This is a corrected Version of:

Derivatives of Markov kernels and their Jordan decomposition.

The version originally published in the journal of Applied Analysis contains some errors that have been corrected in this version.

Bernd Heidergott
Vrije Universiteit Amsterdam, Department of Econometrics
De Boelelaan 1105, 1081 HV Amsterdam, the Netherlands
email: bheidergott@feweb.vu.nl

Arie Hordijk
Leiden University, Mathematical Institute
P.O.Box 9512, 2300 RA Leiden, the Netherlands
email: hordijk@math.leidenuniv.nl

Heinz Weisshaupt*
University of Vienna, Department of Statistics
A-1010 Vienna, Universitaetsstrasse 5/3, Austria†
e-mail: heinz.weisshaupt@univie.ac.at

June 14, 2010

Abstract

We study a particular class of transition kernels that stems from differentiating Markov kernels in the weak sense. Sufficient conditions are established for this type of kernels to admit a Jordan–type decomposition. The decomposition is explicitly constructed.

MSC Index **Primary:** 60J10, 26A26 **Secondary:** 28A15, 47B34

Keyword: Markov kernels, weak differentiation, Jordan decomposition

---

*Responsibility for errors possibly contained in this version, if any.

†Main parts of the paper have been written during a stay of Heinz Weisshaupt at the research Institute EURANDOM located on the campus of the Technische Universiteit Eindhoven (TU/e)

1 Introduction

Let $(P_\vartheta)_{\vartheta\in\Theta}$ be a parametric family of Markov kernels P_θ from a measurable space $(X,\mathcal{X})$ to a locally compact space Y, with $\vartheta \in \Theta \subset \mathbb{R}$, and let $\mathcal{C}_c(Y)$ denote the set of continuous real–valued mappings with compact support on Y. The family of Markov kernels $(P_\vartheta)_{\vartheta\in\Theta}$ is called *weakly differentiable* at ϑ if for any $x \in X$ a finite signed Baire measure $P'_\vartheta(x;\cdot)$ on Y exists such that for any $g \in \mathcal{C}_c(Y)$:

$$\frac{d}{d\vartheta}\int g(y)\,P_\vartheta(x;dy) \;=\; \int g(y)\,P'_\vartheta(x;dy)\,. \tag{1}$$

This definition of weak differentiability differs slightly from the original one in [7]. The concept of weak differentiability of measures has been successfully applied in different mathematical contexts (see [3], [4], [11], [8], [12], [13]).

For applications related to the sensitivity analysis of Markov Chains [5] it is important that we are able to obtain P'_ϑ by a conditional sampling procedure. Conditional sampling procedures are within standard mathematical theory governed by Markov kernels and measurable transformations. It is therefore desirable to show that P'_ϑ can be represented as a measurabely scaled difference of two Markov kernels, i.e., it is desirable to show that

$$P'_\vartheta(x;A) = c_{P_\vartheta}(x)\cdot[Q^+_\vartheta(x;A) - Q^-_\vartheta(x;A)], \tag{2}$$

where Q^+_ϑ and Q^-_ϑ are Markov kernels and $c_{P_\vartheta} : X \to \mathbb{R}$ is a $\mathcal{X}$-measurable function.

In this paper, we give sufficient conditions for P'_ϑ to possess a representation as a scaled difference of two Markov kernels. Specifically, we show that sup-norm boundedness of the linear functional $g \mapsto \int g(y)P'_\vartheta(x;dy)$ on $\mathcal{C}_c(Y)$ together with second countability and compactness of Y is sufficient for P'_ϑ to decompose as a scaled difference of two Markov kernels.

We note that for fixed x equation (2) gives just a scaled decomposition of the signed measures $P'_\vartheta(x;.)$. Hence Q^+_ϑ and Q^-_ϑ are easily obtained from the point-wise Jordan decomposition of P'_ϑ. This does however, not establish that $c_{P_\vartheta}(.)$ is a measurable function and $Q^+_\vartheta(.,A)$ and $Q^-_\vartheta(.,A)$ are Markov kernels, i.e, it does not establish measurability of $c_{P_\vartheta}(.)$, $Q^+_\vartheta(.,A)$ and $Q^-_\vartheta(.,A)$ for any measurable set A. The analysis put forward in this paper will establish sufficient conditions for the measurability of c_{P_ϑ}, $Q^+_\vartheta(.,A)$ and $Q^-_\vartheta(.,A)$, providing thus an answer to the question when (2) holds with Q^+_ϑ and Q^-_ϑ Markov kernels and $c_{P_\vartheta} : X \to \mathbb{R}$ an $\mathcal{X}$-measurable function.

Further we provide by Example 2 a counterexample that indicates that (local) compactness of Y is strictly essential for our results.

The paper is organized as follows. Section 2 introduces measure theoretic and topological concepts (compare with [9] and [15]) and shows that, under suitable conditions, the finite signed Baire measures $P'_\vartheta(x,.)$ constitute indeed a kernel P'_ϑ. In Section 3, a Jordan Type decomposition of P'_ϑ is explicitely constructed. Section 4 is concerned with counterexamples and an extension of our results to infinite products of locally compact second countable spaces.

2 Conditional Integrals and Kernels

Throughout the paper we let Y always denote a locally compact second countable Hausdorff space. In the case that we assume in addition Y to be compact this will be specified. We denote by $\mathcal{Y}$ the σ–field of Baire measurable subsets of Y, i.e., the σ–field generated by the compact subsets of Y.

Remark 1 *Y is a polish (completely metrizable and separable) space.*[1]

On a second countable locally compact space the Borel–field (the σ–field generated by the open or closed sets) and the Baire–field coincide.[2] *Thus, $\mathcal{Y}$ is the σ-field generated by the family of open sets in Y.*

The space $\mathbb{R}^n$ and any submanifold of it constitutes a locally compact second countable space.

Let X be an arbitrary set and let $\mathcal{X}$ be an arbitrary σ–field on X. Let $\mathcal{B}_b(Y)$ denote the family of real–valued bounded $\mathcal{Y}$–measurable functions on Y, let $\mathcal{C}_c(Y)$ denote the family of continuous functions with compact support on Y and let $\mathcal{B}(X)$ denote the family of real–valued $\mathcal{X}$–measurable functions on X. We note that $\mathcal{C}_c(Y) \subseteq \mathcal{B}_b(Y)$

We call a Baire measurable function $g : Y \to \mathbb{R}$ *simple* if an integer $n \in \mathbb{N}$ and, for $i \leq n$, sets $B_i \in \mathcal{Y}$ and constants $\gamma_i \in \mathbb{R}$ exist such that

$$g(y) = \sum_{i=1}^{n} \gamma_i \mathbf{1}_{B_i}(y) , \quad y \in Y .$$

The family of Baire measurable simple functions on Y is denoted by $\mathcal{B}_{simp}(Y)$.

Let $\|\cdot\|$ denote the sup-norm on $\mathcal{B}_b(Y)$. We call a set $\mathcal{G} \subset \mathcal{B}_b(Y)$ *uniformly bounded* or *sup–norm bounded* if $\sup_{g \in \mathcal{G}} \|g\| < \infty$. We say that a sequence $(g_n)_{n\in\mathbb{N}}$ of functions $g_n \in \mathcal{B}_b(Y)$ is uniformly bounded if the set $\{g_n \mid n \in \mathbb{N}\}$ is uniformly bounded.

We say that a linear functional $J : \mathcal{C}_c(Y) \to \mathbb{R}$ is an *integral* if it is sup-norm bounded, i.e, J is bounded on uniformly bounded subsets of $\mathcal{C}_c(Y)$ (see also [1] Section 13.1). We say that a linear functional $\widetilde{J} : \mathcal{B}_b(Y) \to \mathbb{R}$ is an *extended integral* if it is sup-norm bounded on $(\mathcal{B}_b(Y), \|.\|)$.

A measure μ on Y is regular, if for $E \in \mathcal{Y}$ we have that

$$\sup\{\mu(K) \mid K \subseteq E,\ K \text{ compact }\} = \mu(E) = \inf\{\mu(O) \mid O \supseteq E,\ O \text{ open }\}.$$

Definition 1 *A kernel $P(\cdot,\cdot)$ from X to Y is a function $P : X \times \mathcal{Y} \to \mathbb{R}$ such that $P(x,\cdot)$ is for any $x \in X$ a finite signed measure on $(Y,\mathcal{Y})$ and $x \mapsto P(x,B)$ is for any $B \in \mathcal{Y}$ a $\mathcal{X}$–measurable function on X. We say that the kernel is Markov (or a Markov kernel) if for any $x \in X$ the measure $P(x,\cdot)$ is a probability measure. We denote the space of all kernels from X to Y by $\mathcal{P}(X,Y)$.*

[1] The one-point compactification of a second countable locally compact space is again second countable. We thus conclude by Urysohn's metrization theorem ([15] 23.1) that the one-point compactification of Y is metrizable and we conclude further by [15] Exercise 24B 4 that the one-point compactification of Y is even completely metrizable. Since Y is an open subset of its one-point compactification and thus a G_δ-subset of a completely metrizable space, we conclude from [15] 24.17 that Y is itself completely metrizable. Thus by second countability Y is a polish space.

[2] This holds true since any open set in a second countable locally compact space is a countable union of compact sets.

Definition 2 *A conditional integral $I(\cdot,\cdot)$ from X to $\mathcal{C}_c(Y)$ is a function $I : X \times \mathcal{C}_c(Y) \to \mathbb{R}$ such that*

- *$I(x,\cdot)$ is an integral (i.e. a linear functional on $\mathcal{C}_c(Y)$ which is sup-norm bounded) and*
- *$x \mapsto I(x,f)$ is for any $f \in \mathcal{C}_c(Y)$ a $\mathcal{X}$- measurable function on X.*

We denote the space of conditional integrals from X to $\mathcal{C}_c(Y)$ by $\mathcal{I}(X,Y)$.

Definition 3 *Let Z denote an arbitrary Hausdorff space. We say that a function $F : \mathcal{B}_b(Y) \mapsto Z$ is point-wise sequentially continuous on uniformly bounded subsets of $\mathcal{B}_b(Y)$ if for any uniformly bounded point-wise convergent sequence $(g_n)_{n\in\mathbb{N}}$ in $\mathcal{B}_b(Y)$ with limit $g \in \mathcal{B}_b(Y)$ we have that $\lim F(g_n) = F(g)$.*

Given a function space $\mathcal{F} \subseteq \mathbb{R}^X$. We say that a set $S \subset \mathcal{F}$ is point-wise sequentially closed if S contains all the limits (that are in $\mathcal{F}$!) of point-wise convergent sequences $(g_n)_{n\in\mathbb{N}}$ whose elements g_n are in S. We say that a set $\overline{S}$ is the point-wise sequential closure of a set S if $\overline{S}$ is the smallest point-wise sequentially closed set containing S. A set S is point-wise sequentially dense in a set T if T is a subset of the sequential closure $\overline{S}$ of S. (For more details on sequential continuity and measurable functions see [9] Section 3.2.)

Proposition 1 *Let $K \subseteq Y$ be compact and let $O \subseteq Y$ be open with compact closure such that $K \subset O$. Then there exists a continuous function $f : Y \to [0,1]$ such that $f(K) = 1$ and $f(Y \setminus O) = 0$.*

Proof: This follows by an application of the Urysohn Lemma (see [15] 15.6) to K and $Y \setminus O \cup \{\infty\}$ in the one-point compactification (see [15] 19.2 and 19A) $Y \cup \{\infty\}$ of Y, since any compact space is normal (see [15] 17.10). □

Lemma 1 *It holds that:*

(a) The space $\mathcal{B}(X)$ is point–wise sequentially closed in $\mathbb{R}^X$.

(b) The function-space $\mathcal{B}_{simp}(Y)$ is point–wise sequentially dense in $\mathcal{B}_b(Y)$.

(c) The function-space $\mathcal{C}_c(Y)$ is point–wise sequentially dense in $\mathcal{B}_b(Y)$.

Proof: (a) Is the well known fact that a limit of a point–wise convergent sequence of measurable functions is again measurable.

(b) Is a consequence of the fact that any measurable function is the point wise limit of a sequence of simple functions. (See for example Corollary 3.2.1 of [9].)

(c) Given an arbitrary compact set K we can by second countability and local compactness of Y choose a sequence $(O_n)_{n\in\mathbb{N}}$ of open sets such that $O_{n+1} \subset O_n$, $\bigcap_n O_n = K$ and the closures $\overline{O_n}$ are compact. By Proposition 1 we find continuous functions f_n such that $f_n(K) = 1$ and $f_n(Y \setminus O_n) = 0$. Since $\overline{O_n}$ is compact these functions f_n possess compact support. Thus, $1_K = \lim_{n\in\mathbb{N}} f_n(x)$, and 1_K lies in the point-wise sequential closure of $\mathcal{C}_c(Y)$. Since any open set O is - by second countability and local compactness of Y - the countable union of compact sets, we see that also any function 1_O and thus especially the function 1_Y belongs to the sequential closure of $\mathcal{C}_c(Y)$. (That 1_Y belongs to the sequential closure of $\mathcal{C}_c(Y)$ can also be easily seen using a countable partition of unity.)

Hence, any finite linear combination of functions 1_A with $A \in \mathcal{Y}$ belongs to the sequential closure of $\mathcal{C}_c(Y)$ and thus by (b) the space $\mathcal{B}_b(Y)$ is a subset of the sequential closure of $\mathcal{C}_c(Y)$. So we obtain (c) from (b). $\square$

Proposition 2 *(Representation Theorem of Riesz) Let $J : \mathcal{C}_c(Y) \to \mathbb{R}$ be an integral. Then there exists a unique finite signed measure μ on $(Y, \mathcal{Y})$ such that the extended integral $\widetilde{J} : \mathcal{B}_b(Y) \to \mathbb{R}$ given by*

$$g \mapsto \int g(y)\ \mu(dy)$$

is the unique extension of J to $\mathcal{B}_b(Y)$ that is point-wise sequentially continuous on uniformly bounded sets.

Proof: This is a consequence of [9] Section 5.2 Exercise 3 and Lemma 1 of this article.

Remark 2 *Note that any finite signed measure μ on Y is regular, i.e., its Jordan decomposition $\mu = \mu^+ + \mu^-$ decomposes the measure into two regular (positive) measures μ^+ and μ^-. This follows from the usual results on the Jordan decomposition ([9] Section 4.2) and from [9] Scholium 5.2.*

Lemma 2 *Any conditional integral $I \in \mathcal{I}(X, Y)$ extends uniquely to a conditional integral $\widetilde{I} : X \times \mathcal{B}_b(Y) \mapsto \mathbb{R}$ such that for any $x \in X$ the function $\widetilde{I}(x, \cdot)$ is point-wise sequentially continuous on uniformly bounded subsets of $\mathcal{B}_b(Y)$. Moreover, there exists a one-one correspondence between kernels and conditional integrals $G : \mathcal{P}(X, Y) \to \mathcal{I}(X, Y)$ given by*

$$\big[G(P)\big](x, f) \;=\; \int f(y)\ P(x, dy) \quad \textit{for all} \quad f \in \mathcal{C}_c(Y), \tag{3}$$

or, if we prefer to consider the extensions $\widetilde{I}$ of the conditional integrals I, by

$$\big[\widetilde{G(P)}\big](x, g) \;=\; \int g(y)\ P(x, dy)\ ,$$

for all $g \in \mathcal{B}_b(Y)$.

Proof of Lemma 2: For notational convenience we call the above extension $\widetilde{I}$ of a conditional integral I the *extended conditional integral.* The proof consists of 3 steps:

Step 1: By Proposition 2 there exists for an arbitrary conditional integral $I \in \mathcal{I}(X, Y)$ and for any $x \in X$ a unique measure $P(x, .)$ on $(Y, \mathcal{Y})$ and a unique extended integral $\widetilde{I}(x, .)$ on $\mathcal{B}_b(Y)$ such that

$$I(x, f) \;=\; \int f(y)\ P(x, dy) \quad \text{for all} \quad f \in \mathcal{C}_c(Y)\ , \tag{4}$$

$$\widetilde{I}(x, g) \;=\; \int g(y)\ P(x, dy) \quad \text{for all } g \in \mathcal{B}_b(Y)\ , \tag{5}$$

and $\widetilde{I}(x, \cdot)$ is the unique extension of $I(x, \cdot)$ that is sequentially point-wise continuous on uniformly bounded sets.

Step 2: In the second step we show that the functions $x \mapsto \widetilde{I}(x, g)$ are $\mathcal{X}$-measurable, for $g \in \mathcal{B}_b(Y)$ arbitrary, i.e., we show that $\widetilde{I}$ is a conditional extended integral. Further we show that the unique corresponding function $P : X \times \mathcal{Y} \to \mathbb{R}$, defined in the first step, is a kernel.

Let $\mathbb{R}^X$ be endowed with the topology of point-wise convergence. Define an operator $T : \mathcal{B}_b(Y) \to \mathbb{R}^X$ by

$$[T(g)](x) \; = \; \widetilde{I}(x, g) \; .$$

The fact that, for arbitrary $x \in X$, the integral $\widetilde{I}(x, \cdot)$ is point-wise sequentially continuous on uniformly bounded sets of $\mathcal{B}_b(Y)$ implies that T is also point-wise sequentially continuous.

Further, $f \in \mathcal{C}_c(Y)$ implies by definition of T and the fact that $I \in \mathcal{I}(X, Y)$ that

$$T(f) \; = \; \big[x \to I(x, f)\big] \quad \in \mathcal{B}(X) \, , \tag{6}$$

i.e., we have that $T(\mathcal{C}_c(Y)) \subseteq \mathcal{B}(X)$.

By (6) together with Lemma 1 (c) and the point-wise sequential continuity of T, we obtain that $T(\mathcal{B}_b(Y)) \subseteq \mathcal{B}(X)$. In other words, we obtain that $g \in \mathcal{B}_b$ implies that $x \mapsto \widetilde{I}(x, g)$ is $\mathcal{X}$-measurable. The fact that $x \mapsto \widetilde{I}(x, g)$ is $\mathcal{X}$-measurable implies in the case that g is the characteristic function of a set B that $x \mapsto P(x, B)$ is $\mathcal{X}$-measurable. Thus, P is a kernel and (as already noted in the first step) by Proposition 2 unique.

In the first two steps we have shown that any integral $I \in \mathcal{I}(X, Y)$ corresponds with an unique kernel $P \in \mathcal{P}(X, Y)$ and an unique extended integral $\widetilde{I}$. Further we know by equation (4) and (3) that this correspondence is given by G^{-1}. In the third step we show that any $P \in \mathcal{P}(X, Y)$ corresponds with an unique $I = G(P) \in \mathcal{I}(X, Y)$.

Step 3: We show that any kernel P corresponds with an unique conditional integral I. We do this by showing that any kernel P corresponds to a unique extended conditional integral. That any kernel P gives us by formula (5) for any x an extended integral $\tilde{I}(x, .)$ is trivial. To show that $\tilde{I}$ is a conditional extended integral note that for any simple function $g = \sum_{i=1}^{n} \gamma_i \mathbf{1}_{B_i} \in \mathcal{B}_{simp}$ we have:

$$\widetilde{I}(x, g) = \sum_i \gamma_i P(x, B_i) \; .$$

So for $g \in \mathcal{B}_{simp}$ the function $x \mapsto \widetilde{I}(x, g)$ is a finite sum of $\mathcal{X}$-measurable functions and thus itself $\mathcal{X}$- measurable. It remains to be shown that $x \mapsto \widetilde{I}(x, g)$ is for any $g \in \mathcal{B}_b(Y)$ a $\mathcal{X}$-measurable function. We do this by arguments analogous to the arguments provided in step 2 as will be explained in the following.

Let T denote the operator defined in step 2. Recall that T is point-wise sequentially continuous. Furthermore, $f \in \mathcal{B}_{simp}(Y)$ implies (by definition of T and the fact that for $g \in \mathcal{B}_{simp}(Y)$ the function $x \mapsto \widetilde{I}(x, g)$ is $\mathcal{X}$- measurable) that:

$$T(f) \; = \; \big[x \to \widetilde{I}(x, f)\big] \quad \in \mathcal{B}(X) \, , \tag{7}$$

i.e., we have that $T(\mathcal{B}_{simp}(Y)) \subseteq \mathcal{B}(X)$.

By (7) together with Lemma 1 (b) and point-wise sequential continuity of T, we obtain that $T(\mathcal{B}_b(Y)) = \mathcal{B}(X)$. In other words, we obtain that $g \in \mathcal{B}_b(Y)$

implies that $x \mapsto \widetilde{I}(x,g)$ is $\mathcal{X}$-measurable.□

Now we define weak differentiability of conditional integrals and kernels. By an interval we always mean an interval with nonempty interior. A function $\phi : \Theta \to \mathbb{R}$ is called differentiable if it is differentiable in the interior of Θ and one sided differentiable at the boundary points of Θ. Derivatives and one sided derivatives, respectively, are denoted by $\frac{d\phi(\vartheta)}{d\vartheta}$.

Definition 4 *Let Θ be an interval in $\mathbb{R}$ and let $\vartheta \mapsto I_\vartheta$ be a path in (mapping from Θ to) the space $\mathcal{I}(X,Y)$. We say that $\vartheta \mapsto I_\vartheta$ is weakly differentiable if*

$$\frac{dI_\vartheta(x,f)}{d\vartheta} \text{ exists for all } (x,f) \in X \times \mathcal{C}_c(Y)$$

If $\vartheta \to I_\vartheta$ is weakly differentiable then we say that it is bounded weakly differentiable if

$$\sup_{\substack{f \in \mathcal{C}_c(Y) \\ |f| \leq 1}} \left| \frac{dI_\vartheta(x,f)}{d\vartheta} \right| < \infty \,,$$

for any $x \in X$.

We say that a path $\vartheta \mapsto P_\vartheta$ in the space $\mathcal{P}(X,Y)$ of kernels is bounded weakly differentiable if the corresponding path $\vartheta \mapsto G(P_\vartheta)$ in the space $\mathcal{I}(X,Y)$ of conditional integrals is bounded weakly differentiable.

Theorem 1 *If the path $\vartheta \mapsto P_\vartheta$ in the space $\mathcal{P}(X,Y)$ is bounded weakly differentiable, then the weak derivative can be represented by a path $\vartheta \mapsto P'_\vartheta$ in the space $\mathcal{P}(X,Y)$. The connection between $\vartheta \mapsto P_\vartheta$ and $\vartheta \mapsto P'_\vartheta$ is given by*

$$\int f(y)P'_\vartheta(x,dy) \;=\; \frac{d\int f(y)P_\vartheta(x,dy)}{d\vartheta} \quad \text{for } f \in \mathcal{C}_c(Y)\,.$$

Proof: Let $I_\vartheta = G(P_\vartheta)$ be the corresponding path in the space of conditional integrals. Define for any $(x,f) \in X \times \mathcal{C}_c(Y)$ the function $I'_\vartheta(x,f)$ by

$$I'_\vartheta(x,f) := \frac{dI_\vartheta(x,f)}{d\vartheta}$$

Let $(h_n)_{n\in\mathbb{N}}$ be an arbitrary sequence of positive reals which goes to 0 as n tends to ∞. Then for $f \in \mathcal{C}_c(Y)$ we have:

$$[x \mapsto I'_\vartheta(x,f)] = \left[x \mapsto \frac{dI_\vartheta(x,f)}{d\vartheta}\right] = \left[x \mapsto \lim_{n\to\infty} \frac{I_{\vartheta+h_n}(x,f) - I_\vartheta(x,f)}{h_n}\right]\,.$$

Thus, $x \mapsto I'_\vartheta(x,f)$ is for $f \in \mathcal{C}_c(Y)$ a limit of a sequence of $\mathcal{X}$-measurable functions and therefore itself $\mathcal{X}$-measurable. The fact that I is bounded weakly differentiable implies that $I'(x,.)$ is bounded for any $x \in X$. Thus, $I'(x,\cdot)$ is for any $x \in X$ an integral and $I'(\cdot,\cdot)$ is thus itself a conditional integral. By the correspondence between conditional integrals and kernels (Lemma 2) we obtain a kernel $P' = G^{-1}(I')$. The formula connecting P' and P is clear from the correspondence between P', P and I', I and the definition of I'. □

3 Jordan Decomposition of Weak Derivatives of Markov Kernels

Definition 5 *Given a kernel $P \in \mathcal{P}(X, Y)$ we define the absolute value $|P|$ of the kernel as follows:*

$$|P|(x, B) \;=\; \sup_{\substack{A \in \mathcal{Y} \\ A \subseteq B}} 2 \cdot P(x, A) - P(x, B)\,, \quad x \in X\,,\; B \in \mathcal{Y}\,.$$

Lemma 3 *The absolute value $|P|$ of a kernel $P \in \mathcal{P}(X, Y)$ is again a kernel.*

Proof: That for any $x \in X$ the absolute value $|P|(x, .)$ is a finite measure is a well known fact and it remains to be shown that the function

$$x \mapsto |P|(x, B) \tag{8}$$

is $\mathcal{X}$–measurable for any $B \in \mathcal{Y}$. By a monotone class argument it suffices to show that (8) holds for any $B \in \mathcal{A}$ for some set–field $\mathcal{A}$ that generates $\mathcal{Y}$.

Thus let β be a countable basis of Y and let $\mathcal{A}$ be the set–field generated by β. Then, $\mathcal{A}$ is countable and generates the σ–field $\mathcal{Y}$. For any set $C \in \mathcal{Y}$ and any measure μ on $(Y, \mathcal{Y})$ there exists a sequence $(A_n)_{n \in \mathbb{N}}$ of sets $A_n \in \mathcal{A}$ such that $\lim \mu(A_n \triangle C) = 0$ (see [10] Lemma A.24). Thus, the function

$$x \mapsto |P|(x, B)$$

is for any $B \in \mathcal{A}$ the point–wise supremum over the countable family

$$\Big\{ x \mapsto 2 \cdot P(x, A) - P(x, B) \;:\; A \in \mathcal{A} \;\text{ and }\; A \subseteq B \Big\}$$

of $\mathcal{X}$-measurable functions and thus itself $\mathcal{X}$-measurable. □

Definition 6 *We say that a kernel is positive if $P(x, B) \geq 0$ for all $(x, B) \in X \times \mathcal{Y}$. We say that a pair of kernels (P^+, P^-) forms a decomposition of a kernel P if P^+ and P^- are positive kernels and $P(x, B) = P^+(x, B) - P^-(x, B)$. We say that this decomposition is minimal or Jordan if for any other decomposition (Q^+, Q^-) of P we have $P^+(x, B) \leq Q^+(x, B)$ and $P^-(x, B) \leq Q^-(x, B)$.*

Corollary 1 *Any kernel $P \in \mathcal{P}(X, Y)$ possesses a Jordan decomposition.*

Proof: For $(x, B) \in X \times \mathcal{Y}$ define

$$P^+(x, B) \;:=\; \frac{|P|(x, B) + P(x, B)}{2}$$

and

$$P^-(x, B) := \frac{|P|(x, B) - P(x, B)}{2}\,.$$

Then, $P^+(x, B), P^-(x, B) \geq 0$, $P^+(x, \cdot), P^-(x, \cdot)$ are measures, and by an application of Lemma 3 the functions $x \mapsto P^+(x, B)$ and $x \mapsto P^-(x, B)$ are $\mathcal{X}$-measurable. It is also clear that the decomposition is minimal. □

Theorem 2 *Suppose that the path $\vartheta \mapsto P_\vartheta$ in the space $\mathcal{P}(X,Y)$ is bounded weakly differentiable and that for any ϑ the kernel P_ϑ is Markov. Suppose further that Y is compact. Then there exist for any ϑ Markov kernels Q_ϑ^+ and Q_ϑ^- from X to Y and a $\mathcal{X}$-measurable function $c_\vartheta : X \to \mathbb{R}$ such that the weak derivative P'_ϑ of P_ϑ decomposes in the form*

$$P'_\vartheta(x,B) \;=\; c_\vartheta(x)\left(Q_\vartheta^+(x,B) - Q_\vartheta^-(x,B)\right) \qquad \forall (x,B) \in X \times \mathcal{Y}\,.$$

Proof: By Theorem 1, the weak derivative P'_ϑ is for any ϑ a kernel and by Corollary 1, P'_ϑ possesses a Jordan decomposition $(P'^+_\vartheta, P'^-_\vartheta)$, i.e., $P'_\vartheta = P'^+_\vartheta - P'^-_\vartheta$ with $P'^+_\vartheta, P'^-_\vartheta$ positive kernels. Since the P_ϑ are Markov kernels and Y is compact, we have $P'^+_\vartheta(x,Y) = P'^-_\vartheta(x,Y)$ by [14] Remark 2.11. Let $c_\vartheta : X \to \mathbb{R}$ be defined by

$$c_\vartheta(x) \;:=\; P'^+_\vartheta(x,Y) = P'^-_\vartheta(x,Y)\,.$$

Since P_ϑ^+ is a kernel, the function $c_\vartheta(\cdot)$ is $\mathcal{X}$- measurable. For $B \in \mathcal{Y}$ let

$$Q_\vartheta^+(x,B) \;:=\; \frac{1}{c(x)} P'^+_\vartheta(x,B) \;\text{ for all }\; x \;\text{ with }\; c_\vartheta(x) > 0\,,$$

$$Q_\vartheta^-(x,B) \;:=\; \frac{1}{c(x)} P'^-_\vartheta(x,B) \;\text{ for all }\; x \;\text{ with }\; c_\vartheta(x) > 0\,.$$

For $x \in X$ with $c_\vartheta(x) = 0$ and $B \in \mathcal{Y}$ set

$$Q_\vartheta^+(x,B) = Q_\vartheta^-(x,B) = \mu(B)\,,$$

where μ is an arbitrary probability measure. Then Q_ϑ^+ as well as Q_ϑ^- are Markov kernels. □

Remark 3 *That P'_ϑ decomposes according to (2) is due to the fact that the kernels P_ϑ are Markov and Y is compact. Formula (2) is not any more true for the decomposition of derivatives of general (non Markovian) kernel valued functions $\vartheta \mapsto P_\vartheta$ or non compact Y.*

4 Examples and an Extension of our Results

Weak differentiability with respect to $\mathcal{C}_c(Y)$ is even in the case $Y = \mathbb{R}$ not the same as differentiability with respect to the Baire σ-field as is shown by the following example concerning just measures that can be interpreted as constant Markov kernels.

Example 1 *Let $\Theta = \mathbb{R}$ and let ν_θ be the uniform distribution on $[\theta, 1+\theta]$. Then ν'_θ exists weakly and $\nu'_\theta = -\delta_\theta + \delta_{(1+\theta)}$, while the derivative with respect to the Baire σ-field does not exist. Further examples of this phenomenon are provided in [11], [12], [13] and [8].*

Inspecting the proofs of the Theorems 1 and 2 one may be under the impression that (local) compactness of Y is not essential for our analysis and that the theorems could be obtained via the Daniell-Stone theorem ([9] Theorem 3.3 or [2] Theorem 4.5.2) instead of the Riesz representation theorem. In the remainder of this section we will show that this is not true (not even for probability measures). Moreover we will present an extension of our result to products of infinitely many (locally) compact spaces.[3]

Definition 7 *Let I be some set and let $I_0 \subseteq I$ be an arbitrary finite subset. Let $pr_{I_0} : Y^I \to Y^{I_0}$ be the projection onto the coordinates in I_0. Let $\otimes_I \mathcal{Y} := \sigma(\bigcup_{i\in I} pr_i^{-1}(\mathcal{Y}))$ be the product σ-algebra on Y^I. We call a set $Z \in \otimes_I \mathcal{Y}$ a cylinder set if $Z = pr_{I_0}^{-1}(B)$ for some finite set $I_0 \subseteq I$ and some arbitrary $B \in \otimes_{I_0} \mathcal{Y}$ ([6] Section 2.2). We call a function $f : Y^I \to \mathbb{R}$ a $\mathcal{C}_c$-cylinder function if there exists a finite set $I_0 \subseteq I$ and a function $f_0 \in \mathcal{C}_c(Y^{I_0})$ such that $f = f_0 \circ pr_{I_0}$.*

Theorem 3 *Let $\Theta \subset \mathbb{R}$ be an interval and let for $\vartheta \in \Theta$ $P_\vartheta \in \mathcal{P}(X, Y^I)$ and suppose that for any finite set $I_0 \subseteq I$ the path $\vartheta \mapsto P_\vartheta \circ pr_{I_0}^{-1}$ is bounded weakly differentiable. Then the weak derivative can be represented by a path $\vartheta \mapsto P'_\vartheta$ in the space $\mathcal{P}(X, Y^I)$. The connection between $\vartheta \mapsto P_\vartheta$ and $\vartheta \mapsto P'_\vartheta$ is given by*

$$\int f(y) P'_\vartheta(x, dy) \;=\; \frac{d \int f(y) P_\vartheta(x, dy)}{d\vartheta} \quad \text{for } \mathcal{C}_c\text{-cylinder functions } f\,.$$

Moreover if P_ϑ is a Markov kernel and Y is compact then the derivative decomposes in the form

$$P'_\vartheta(x, B) \;=\; c_{P_\vartheta}(x)\left(Q_\vartheta^+(x, B) - Q_\vartheta^-(x, B)\right) \quad \forall (x, B) \in X \times (\otimes_I \mathcal{Y})$$

where Q_ϑ^+ and Q_ϑ^- are Markov kernels and $c_{P_\vartheta}(.)$ is a measurable function.

Proof: This theorem is easily derived from Theorem 1 and Theorem 2 together with Kolmogorovs consistency theorem ([10] Theorem B.133).

The following example illustrates that (local) compactness is a necessary hypothesis in Theorems 1 and 2. More specifically, an example is provided where the derivative of a path of probability measures $\vartheta \mapsto \nu_\vartheta$ on the Hilbert space ℓ^2 fails to be a curve of signed measures, but incorporates cylindrical signed measures, i.e., set-functions such that only their finite dimensional projections are σ-additive.

Example 2 *Let $N(0, \sigma^2)$ denote the normal distribution with mean 0 and variance σ^2. Let δ_0 denote Dirac measure at the point $(0, 0, 0, \dots) \in \mathbb{R}^{\mathbb{N}}$. Let for $s \in [0, 1]$ measures m_s on the Borel σ-algebra $\mathcal{B}$ of $\mathbb{R}^{\mathbb{N}}$ be defined by*

$$m_s(B) := \left[\prod_{n\in\mathbb{N}} N\left(0, n^{-(1+s)}\right)\right](B).$$

Let $\Theta = [0, 1]$ and let measures μ_ϑ be defined by

$$\mu_\vartheta(B) := \int_\vartheta^1 m_s(B)\, ds + \vartheta \cdot \delta_0(B).$$

Then $\vartheta \mapsto \mu_\vartheta(Z)$ is for any cylinder set Z differentiable on $[0, 1]$ and

$$\mu'_\vartheta(Z) = -m_\vartheta(Z) + \delta_0(Z).$$

[3]Note that the product of infinitely many locally compact spaces is locally compact if and only if all but finitely many factors are compact ([15] 18.6).

By Kolmogorovs consistency theorem ([10] Theorem B.133) the signed cylinder measure μ'_ϑ extends to a signed measure $\widetilde{\mu'_\vartheta}$.

Define further set-functions ν_ϑ on $\ell^2 := \{y \in \mathbb{R}^\mathbb{N} \mid \sum_{n\in\mathbb{N}} y_n^2 < \infty\}$ by

$$\nu_\vartheta(Z \cap \ell^2) = \mu_\vartheta(Z) \quad \textit{for cylinder sets} \quad Z.$$

Then all set-functions ν_ϑ extend uniquely to measures on the Borel σ-algebra $\mathcal{B}(\ell^2)$ of ℓ^2. We denote these measures again by ν_ϑ. The measures ν_ϑ are differentiable with derivatives ν'_ϑ such that for restrictions of $\mathcal{C}_c$-cylinder functions $f : \mathbb{R}^\mathbb{N} \to \mathbb{R}$ to ℓ^2 we have that

$$\int f|_{\ell^2}\ d\nu'_\vartheta = \int f\ d\mu'_\vartheta = \int f\ d[m_\vartheta + \delta_0]$$

and for cylinder sets Z we have that $\vartheta \mapsto \nu_\vartheta(Z \cap \ell^2)$ is differentiable and

$$\nu'_\vartheta(Z) = \mu'_\vartheta(Z) = -m_\vartheta(Z) + \delta_0(Z).$$

But ν'_0 does not extend to a signed measure $\widetilde{\nu'_0}$ on the Borel sets of ℓ^2. An extension $\widetilde{\nu'_0}$ would have to coincide with $\widetilde{\mu'_0}$ on the Borel sets of ℓ^2. Thus we would obtain the contradiction

$$0 = \mu'_0(\mathbb{R}^\mathbb{N}) = \nu'_0(\mathbb{R}^\mathbb{N} \cap \ell^2) = \widetilde{\nu'_0}(\ell^2) = \widetilde{\mu'_0}(\ell^2) = -m_0(\ell^2) + \delta_0(\ell^2) = 0 + 1 = 1.$$

□

Remark 4 *All theorems, lemmas, propositions and examples remain true if $\mathcal{C}_c(Y)$ is replaced by a uniformly dense subspace of $\mathcal{C}_c(Y)$.*

Acknowledgment We would like to thank the referees for their valuable remarks that helped improving the paper.

References

[1] J. Dieudonne *Grundzüge der modernen Analysis. Band 2. Übers. aus dem Französischen von Ludwig Boll und Horst Antelmann. 2., ber. u. erg. Aufl. (Foundations of modern analysis. Vol. 2). (German)* Wiesbaden etc.: Vieweg & Sohn. (1987).

[2] R.M. Dudley *Real analysis and probability. Repr.* Cambridge Studies in Advanced Mathematics. 74. Cambridge: Cambridge University Press, 2002.

[3] B. Heidergott. A weak derivative approach to optimization of threshold parameters in a multi-component maintenance system. *Journal of Applied Probability*, 38:386-406, 2001.

[4] B. Heidergott. Option pricing via Monte Carlo Simulation: A weak derivative approach. *Probability in Engineering and Informational Sciences*, 15:335- 349, 2001.

[5] B. Heidergott, A. Hordijk and H. Weisshaupt. Measure–Valued Differentiation for Stationary Markov Chains. *Mathematics of Operations Research* 31:154-172, 2006.

[6] I. Karatzas, S. E. Shreve. *Brownian motion and stochastic calculus. 2nd ed.* Graduate Texts in Mathematics, 113. New York etc.: Springer-Verlag, 1991

[7] G. Pflug. *Optimisation of Stochastic Models.* Kluwer Academic Publishers, Boston, 1996.

[8] G. Pflug, H. Weisshaupt. Probability Gradient Estimation by Set-Valued calculus and applications in Network Design. *SIAM Journal of Optimization* 15:898-914, 2005.

[9] I. Segal and R. Kunze. *Integrals and operators.* (2nd rev. and enl. ed.) Springer, Berlin, 1978.

[10] M. Schervish. *Theory of statistics.* Springer, New York, 1995.

[11] H. Weisshaupt. A measure-valued approach to convex set-valued dynamics. *Set-Valued Analysis.* 9:337-373, 2001.

[12] H. Weisshaupt. A characterization of parallelepipeds related to weak derivatives. *Acta Math. Hung.* 107:319-328, 2005.

[13] H. Weisshaupt A generalization of Lehmann's theorem on the comparison of uniform location experiments. *Acta Math. Hung.* 107:329-336, 2005.

[14] H. Weisshaupt On the differentiability of parameterized families of linear operators and the sensitivity of their stationary vectors. *Periodica Math. Hung.* Vol. 59 (2), 2009, pp 147-171

[15] S. Willard. *General topology.* Addison-Wesley, Mass., 1970.