon & \varepsilon & \varepsilon \\ X_2 & S_3^{n+1} & 0 & \varepsilon & \varepsilon \\ \varepsilon & \varepsilon & \varepsilon & 0 & \varepsilon \\ \varepsilon & \varepsilon & \varepsilon & \varepsilon & 0 \end{pmatrix},$$

for n odd and

$$A_\theta^{(n,+)}(3) = \begin{pmatrix} 0 & \varepsilon & \varepsilon & \varepsilon & \varepsilon \\ \varepsilon & 0 & \varepsilon & \varepsilon & \varepsilon \\ X_2 & S_3^{n+1} & 0 & \varepsilon & \varepsilon \\ \varepsilon & \varepsilon & \varepsilon & 0 & \varepsilon \\ \varepsilon & \varepsilon & \varepsilon & \varepsilon & 0 \end{pmatrix}, \; A_\theta^{(n,-)}(3) = \begin{pmatrix} 0 & \varepsilon & \varepsilon & \varepsilon & \varepsilon \\ \varepsilon & 0 & \varepsilon & \varepsilon & \varepsilon \\ X_2 & S_3^{n} & 0 & \varepsilon & \varepsilon \\ \varepsilon & \varepsilon & \varepsilon & 0 & \varepsilon \\ \varepsilon & \varepsilon & \varepsilon & \varepsilon & 0 \end{pmatrix},$$

for n even. One can proceed similarly for calculating the derivatives $A_\theta^{(n)}(2)$ of $A_\theta(2)$, for $n \geq 1$.

Now Theorem 5.3 allows us to compute the higher-order derivatives of the product A_θ in (5.24). Since only $A(2)$ and $A(3)$ depend on θ, it follows that the higher-order derivatives of $\mathbb{E}[T_\theta]$ can be written as follows:

$$\begin{aligned} \forall n \geq 1: \; \frac{d^n}{d\theta^n}\mathbb{E}[T_\theta] &= \frac{d^n}{d\theta^n}\mathbb{E}\left[g_5\left(A_\theta^{(n)}\right)\right] \\ &= \frac{d^n}{d\theta^n}\mathbb{E}\left[g_5\left(A_\theta(5) \odot A_\theta(4) \odot A_\theta(3) \odot A_\theta(2)\right)\right] \\ &= \sum_{j+k=n} \mathbb{E}\left[g_5^\iota\left(A_\theta(5) \odot A_\theta(4) \odot A_\theta^{(j)}(3) \odot A_\theta^{(k)}(2)\right)\right] \\ &= (-1)^n \frac{n!}{\theta^n} \sum_{j+k=n} \mathbb{E}[T_{j+1,k+1} - T_{j+1,k} - T_{j,k+1} + T_{j,k}], \end{aligned}$$

where $g_5 : \mathfrak{M}_5 \to \mathbb{R}$, $g_5(X) = [X \odot \mathbf{e}]_5$, i.e., we obtain the following sequence of Taylor polynomials

$$\mathbf{T}_n(\theta, \xi) := \sum_{m=0}^{n} (-1)^m \left(\frac{\xi}{\theta}\right)^m \sum_{j+k=m} \mathbb{E}_\theta[T_{j+1,k+1} - T_{j+1,k} - T_{j,k+1} + T_{j,k}],$$

for $n \geq 0$ and $|\xi| < \theta$. The above Taylor series is identical to the one in (4.17). Hence, we can proceed just like in Section 4.3.

5.6 Concluding Remarks

In this chapter we have considered parameter-dependent stochastic systems whose physical evolution is modeled by a matrix-vector multiplication in some general algebra. To analyze these systems we have adapted the measure-valued differentiation theory to random matrices and it turned out that, in the $\mathcal{D}_p$-space setting, a weak differential calculus, similar to the classical one, holds true for random matrices. The key result of this chapter states that $\mathcal{D}_p$-differentiability (resp. analyticity) of random matrices X_θ and Y_θ is inherited by their generalized product $X_\theta \odot Y_\theta$ for a certain class of matrix multiplication operators $\odot$. Based on this differential calculus we have derived Taylor series approximations for DES.

As illustrated in this chapter, Taylor series approximations provide rather accurate estimations. A similar problem, for (max-plus)-linear stochastic systems with parameter-dependent Poisson input has been addressed in [7] where the coefficients of the Taylor series appear as the expectations of polynomials of some input variables of the system. In addition, the method was successfully applied to derive Taylor series expansions for Lyapunov exponents of ergodic (max-plus)-linear systems. In [6] Taylor series expansions are obtained for the max-plus Lyapunov exponent of an i.i.d. sequence of Bernoulli distributed random matrices (in particular for the network with breakdowns presented in Section 5.5.1), where the derivatives are evaluated using specific max-plus techniques such as backwards coupling. A theory of Taylor series expansion of products in the (max-plus) algebra is provided in [21].

The analysis put forward in this chapter is meant to be a first step into developing a general theory to comprise a wider range of applications. In this sense, challenging topics for future research are, for instance, to adapt the theory of weak differentiation to the random horizon setting (in order to construct Taylor series approximations for Lyapunov exponents, whose existence in the case of generalized linear stochastic systems can be shown by using sub-additive ergodic theory; see, e.g., [37, 21]), to develop efficient algorithms for evaluating the derivatives based on the particularities of the model and to obtain accurate estimates for the error of the Taylor polynomials.

APPENDIX

A. Convergence of Infinite Series of Real Numbers

Let us consider a sequence $\{a_n : n \geq 0\}$ of real numbers and consider the infinite series $\sum_{n=0}^{\infty} a_n$. The series is said to be convergent if the sequence S_n defined as

$$\forall n \geq 0: \; S_n := \sum_{k=0}^{n} a_k$$

is convergent in $\mathbb{R}$ and is said to be divergent otherwise. The limit of the sequence $\{S_n\}_{n\geq 0}$ (provided that it exists) is called the sum of the series. In addition, the convergent series $\sum_{n=0}^{\infty} a_n$ is said to be absolutely convergent if the series $\sum_{n=0}^{\infty} |a_n|$ is convergent and we call it conditionally convergent if $\sum_{n=0}^{\infty} |a_n|$ is divergent.

Theorem A.1. ***The Rearrangements Theorem***: *Let $\sum_{n=0}^{\infty} a_n$ be a convergent series. Then,*

(i) If the series is absolutely convergent then for any permutation σ of the set of non-negative integers the series $\sum_{n=0}^{\infty} a_{\sigma(n)}$ is convergent and has the same sum as the original series.

(ii) If the series is conditionally convergent then for any $S \in \mathbb{R}$ there exists a permutation σ such that the series $\sum_{n=0}^{\infty} a_{\sigma(n)}$ converges to S.

Theorem A.2. ***Cauchy-Hadamard Theorem***: *Let $\{a_n : n \geq 0\}$ be a sequence of real numbers and let*

$$R := \frac{1}{\limsup_{n\to\infty} \sqrt[n]{|a_n|}},$$

where we agree that $1/\infty = 0$ and $1/0 = \infty$. Then, the power series

$$\sum_{n=0}^{\infty} a_n \xi^n$$

is absolutely convergent, uniformly with respect to $|\xi| < R$.

For a proof of these results see, e.g., [53].

B. Interchanging Limits

In this section we state two standard results from classical analysis which establish sufficient conditions for interchanging limits with continuity and differentiability.

Theorem B.1. ***Interchanging Limits**: Let $\{a_{m,n} : m, n \in \mathbb{N}\}$ be a double-indexed sequence of real numbers such that $a_{m,n}$ converges to some limit c_n, for $m \to \infty$, uniformly with respect to $n \in \mathbb{N}$. If, in addition, the limit*

$$b_m := \lim_{n\to\infty} a_{m,n}$$

exists for each $m \in \mathbb{N}$, then the sequence $\{b_m\}_m$ converges and it holds that

$$\lim_{n\to\infty} c_n = \lim_{m\to\infty} b_m.$$

That is, interchanging of limits is justified and we have

$$\lim_{n\to\infty}\lim_{m\to\infty} a_{m,n} = \lim_{m\to\infty}\lim_{n\to\infty} a_{m,n}.$$

Theorem B.2. ***Interchanging Limit and Continuity**: Assume that $f_n : A \subset \mathbb{S} \to \mathbb{R}$, for $n \in \mathbb{N}$, defines a sequence of functions which converge uniformly on A to some function f and let x be an accumulation point of A. If, in addition, the limit*

$$L_n := \lim_{s\to x} f_n(s)$$

exists for each $n \in \mathbb{N}$, then the sequence $\{L_n\}_n$ converges and it holds that

$$\lim_{s\to x} f(s) = \lim_{n\to\infty} L_n.$$

That is, interchanging the limit with continuity is justified and we have

$$\lim_{s\to x}\lim_{n\to\infty} f_n(s) = \lim_{n\to\infty}\lim_{s\to x} f_n(s).$$

Theorem B.3. ***Interchanging Limit and Differentiation**: Let $A \subset \mathbb{R}$ be a compact interval and consider a sequence of functions $f_n : A \to \mathbb{R}$, for $n \geq 1$, satisfying:*

(i) f_n is differentiable on A for each $n \geq 1$,

(ii) there exists some $x_0 \in A$ such that the sequence $\{f_n(x_0)\}_{n\geq 1}$ converges.

If the sequence $\{f'_n\}_{n\geq 1}$ converges uniformly on A, then the sequence $\{f_n\}_{n\geq 1}$ converges uniformly on A, to some function f, and we have

$$\forall x \in A : \; f'(x) = \lim_{n\to\infty} f'_n(x).$$

Theorem B.4. ***Interchanging Limit and Integration**: If $f_n : [a, b] \to \mathbb{R}$, for $n \geq 1$, is a sequence of Riemann integrable functions that converges uniformly on $[a, b]$ then the limit $f : [a, b] \to \mathbb{R}$ is Riemann integrable and it holds that*

$$\int_a^b f(x)dx = \lim_{n\to\infty} \int_a^b f(x)dx.$$

For a proof of these results we refer to [53].

C. Measure Theory

In this section we list a few standard results on measure theory. In what follows we assume that $\mathbb{S}$ is a metric space endowed with the Borel σ-field $\mathcal{S}$. If $\mu \in \mathcal{M}$ we say that a property holds true almost everywhere with respect to $|\mu|$, and we write $|\mu|$-a.e., if the property holds true for each $s \in \mathbb{S}$ except on a set $A \in \mathcal{S}$ such that $|\mu|(A) = 0$.

Theorem C.1. ***Dominated Convergence Theorem****: Let $\mu \in \mathcal{M}$ and assume that $\{f_n : n \in \mathbb{N}\} \subset \mathcal{L}^1(|\mu|)$ is a sequence of functions such that $f_n \to f$, $|\mu|$-a.e. and there exist $g \in \mathcal{L}^1(|\mu|)$ such that $|f_n| \leq g$, $|\mu|$-a.e., for all $n \in \mathbb{N}$. Then,*

$$\lim_{n\to\infty} \int f_n(x)\mu(dx) = \int f(x)\mu(dx).$$

Theorem C.2. ***Monotone Convergence Theorem****: Let $\mu \in \mathcal{M}$ and assume that $\{f_n : n \in \mathbb{N}\} \subset \mathcal{L}^1(|\mu|)$ is a sequence of non-negative functions such that $f_n \to f$, $|\mu|$-a.e. and satisfying $f_n \leq f_{n+1}$, $|\mu|$-a.e., for all $n \in \mathbb{N}$. Then,*

$$\lim_{n\to\infty} \int f_n(x)\mu(dx) = \int f(x)\mu(dx).$$

Theorem C.3. ***Radon-Nikodym Theorem****: Let μ be a positive measure on $\mathcal{S}$ and ν be a finite signed measure on $\mathcal{S}$. If $|\nu|$ is absolutely continuous with respect to μ then there exists $f \in \mathcal{L}^1(\mu)$ such that*

$$\forall A \in \mathcal{S} : \ \nu(A) = \int f(x)\mathbb{I}_A(x)\mu(dx).$$

f is called the Radon-Nikodym derivative and is unique μ-a.e.

Theorem C.4. ***Lebesgue Decomposition Theorem****: Let λ be a positive measure on $\mathcal{S}$ and ν be a finite signed measure on $\mathcal{S}$. Then there exist some uniquely determined finite signed measures μ, κ such that*

- *$|\mu|$ is absolutely continuous with respect to λ,*
- *$|\kappa|$ and λ are orthogonal,*
- *$\nu = \mu + \kappa$.*

For the proof of the above results we refer to [14].

D. Conditional Expectations

This section is intended to list the definition and the basic properties of the conditional expectation as stated in any standard textbook on probability theory.

Theorem D.1. ***Existence of the Conditional Expectation***: *Let $(\Omega, \mathcal{K}, \mathbb{P})$ be a probability field and $\mathcal{B} \subset \mathcal{K}$ be a σ-field. For any random variable $X \in \mathcal{L}^1(\mathcal{K}, \mathbb{P})$ there exists a $\mathbb{P}$-a.s. unique random variable $Z \in \mathcal{L}^1(\mathcal{B}, \mathbb{P})$, denoted by $\mathbb{E}[X|\mathcal{B}]$, such that*

$$B \in \mathcal{B}: \ \mathbb{E}[Z\mathbb{I}_B] = \mathbb{E}[X\mathbb{I}_B].$$

The random variable Z is called the conditional expectation of X with respect to $\mathcal{B}$. If $Y \in \mathcal{L}^1(\mathcal{K}, \mathbb{P})$ then we define the conditional expectation of X with respect to Y as $\mathbb{E}[X|Y] := \mathbb{E}[X|\sigma(Y)]$, where $\sigma(Y)$ denotes the σ-field generated by Y.

The conditional expectation acts as a projection operator from $\mathcal{L}^1(\mathcal{K}, \mathbb{P})$ onto $\mathcal{L}^1(\mathcal{B}, \mathbb{P})$. As an operator, the conditional expectation enjoys the following basic properties[3]:

(a) Is identic when restricted to $\mathcal{L}^1(\mathcal{B}, \mathbb{P})$, i.e., if $X \in \mathcal{L}^1(\mathcal{B}, \mathbb{P})$ then $\mathbb{E}[X|\mathcal{B}] = X$.

(b) Is idempotent, i.e., $\mathbb{E}\left[\mathbb{E}[X|\mathcal{B}]\,|\mathcal{B}\right] = \mathbb{E}[X|\mathcal{B}]$

(c) Preserves the total expectation, i.e., $\mathbb{E}\left[\mathbb{E}[X|\mathcal{B}]\right] = \mathbb{E}[X]$.

(d) Is linear, i.e., $\forall u, v \in \mathbb{R}: \ \mathbb{E}[uX + vY|\mathcal{B}] = u\mathbb{E}[X|\mathcal{B}] + v\mathbb{E}[Y|\mathcal{B}]$.

(e) It is positive (monotone), i.e., if $X \geq 0$ then $\mathbb{E}[X|\mathcal{B}] \geq 0$ and, more generally

$$X \leq Y \implies \mathbb{E}[X|\mathcal{B}] \leq \mathbb{E}[Y|\mathcal{B}].$$

(f) Is contractive (in particular, continuous), i.e., $|\mathbb{E}[X|\mathcal{B}]| \leq \mathbb{E}[|X||\mathcal{B}]$, which implies

$$\|\mathbb{E}[X|\mathcal{B}]\|_{\mathcal{L}^1} \leq \|X\|_{\mathcal{L}^1}.$$

(g) Is consistent with σ-fields embedding, i.e., if $\mathcal{A} \subset \mathcal{B}$ is a subfield then

$$\mathbb{E}\left[\mathbb{E}[X|\mathcal{B}]\,|\mathcal{A}\right] = \mathbb{E}[X|\mathcal{A}].$$

(h) If $X \in \mathcal{L}^1(\mathcal{B}, \mathbb{P})$ is bounded and $Y \in \mathcal{L}^1(\mathcal{K}, \mathbb{P})$ then $\mathbb{E}[XY|\mathcal{B}] = X\mathbb{E}[Y|\mathcal{B}]$ and it follows that $\mathbb{E}\left[X\mathbb{E}[Y|\mathcal{B}]\right] = \mathbb{E}[XY]$.

(i) If X is independent of $\mathcal{B}$ then $\mathbb{E}[X|\mathcal{B}] = \mathbb{E}[X]$.

(j) $Z = \mathbb{E}[X|\mathcal{B}]$ if and only if $Z \in \mathcal{L}^1(\mathcal{B}, \mathbb{P})$ and for each $\mathcal{B}$-measurable random variable Y it holds that $\mathbb{E}[ZY] = \mathbb{E}[XY]$. In particular, $Z = \mathbb{E}[X|Y]$, for some $Y \in \mathcal{L}^1(\mathcal{K}, \mathbb{P})$, if and only if $Z \in \mathcal{L}^1(\sigma(Y), \mathbb{P})$ and for each Borel measurable function f

$$\mathbb{E}[Zf(Y)] = \mathbb{E}[Xf(Y)].$$

[3] Note that the above equalities hold $\mathbb{P}$-a.s.

E. Fubini Theorem and Applications

The following theorem is the basis for calculating multiple integrals, i.e., integrals with respect to finite products of measures, in measure and probability theory (for a proof see, e.g., [14]).

Theorem E.1. ***Fubini Theorem***: *Let* $(\mathbb{S}, \mathcal{S}, \mu)$ *and* $(\mathbb{T}, \mathcal{T}, \eta)$ *be* σ*-finite measure spaces. If* $g \in \mathcal{L}^1(\mu \times \eta)$ *then* $g(s, \cdot) \in \mathcal{L}^1(\eta)$ μ*-a.e. and* $g(\cdot, t) \in \mathcal{L}^1(\mu)$ η*-a.e. Furthermore, the mappings* $s \mapsto \int g(s,t)\eta(dt)$ *and* $t \mapsto \int g(s,t)\mu(ds)$ *belong to* $\mathcal{L}^1(\mu)$ *and* $\mathcal{L}^1(\eta)$*, respectively and it holds that*

$$\int g(s,t)(\mu \times \eta)(ds, dt) = \int \left[\int g(s,t)\eta(dt) \right] \mu(ds) = \int \left[\int g(s,t)\mu(ds) \right] \eta(dt).$$

The following result, which is useful for calculating conditional expectations, follows from Fubini Theorem.

Lemma E.2. *Let* X *and* Z *be independent random variables, defined on a common probability space, taking values in measurable spaces* $(\mathbb{S}, \mathcal{S})$ *and* $(\mathbb{T}, \mathcal{T})$*, respectively. For any bounded Borel measurable mapping* Φ *defined on* $\mathbb{S} \times \mathbb{T}$ *the function*

$$\forall x \in \mathbb{S}: \ \phi(x) := \mathbb{E}\left[\Phi(x, Z)\right]$$

is measurable on $\mathbb{S}$ *and it holds that*

$$\mathbb{E}\left[\Phi(X, Z) \middle| X\right] = \phi(X), \ a.s.$$

Proof. Let us denote by μ and η the distributions of X and Z, respectively. It follows that ϕ satisfies

$$\forall x \in \mathbb{S}: \ \phi(x) = \int \Phi(x, z)\eta(dz)$$

and measurability of ϕ follows from Fubini Theorem. Therefore, $\phi(X)$ is $\sigma(X)$-measurable.

Let us consider an arbitrary $\sigma(X)$-measurable random variable Y, i.e., $Y = f(X)$, for some Borel function f. Then, using again Fubini Theorem, one can show that

$$\begin{aligned} \mathbb{E}\left[\Phi(X, Z)Y\right] = \mathbb{E}\left[\Phi(X, Z)f(X)\right] &= \int \left[\int \Phi(x, z)f(x)\mu(dx) \right] \eta(dz) \\ &= \int f(x) \left[\int \Phi(x, z)\eta(dz) \right] \mu(dx) \\ &= \mathbb{E}\left[f(X)\phi(X)\right] = \mathbb{E}\left[\phi(X)Y\right], \end{aligned}$$

which, in accordance with property (j) of conditional expectations (see Section D of the Appendix), concludes the proof. □

F. Weak Convergence of Measures

In the following we assume that $\{\mu, \mu_n : n \geq 1\}$ are probability measures on some metric space $(\mathbb{S}, d)$ and we denote by "$\Rightarrow$" the classical weak convergence of probability measures, i.e., $\mu_n \Rightarrow \mu$ if

$$\forall g \in \mathcal{C}_B(\mathbb{S}) : \lim_{n\to\infty} \int g(s)\mu_n(ds) = \int g(s)\mu(ds).$$

Theorem F.1. ***The Portmanteau Theorem***: *The following assertions are equivalent:*

(i) $\mu_n \Rightarrow \mu$.

(ii) For any closed subset F of $\mathbb{S}$ it holds that $\limsup_n \mu_n(F) \leq \mu(F)$,

(iii) For any open subset G of $\mathbb{S}$ it holds that $\liminf_n \mu_n(G) \geq \mu(G)$,

(iv) If A is a continuity set of μ, i.e., $\mu(\partial A) = 0$ then $\lim_n \mu_n(A) = \mu(A)$.

Theorem F.2. ***The Extension Theorem***: *If $\mu_n \Rightarrow \mu$ then for any measurable mapping g satisfying:*

(i) g is uniformly integrable with respect to $\{\mu_n : n \geq 1\}$,

(ii) the set of discontinuities D_g of g satisfies $\mu(D_g) = 0$,

it holds that

$$\lim_{n\to\infty} \int g(s)\mu_n(ds) = \int g(s)\mu(ds).$$

Theorem F.3. ***Prokhorov Theorem***: *Assume that the metric space $(\mathbb{S}, d)$ is separable.*

(i) Any tight family of probability measures is relatively compact[4] *with respect to the topology induced by the weak convergence.*

(ii) If, in addition, $(\mathbb{S}, d)$ is complete then any relatively compact family of probability measures is tight.

For a proof of these results see, e.g., [8].

[4] i.e., every sequence has an weakly convergent subsequence.

G. Functional Analysis

Here we list several standard results in functional analysis which are mentioned in this thesis. For the proofs of the results stated below we refer to [19].

Theorem G.1. ***Banach-Steinhaus Theorem****: Let $\mathcal{U}$ be a norm space, $\mathcal{V}$ be a Banach space and let*

$$\{\Phi_n : n \in \mathbb{N}\} \subset \mathfrak{L}_B(\mathcal{V}, \mathcal{U})$$

be a sequence of bounded operators such that

$$\forall \boldsymbol{x} \in \mathcal{V} : \sup_{n \in \mathbb{N}} \|\Phi_n(\boldsymbol{x})\|_{\mathcal{U}} < \infty.$$

Then it holds that $\sup_n \|\Phi_n\| < \infty$.

Theorem G.2. ***Banach-Alaoglu Theorem****: Let $\mathcal{V}$ be a norm space and let us denote by $\mathcal{V}^*$ its topological dual. Then, the set*

$$\{\Phi \in \mathfrak{L}(\mathcal{V}, \mathbb{R}) : \|\Phi\| \leq 1\} \subset \mathcal{V}^*$$

is compact in the weak- topology. In particular, it follows that any strongly bounded subset of $\mathcal{V}^*$ is relatively compact, i.e., its closure is compact, in the weak-* topology.*

In the following, we assume that $(\mathbb{S}, d)$ is a locally compact metric space and we denote by $\mathcal{C}_K(\mathbb{S}) \subset \mathcal{C}(\mathbb{S})$ the linear space of continuous functions f with compact support, i.e., there exists some compact $K \subset \mathbb{S}$ such that $f(s) = 0$ for each $s \notin K$. Note that, by Weierstrass's Theorem, any continuous function is bounded on compact sets and it follows that $\mathcal{C}_K(\mathbb{S}) \subset \mathcal{C}_B(\mathbb{S})$. The following result shows that the topological dual of any linear space which includes $\mathcal{C}_K(\mathbb{S})$ is a subspace of $\mathcal{M}(\mathcal{S})$, i.e., a space of measures.

Theorem G.3. ***Riesz Representation Theorem****: If $T : \mathcal{C}_K(\mathbb{S}) \to \mathbb{R}$ is a linear functional on $\mathcal{C}_K(\mathbb{S})$ there exist a unique Radon measure $\mu \in \mathcal{M}(\mathcal{S})$ such that*

$$\forall f \in \mathcal{C}_K(\mathbb{S}) : Tf = \int f(s)\mu(ds).$$

In addition, the operator norm of T coincides with the total variation norm of μ, i.e.,

$$\|T\| = \|\mu\| = |\mu|(\mathbb{S})$$

and it follows that the functional T is bounded (in particular, continuous) if and only if μ is a finite measure, i.e., $\|\mu\| < \infty$.

H. Overview of Weakly Differentiable Distributions

Name (Base)	Distribution (μ_θ)	Weak Derivative (μ'_θ)	c_θ	μ_θ^+	μ_θ^-
Bernoulli: β_θ $(\mathcal{F}, v_p), p \geq 0$	$(1-\theta)\cdot\delta_{x_1} + \theta\cdot\delta_{x_2}$	$\delta_{x_2} - \delta_{x_1}$	1	δ_{x_2}	δ_{x_1}
Binomial: $B^0_{n,\theta}$ $(\mathcal{F}, v_p), p \geq 0$	$\sum_{j=0}^n \binom{n}{j}(1-\theta)^j\theta^{n-j}\cdot\delta_j$	$\sum_{j=0}^n \binom{n}{j}\frac{(1-\theta)^{j-1}}{\theta^{1+j-n}}[n(1-\theta)-j]\cdot\delta_j$	$\frac{n}{\theta}$	$B^0_{n,\theta}$	$B^1_{n,\theta}$
Poisson: P^0_θ $(\mathcal{F}, v_p), p \geq 0$	$\sum_{n=0}^\infty \frac{\theta^n}{n!}e^{-\theta}\cdot\delta_n$	$\sum_{n=0}^\infty \frac{n\theta^{n-1}-\theta^n}{n!}e^{-\theta}\cdot\delta_n$	1	P^1_θ	P^0_θ
mixed: μ_θ $(\mathcal{F}, v_p), p \geq 0$	$(1-\theta)\cdot\mu + \theta\cdot\eta$	$\eta - \mu$	1	η	μ
exponent: $\varepsilon_{1,\theta}$ $(\mathcal{F}, v_p), p \geq 0$	$\theta e^{-\theta x}\mathbb{I}_{(0,\infty)}(x)dx$	$(1-\theta x)e^{-\theta x}\mathbb{I}_{(0,\infty)}(x)dx$	$\frac{1}{\theta}$	$\varepsilon_{1,\theta}$	$\varepsilon_{2,\theta}$
uniform: ψ_θ $(\mathcal{C}, v_p),\ p \geq 0$	$\frac{1}{\theta}\mathbb{I}_{[0,\theta)}(x)dx$	$\frac{1}{\theta}\cdot\delta_\theta - \frac{1}{\theta^2}\mathbb{I}_{[0,\theta)}(x)dx$	$\frac{1}{\theta}$	δ_θ	ψ_θ
Pareto: π_θ $(\mathcal{C}, v_p),\ p < \beta$	$\frac{\beta\theta^\beta}{x^{\beta+1}}\mathbb{I}_{(\theta,\infty)}(x)dx$	$\frac{\beta^2\theta^{\beta-1}}{x^{\beta+1}}\mathbb{I}_{(\theta,\infty)}(x)dx - \frac{\beta}{\theta}\cdot\delta_\theta$	$\frac{\beta}{\theta}$	π_θ	δ_θ
Gaussian: γ_θ $(\mathcal{F}, v_p), p \geq 0$	$\frac{1}{\theta\sqrt{2\pi}}e^{-\frac{(x-a)^2}{2\theta^2}}dx$	$\frac{(x-a)^2-\theta^2}{\theta^4\sqrt{2\pi}}e^{-\frac{(x-a)^2}{2\theta^2}}dx$	$\frac{1}{\theta}$	m_θ	γ_θ

Tab. 5.1: An overview on differentiability properties.

Table 5.1 presents weak derivatives of some distributions on $\mathbb{R}$, commonly used in practice. For each distribution, an instance of a weak derivative and a suitable Banach-base are provided. Continuous distributions are given by means of their Lebesgue densities. The following notation has been used:

- $B^k_{n,\theta}$, for $0 \leq k \leq n$, denotes the distribution of the number of successes in a sequence of n independent Bernoulli experiments with probability of success θ, conditioned on the event that the first k experiments were successful, i.e., $B^k_{n,\theta} = \sum_{j=0}^{n-k}\binom{n-k}{j}\theta^{n-k-j}(1-\theta)^j\cdot\delta_{k+j}$.
- P^k_θ, for $k \geq 0$, denotes the k-units shift of the Poisson distribution, i.e., the distribution of $X + k$, where X is an Poisson variable with rate θ. In formula: $P^k_\theta := \sum_{n=0}^\infty \frac{\theta^n}{n!}e^{-\theta}\cdot\delta_{k+n}$.
- $\varepsilon_{n,\theta}$, for $n \geq 1$, denotes the Erlang distribution, i.e., the distribution of the sum of n independent exponential variables with rate θ. In formula: $\frac{\theta^n x^{n-1}}{(n-1)!}e^{-\theta x}\mathbb{I}_{(0,\infty)}(x)dx$.
- m_θ denotes the double-sided Maxwell distribution. Precisely, if (X, Y, Z) is a 3-dimensional vector such that its components are independent standard gaussian variables and V denotes its magnitude, i.e., $V = \sqrt{X^2+Y^2+Z^2}$ then m_θ denotes the distribution of $a+\theta SV$, where S is a variable taking values $\{\pm 1\}$ with probability 1/2, independent of V. In formula: $m_\theta(dx) := \frac{(x-a)^2}{\theta^3\sqrt{2\pi}}e^{-\frac{(x-a)^2}{2\theta^2}}dx$.

SUMMARY

MEASURE–VALUED DIFFERENTIATION FOR FINITE PRODUCTS OF MEASURES: THEORY AND APPLICATIONS

This thesis is devoted to the theory of weak differentiation of measures. The basic observation is that, formally, the weak derivative of a parameter-dependent probability distribution μ_θ is in general a finite signed measure which can be represented as the re-scaled difference between two probability distributions. This fact allows for a useful representations of the derivative $\frac{d}{d\theta}\mathbb{E}_\theta[g(X)]$ of the expected value $\mathbb{E}_\theta[g(X)]$, for some predefined class $\mathcal{D}$ of cost-functions g, where X is a random variable with distribution μ_θ.

Many mathematical models are described by a finite family of independent random variables and this is the reason why differentiability properties as well as representations for weak derivatives of product measures are studied in this thesis. To develop the theory, concepts and results from measure theory and functional analysis are required and the necessary prerequisites are presented in Chapter 1.

In Chapter 2 we develop the theory of first-order differentiation. Main results, such as the product rule of weak differentiation and a representation theorem for the weak derivatives of product measures, are established. A product rule for weak differentiation of probability measures was conjectured (without a proof) in [48]. At the end of the chapter two gradient estimation examples are provided.

In Chapter 3 we illustrate how the theory of measure-valued differentiation can be applied in order to establish bounds on perturbations for general stochastic models. Special attention is paid to the sequence of waiting times in the G/G/1 queue for which we show that the strong stability property holds true provided that the service-time distribution is weakly differentiable with respect to some class of sub-exponential cost-functions.

In Chapter 4 we extend our analysis to higher-order differentiation, which leads us to establish a measure-valued differential calculus. Analyticity issues are also treated and Taylor series approximation examples are provided.

Eventually, in Chapter 5 we apply the results established in Chapter 4 to the class of discrete event systems whose state dynamic can be formalized into a matrix-vector multiplication in some general, non-conventional algebra (e.g., max-plus or min-plus algebra). A key result shows that, for some class of polynomially bounded cost-functions, weak differentiability of two random matrices X_θ and Y_θ is inherited by their generalized product $X_\theta \odot Y_\theta$, which allows us to develop a weak differential calculus for random matrices.

SAMENVATTING

MAATWAARDIGE DIFFERENTIATIE VOOR EINDIGE PRODUCTMATEN: THEORIE EN TOEPASSINGEN

In dit proefschrift wordt een theorie van zwakke differentiatie van kansverdelingen gepresenteerd. De fundamentele observatie is dat de zwakke afgeleide van een kansverdeling μ_θ, die afhankelijk is van een parameter θ, kan herschreven worden als het gescaleerd verschil tussen twee kansverdelingen. Dit feit leidt naar nuttige representaties van de afgeleide $\frac{d}{d\theta}\mathbb{E}_\theta[g(X)]$ van de gemiddelde waarde $\mathbb{E}_\theta[g(X)]$, voor iedere g uit een vooraf gedefinieerd onderverdeling $\mathcal{D}$ van kostenfuncties, waarbij X is een stochastische variabele met kansverdeling μ_θ.

Veel wiskundige modellen zijn beschreven door een eindige verzameling van onafhankelijk stochastische variabelen en dit is de reden waarom zwakke afgeleiden van producten van kansverdelingen in dit proefschrift onderzocht worden. Voor de opbouw van de theorie zijn resultaten uit de maattheorie en de functionaal analyse nodig. Deze worden dan ook in Hoofdstuk 1 voorgesteld.

Hoofdstuk 2 behandelt eerste-orde zwakke differentiatie. Hoofdresultaten zoals de productregel voor zwakke differentiatie en de representatiestelling van het zwakke afgeleiden van productmaten worden aangetoond. Een productregel voor zwakke differentiatie van kansverdelingen was verondersteld (zonder bewijs) in [48]. Twee voorbeelden van gradiënt schatters beëindingen dit hoofdstuk.

In Hoofdstuk 3 laten we zien hoe de theorie van de differentiatie van kansverdelingen kan worden toegepast om bovengrenzen voor storingen van parameter-afhankelijke stochastische modellen te berekenen. Bijzondere aandacht wordt besteed aan de wachttijden van het G/G/1 wachtrijsysteem. Het hoofdresultaat van dit hoofdstuk laat zien dat zwakke differentieerbaarheid van de bedientijden, "sterke stabiliteit" van de stationaire kansverdeling van de wachttijden geeft, met betrekking tot een bepaalde klasse van sub-exponentiële kostenfuncties.

In Hoofdstuk 4 breiden wij onze analyse uit naar hogere-orde differentiatie en een zwakke differentiaalrekening voor maatwaardige functies wordt voorgesteld. Een onderzoek op het gebied van Taylor-reeks ontwikkelingen gebaseerd op zwakke afgeleiden wordt ook uitgevoerd.

Afsluitend passen wij in Hoofdstuk 5 de resultaten uit Hoofdstuk 4 toe op discrete-tijd systemen die kunnen worden beschreven door een matrix-vector vermenigvuldiging in een aantal algemene, niet-conventionele algebras (bv. max-plus of min-plus algebra). Een belangrijk resultaat is dat voor sommige klassen van polynomiaal begrensde kostenfuncties zwakke differentieerbaarheid van twee stochastische matrices X_θ en Y_θ de zwakke differentieerbaarheid van het algemeen product $X_\theta \odot Y_\theta$ impliceerd. Dit feit laat ons toe om een zwakke differentiaalrekening voor stochastische matrices te ontwikkelen.

BIBLIOGRAPHY

[1] Ayhan, H. and Baccelli, F. Expressions for joint Laplace transforms for stationary waiting times in $(\max, +)$-linear systems with Poisson input. *Queueing Systems - Theory and Applications*, 37(4), pp. 291–328, 2001.

[2] Ayhan, H. and Seo, D. Tail probability of transient and stationary waiting times in $(\max, +)$-linear systems. *IEEE Transaction on Automatic Control*, 47, pp. 151–157, 2000.

[3] Ayhan, H. and Seo, D. Laplace transform and moments of waiting times in Poisson driven $(\max, +)$-linear systems. *Queueing Systems - Theory and Applications*, 37, pp. 405–436, 2001.

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

[5] Baccelli, F., Hasenfuß, S., and Schmidt, V. Expansions for steady-state characteristics of (max,+)-linear systems. *Stochastic Models*, 14, pp. 1–24, 1998.

[6] Baccelli, F. and Hong,D. Analytic expansions of (max,+)-linear Lyapunov exponents. *Annals of Applied Probability*, 10, pp. 779–827, 2000.

[7] Baccelli, F. and Schmidt, V. Taylor series expansions for Poisson-driven (max,+)-linear systems. *Annals of Applied Probability*, 6, pp. 138–185, 1996.

[8] Billingsley, B. *Weak Convergence of Probability Measures.* J.Wiley and Sons, New York, 1966.

[9] Bobrowski, A. *Functional Analysis for Probability and Stochastic Processes.* Cambridge University Press, Cambridge, 2005.

[10] Bourbaki, N. *Integration I.* Springer Verlag, New York, 2004.

[11] Buck, R.C. Bounded continuous functions on a locally compact space. *Michigan Math. Journal*, 5(2), pp. 95–104, 1958.

[12] Cao, X.R. The Mac-Laurin series for performance functions of Markov chains. *Adv.Appl.Prob*, 30, pp. 676–692, 1998.

[13] Cohen, J.E. Subadditivity, generalized products of random matrices and operation research. *SIAM Review*, 30(1), pp. 69–86, 1988.

[14] Cohn, D. *Measure Theory.* Birkäuser, Stuttgart, 1980.

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

[16] Dekker, R. and Hordijk, A. Average, sensitive and Blackwell optimal policies in denumerable Markov decision chains with unbounded rewards. Tech. report no. 83-36, Institute of Applied Mathematics and Computing Science, 1983.

[17] Dekker, R. and Hordijk, A. Average, sensitive and Blackwell optimal policies in denumerable Markov decision chains with unbounded rewards. *Mathematics of Operations Research*, 13, pp. 395–421, 1988.

[18] Devroye, L.P. Inequalities for completion times of stochastic pert networks. *Mathematics of Operation Research*, 4, pp.441–447, 1979.

[19] Dunford, N. *Linear Operators.* Pure and applied mathematics. New York, Interscience Publishers, 1971.

[20] Heidergott, B. Variability expansion for performance characteristics $(\max, +)$-linear systems. *Proceedings of the 6th Workshop on Discrete Event Systems (WODES)*, pp. 245–250, Zaragoza/Spain, 10/2002.

[21] Heidergott, B. *Max-Plus linear Stochastic Systems and Perturbation Analysis.* The International Series of Discrete Event Systems, 15. Springer-Verlag, Berlin, 2006.

[22] Heidergott, B. and H.Leahu. Bounds on perturbations for discrete event systems. *Proceedings of the 8th Workshop on Discrete Event Systems (WODES)*, pp. 378–383, Ann Arbor/Michigan, 07/2006.

[23] Heidergott, B. and H.Leahu. Series expansions of generalized matrix products. *Proceedings of the 44th IEEE Conference on Decision and Control and European Control Conference*, pp. 7793–7798, Sevilla/Spain, 12/2005.

[24] Heidergott, B. and Hordijk, A. Taylor series expansions for stationary Markov chains. *Advances in Applied Probability*, 23, pp.1046–1070, 2003.

[25] Heidergott, B. and Hordijk, A. Single-run gradient estimation via measure-valued differentiation. *IEEE Transactions on Automatic Control*, 49, pp. 1843–1846, 2004.

[26] Heidergott, B., Hordijk, A., and Leahu, H. Strong bounds on perturbations. *(to appear)*, 2007.

[27] Heidergott, B., Hordijk, A., and Weißhaupt, H. Measure-valued differentiation for stationary Markov chains. *Mathematics of Operations Research*, 31, pp. 154–172, 2006.

[28] Heidergott, B. and Leahu, H. Differentiability of product measures. Research Memorandum 2008-5, Vrije Universiteit Amsterdam, The Netherlands, 2008.

[29] Heidergott, B., Olsder, G.J., and van der Woude, J. *Max Plus at Work: Modelling and Analysis of Synchronized Systems.* Princeton University Press, 2006.

[30] Heidergott, B. and Vázquez-Abad, F. Gradient estimation for a class of systems with bulk services: a problem in public transportation. Tech. report no. 057/4, Tinbergen Institute, Amsterdam, 2003.

[31] Heidergott, B. and Vazquez-Abad. F. Measure-valued differentiation for random horizon problems. *Markov Processes and Related Fields*, 12, pp. 509–536, 2006.

[32] Heidergott, B. and Vázquez-Abad, F. Measure valued differentiation for Markov chains. *Journal of Optimization and Applications*, 136, pp. 187–209, 2008.

[33] Heidergott, B., Vázquez-Abad, F., and Volk-Makarewicz, W. Sensitivity estimation for gaussian systems. *European Journal of Operational Research*, 187, pp. 193–207, 2008.

[34] Hordijk, A. and Yushkevich, A.A. Blackwell optimality in the class of all policies in Markov decision chains with a Borel state space and unbounded rewards. *Mathematics of Operations Research*, 50, pp. 421–448, 1999.

[35] Kartashov, N. *Strong Stable Markov Chains.* VSP, Utrecht, 1996.

[36] Kelley, J.L. *General Topology.* Springer, New-York, 1975.

[37] Kingman, J.F.C. Subadditive ergodic theory. *The Annals of Probability*, 1(6), pp. 883–909, 1973.

[38] Kumagai, S. An implicit function theorem: Comment. *Journal of Optimization Theory and Applications*, 31(2), pp. 285–288, 1980.

[39] Kushner, H. and Vázquez-Abad, F. Estimation of the derivative of a stationary measure with respect to a control parameter. *Journal of Applied Probability*, 29, pp. 343–352, 1992.

[40] Kushner, H. and Vázquez-Abad, F. Stochastic approximations for systems of interest over an infinite time interval. *SIAM Journal on Control and Optimization*, 29, pp. 712–756, 1996.

[41] Lewis, D.R. Integration with respect to vector measures. *Pacific Journal of Mathematics*, 33(1), pp. 157–165, 1970.

[42] Lipman, S. On dynamic programming with unbounded rewards. *Mangement Science*, 21, pp. 1225–1233, 1974.

[43] Loynes, R.M. The stability of a queue with non-independent inter-arrival and service times. *Proceedings of the Cambridge Philosophical Society*, 58, pp. 497-520, 1962.

[44] Meyn, S.P. and Tweedie, R.L. *Markov Chains and Stochastic Stability.* Springer, London, 1993.

[45] Moisil, Gr.C. Sur une représentation des graphes qui interviennent dans l'économie des transports. *Communications de l'Académie de la R.P. Roumaine*, 10, pp.647–652, 1960.

[46] Nachbin, L. Weighted approximation for algebras and modules of continuous functions: real and self-adjoint complex cases. *Annals of Mathematics*, 81(2), pp. 289–302, 1965.

[47] Pflug, G. Derivatives of probability measures - concepts and applications to the optimization of stochastic systems. In *Lecture Notes in Control and Information Science 103*, pages 252–274. Springer, Berlin, 1988.

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

[49] Pich, M., Loch, C., and de Meyer, A. On uncertainty, ambiguity and complexity in project management. *Management Science*, 75(2), pp. 137–176, 1996.

[50] Prolla, J.B. Bishop's generalized Stone-Weierstrass theorem for weighted spaces. *Mathematische Annalen*, 191(4), pp. 283–289, 1971.

[51] Prolla, J.B. Weighted space of vector-valued continuous functions. *Annali di matematica pura ed applicata*, 89(1), pp. 145–157, 1971.

[52] Rao, R.R. Relations between weak and uniform convergence of measures with applications. *The Annals of Mathematical Statistics*, 33(2), pp. 659–680, 1962.

[53] Rudin, W. *Principles of Mathematical Analysis*. McGraw-Hill, 1976.

[54] Rudin, W. *Functional Analysis*. McGraw-Hill, 1991.

[55] Seidel, W., Kocemba, K.v., and Mitreiter, K. On Taylor series expansions for waiting times in tandem queues: An algorithm for calculating the coefficients and an investigation of the approximation error. *Performance Evaluation*, 38(3), pp. 153–173, 1999.

[56] Semadeni, Z. *Banach Spaces of Continuous Functions*. Polish Scientific Publishers, Warszawa, 1971.

[57] Summers, W.H. Dual spaces of weighted spaces. *Transactions of the American Mathematical Society*, 151(1), pp. 323–333, 1970.

[58] Summers, W.H. A representation theorem for biequicontinuous completed tensor products of weighted spaces. *Transactions of the American Mathematical Society*, 146, pp. 121–131, 1970.

[59] van den Boom, T., De Schutter, B., and Heidergott, B. Complexity reduction in MPC for stochastic $(\max, +)$-linear systems by variability expansion. *Proceedings of the 41st IEEE Conference on Decision and Control*, pp. 3567–3572, Las Vegas/Nevada, 12/2002.

[60] Wells, J. Bounded continuous vector-valued functions on a locally compact space. *Michigan Math. Journal*, 12(1), pp. 119–126, 1965.

INDEX

LIST OF SYMBOLS AND NOTATIONS

$D_\theta^v(\cdot)$: $[\mathcal{D}]_v$-domain of convergence of the weak Taylor series, 89

D_g: set of discontinuities of g, 48

$P_g(\theta)$: performance measure, iii

$R_\theta^v(\cdot)$: $[\mathcal{D}]_v$-radius of convergence of the weak Taylor series, 89

T: the completion time of a SAN, 95

X^ι: canonical embedding of X into the extended algebra, 109

$X_\theta^{(n)}$: the n^{th}-order derivative of the random matrix X_θ, 107

$[\cdot]^\pm$: Hahn-Jordan decomposition, 7

Π_*: product measure mapping, 46

$\Theta \subset \mathbb{R}$: set of parameters, iii

$\Theta_s \subset \Theta$: the stability set, 77

$\bar{\tau}$: the conjugate of τ in the extended algebra, 109

β_θ: Bernoulli distribution, 41

χ^0: initial distribution (Markov chain), 69

δ_θ: Dirac distribution, 13

ℓ: the Lebesgue measure on a Euclidean space, 7

$\equiv_\mathcal{D}$: weak equality w.r.t. the space of test functions $\mathcal{D}$, 111

$\hat{\partial}_\theta$: gradient estimator for $P_g(\theta)$, iv

$\mathbb{N}$: the set of natural numbers, 5

$\mathbb{R}$: the set of real numbers, iii

$\mathbb{S}$: complete, separable metric space, iv

$\mathbb{S}_v \subset \mathbb{S}$: support of v, 9

$\mathbf{L}^*$: Lipschitz constant corresponding to the product measure, 65

$\mathbf{L}_\mu^v$: Lipschitz constant of μ_* in v-norm, 63

$\mathbf{M}^*$: Lipschitz constant corresponding to the product measure (non-negative cost-functions), 65

$\mathbf{M}_\mu^v$: Lipschitz constant of μ_* in v-norm for non-negative cost-functions, 63

$\mathbf{T}_n(\mu, \theta, \xi)$: n^{th} Taylor polynomial, 88

$\mathcal{A}$: algebra of matrices, 101

$\mathcal{A}_*$: extended algebra of matrices, 108

$\mathcal{C}$: space of continuous mappings, 4

$\mathcal{C}^+ \subset \mathcal{C}$: non-negative mappings, 4

$\mathcal{C}_B \subset \mathcal{C}$: bounded mappings, 4

$\mathcal{C}_v \subset \mathcal{C}$: v-bounded mappings, 9

$\mathcal{D}$: space of test functions, with typically $\mathcal{D} = \mathcal{C}, \mathcal{F}$, 19

$\mathcal{D}_p$: the $[\mathcal{D}]_v$-space induced on $\mathcal{D}$ by the weight v_p, 104

$\mathcal{F}$: space of Borel measurable mappings, 5

$\mathcal{F}_B \subset \mathcal{F}$: bounded mappings, 5

$\mathcal{L}^p$: space of p-integrable mappings, 6

$\mathcal{M}$: space of regular measures, 8

$\mathcal{M}^+ \subset \mathcal{M}$: positive measures, 8

$\mathcal{M}^1 \subset \mathcal{M}$: probability measures, 8

$\mathcal{M}_B \subset \mathcal{M}$: finite measures, 8

$\mathcal{M}_v \subset \mathcal{M}$: v-finite measures, 20

$\mathcal{M}_v^1 \subset \mathcal{M}$: $\mathcal{M}_v \cap \mathcal{M}^1$, 20

$\mathcal{P}$: the set of paths through a SAN, 94

$\mathcal{S}$: Borel field on $\mathbb{S}$, 5

$\mathcal{U}_A$: uniform distribution on A, 40

$\mathcal{V}^*$: topological dual of $\mathcal{V}$, 16

$\mathfrak{M}_{m,n}$: class of m, n matrices, 100

μ_*: measure-valued mapping, 9

$\odot$: generalized matrix product, 99

$\otimes$: functional tensor product, 24

$\overline{\mathfrak{M}}_{m,n}$: extended algebra of matrices, 108

$\overset{\mathcal{D}}{\Longrightarrow}$: weak convergence of measures w.r.t. the space of test-functions $\mathcal{D}$, 11

π_θ: Pareto distribution, 52

ψ_θ: uniform distribution, 42

$\varepsilon_{n,\theta}$: Erlang distribution, 42

$\varnothing$: the null measure, 43

$\vec{v}$: tensor product $v_1 \otimes \ldots \otimes v_n$, 46

$\wp_{m,n}$: canonical metric on $\mathfrak{M}_{m,n}$, 100

g^ι: ι-extension of the mapping g, 109

v_p: polynomial weight of degree p, 104

ACKNOWLEDGMENTS

This thesis is the result of the research carried out at Vrije Univeristeit Amsterdam and Technische Universiteit Delft. Both institutions have offered me a very supporting and stimulating work environment. I would also like to acknowledge the contribution of Dutch Foundation for Technology (Technologiestichting STW) which supported financially my four years contract with VU Amsterdam, within the research project "Modeling and Analysis of Operations in Railway Networks: the Influence of Stochasticity" (joint project between VU Amsterdam and TU Delft). However, many individuals deserve my gratitude.

I am especially grateful to my supervisor Bernd Heidergott for both professional and personal aspects of our collaboration which led to the completion of this monograph. While his remarks and suggestions over the material were determinant for me in finding (I hope) the best way to put forward the results of my research I would also like to mention that both his optimism and faith in me helped me to get over the critical moments. Therefore, I would like to take this opportunity to thank him for guiding me in the last four years and for his notable contribution in the development of my research profile.

I am indebted to Prof.Dr. F.M. Dekking, Prof.Dr. G.M. Koole, Prof.Dr. G.Ch. Pflug, Dr. A.N. de Ridder and Dr. F.M. Spieksma for taking the time to read the manuscript and providing me with valuable feedback.

I am thankful to all my friends for their constant support. The list being quite long, I would like, however, to mention my good friends Daniel and Vlad for making my accommodation period in the Netherlands smoother and for making themselves available whenever I needed their help. Special thanks to my friend Ana, which apart from being a very good friend, helped me with virtually any computer-related problem that I have encountered in my work.

Finally, my gratitude goes to my beloved parents, Veve and Viorel, for their unconditional support. Thank you for being there for me and for backing up and understanding my decisions, regardless of your will.

TINBERGEN INSTITUTE RESEARCH SERIES

The Tinbergen Institute is the Institute for Economic Research, which was founded in 1987 by the Faculties of Economics and Econometrics of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam and Vrije Universiteit Amsterdam. The Institute is named after the late Professor Jan Tinbergen, Dutch Nobel Prize laureate in economics in 1969. The Tinbergen Institute is located in Amsterdam and Rotterdam. The following books recently appeared in the Tinbergen Institute Research Series:

378 M.R.E. BRONS, Meta-analytical studies in transport economics: Methodology and applications.

379 L.F. HOOGERHEIDE, Essays on neural network sampling methods and instrumental variables.

380 M. DE GRAAF-ZIJL, Economic and social consequences of temporary employment.

381 O.A.C. VAN HEMERT, Dynamic investor decisions.

382 Z. AOVOV, Liking and disliking: The dynamic effects of social networks during a large-scale information system implementation.

383 P. RODENBURG, The construction of instruments for measuring unemployment.

384 M.J. VAN DER LEIJ, The economics of networks: Theory and empirics.

385 R. VAN DER NOLL, Essays on internet and information economics.

386 V. PANCHENKO, Nonparametric methods in economics and finance: dependence, causality and prediction.

387 C.A.S.P. S, Higher education choice in The Netherlands: The economics of where to go.

388 J. DELFGAAUW, Wonderful and woeful work: Incentives, selection, turnover, and workers' motivation.

389 G. DEBREZION, Railway impacts on real estate prices.

390 A.V. HARDIYANTO, Time series studies on Indonesian rupiah/USD rate 1995 2005.

391 M.I.S.H. MUNANDAR, Essays on economic integration.

392 K.G. BERDEN, On technology, uncertainty and economic growth.

393 G. VAN DE KUILEN, The economic measurement of psychological risk attitudes.

394 E.A. MOOI, Inter-organizational cooperation, conflict, and change.

395 A. LLENA NOZAL, On the dynamics of health, work and socioeconomic status.

396 P.D.E. DINDO, Bounded rationality and heterogeneity in economic dynamic models.

397 D.F. SCHRAGER, Essays on asset liability modeling.

398 R. HUANG, Three essays on the effects of banking regulations.

399 C.M. VAN MOURIK, Globalisation and the role of financial accounting information in Japan.

400 S.M.S.N. MAXIMIANO, Essays in organizational economics.

401 W. JANSSENS, Social capital and cooperation: An impact evaluation of a womens empowerment programme in rural India.

402 J. VAN DER SLUIS, Successful entrepreneurship and human capital.

403 S. DOMINGUEZ MARTINEZ, Decision making with asymmetric information.

404 H. SUNARTO, Understanding the role of bank relationships, relationship marketing, and organizational learning in the performance of peoples credit bank.

405 M.Â. DOS REIS PORTELA, Four essays on education, growth and labour economics.

406 S.S. FICCO, Essays on imperfect information-processing in economics.

407 P.J.P.M. VERSIJP, Advances in the use of stochastic dominance in asset pricing.

408 M.R. WILDENBEEST, Consumer search and oligopolistic pricing: A theoretical and empirical inquiry.

409 E. GUSTAFSSON-WRIGHT, Baring the threads: Social capital, vulnerability and the well-being of children in Guatemala.

410 S. YERGOU-WORKU, Marriage markets and fertility in South Africa with comparisons to Britain and Sweden.

411 J.F. SLIJKERMAN, Financial stability in the EU.

412 W.A. VAN DEN BERG, Private equity acquisitions.

413 Y. CHENG, Selected topics on nonparametric conditional quantiles and risk theory.

414 M. DE POOTER, Modeling and forecasting stock return volatility and the term structure of interest rates.

415 F. RAVAZZOLO, Forecasting financial time series using model averaging.

416 M.J.E. KABKI, Transnationalism, local development and social security: the functioning of support networks in rural Ghana.

417 M. POPLAWSKI RIBEIRO, Fiscal policy under rules and restrictions.

418 S.W. BISSESSUR, Earnings, quality and earnings management: the role of accounting accruals.

419 L. RATNOVSKI, A Random Walk Down the Lombard Street: Essays on Banking.

420 R.P. NICOLAI, Maintenance models for systems subject to measurable deterioration.

421 R.K. ANDADARI, Local clusters in global value chains, a case study of wood furniture clusters in Central Java (Indonesia).

422 V.KARTSEVA, Designing Controls for Network Organizations: A Value-Based Approach.

423 J. ARTS, Essays on New Product Adoption and Diffusion.

424 A. BABUS, Essays on Networks: Theory and Applications.

425 M. VAN DER VOORT, Modelling Credit Derivatives.

426 G. GARITA, Financial Market Liberalization and Economic Growth.

427 E.BEKKERS, Essays on Firm Heterogeneity and Quality in International Trade.