Derivatives of Markov kernels and their Jordan decomposition
Heidergott, B.F.; Hordijk, A.; Weisshaupt, H.

Published: 01/01/2003

Document Version
Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

• A submitted manuscript is the author's version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

Citation for published version (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 10. Oct. 2018
Derivatives of Markov kernels and their Jordan decomposition

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 Düsseldorf, Mathematical Institute
Universitätsstraße 1, D-40225 Düsseldorf, Germany
email: Heinz.Weisshaupt@uni-duesseldorf.de

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 chain, weak differentiation, Jordan decomposition

*Main parts of this article have been written during a stay of Heinz Weisshaupt at EURANDOM, TU-Eindhoven
1 Introduction

Let $P_\theta$ be a family of Markov kernels from a measurable space $(X, \mathcal{X})$ to a locally compact space $Y$ (a precise definition will be given later in the text), with $\theta \in \Theta \subset \mathbb{R}$, and let $C_c(Y)$ denote the set of continuous real-valued mappings with compact support on $Y$. The Markov kernel $P_\theta$ is called \textit{weakly differentiable} at $\theta$ if for any $x \in X$ a finite signed measure $P'_\theta(x; \cdot)$ on $(Y, \mathcal{Y})$ exists such that for any $g \in C_c(Y)$:

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

This definition of weak differentiability is slightly more general than the original one in [4]: there (1) has to hold for any continuous bounded mapping $g$. Weak differentiability has been successfully applied to the theory of Markov chains. See [1] for an application to a problem in maintenance theory and [2] for an application to option pricing. The concept of weak differentiation is also related to finding optimal statistical tests, see [7]. For Markov chains, the following result is of particular interest: let $\pi_\theta$ denote the (unique) invariant distribution of $P_\theta$ (existence is assumed here), then it can be shown that

$$
\pi'_\theta = \pi_\theta \sum_{n=0}^{\infty} P'_\theta P^n_\theta, \quad (2)
$$

where $P'_\theta$ is defined through (1) and $P^n_\theta$ denotes the $n$ fold product of $P_\theta$, see [4, 3] for a proof and more details on weak differentiability. If $P'_\theta$ exists, then the fact that $P'_\theta(x; \cdot)$ fails to be a probability measure poses the problem of sampling from $P'_\theta$. For $x \in X$ fixed, we can represent $P'_\theta(x; \cdot)$ by its Jordan decomposition as a difference between two probability measures as follows. For a finite signed measure $\mu$ denote its Jordan decomposition by $[\mu]^+ - [\mu]^-$, i.e., $\mu = [\mu]^+ - [\mu]^-$ and $[\mu]^+, [\mu]^-$ are positive measures. Let

$$
c_{P_\theta}(x) = [P'_\theta]^+(x; X) = [P'_\theta]^-(x; X) \quad (3)
$$

and

$$
P'^+_{\theta}(x; \cdot) = \frac{[P'_\theta]^+(x; \cdot)}{c_{P_\theta}(x)}, \quad P^-_{\theta}(x; \cdot) = \frac{[P'_\theta]^-(x; \cdot)}{c_{P_\theta}(x)},
$$

then it holds, for all $g \in C_c(Y)$, that

$$
\int g(y) P'_\theta(x; dy) = c_{P_\theta}(x) \left( \int g(y) P'^+_{\theta}(x; dy) - \int g(y) P^-_{\theta}(x; dy) \right). \quad (4)
$$

For the above line of argument we fixed $x$. For $P'^+_{\theta}$ and $P^-_{\theta}$ to be Markov kernels, we have to consider $P'^+_{\theta}$ and $P^-_{\theta}$ as functions in $x$ and have to establish
measurability of $P_\theta^+(\cdot; A)$ and $P_\theta^-(\cdot; A)$ for any $A \in \mathcal{Y}$. The solution of this problem implies that $c_{P_\theta}(\cdot)$ in (3) is measurable as a mapping from $X$ to $\mathbb{R}$. A representation of $P_\theta^\prime$ through $(c_{P_\theta}(\cdot), P_\theta^+, P_\theta^-)$, with $c_{P_\theta}$ measurable and $P_\theta^\pm$ Markov kernels, is called a weak derivative of $P_\theta$. The existence of a weak derivative is of key importance for the statistical interpretation of (2) and for obtaining efficient unbiased gradient estimators.

In this paper, we give sufficient conditions for $P_\theta^\prime$ to possess a representation as scaled difference of two Markov kernels. Specifically, we show that uniform boundedness of $P_\theta^\prime (\cdot)$ (i.e., the supremum of $|\int g(y)P_\theta(x;dy)|$ over $g \in \mathcal{C}_c(Y)$ with $|g| \leq 1$ and $x \in X$ is finite) is together with a topological condition on $Y$ sufficient for $c_{P_\theta}(\cdot)$ in (3) to be measurable (and for $P_\theta^+$ and $P_\theta^-$ to be Markov kernels again). In conclusion we will show that uniform boundedness is sufficient for $P_\theta^\prime$ to admit a weak derivative.

The paper is organized as follows. Section 1 introduces the basic concepts and definitions. Section 2 shows that, under suitable conditions, the kernel $P_\theta^\prime$ as defined in (1) can be uniquely extended to the bounded Borel–measurable mappings. In Section 3 an explicit construct of a Jordan–type decomposition of $P_\theta^\prime$ is given.

2 Conditional Integrals and Kernels

We say that a topological space is second countable if its topology is generated by a countable basis, i.e., if there exists a countable family of open (or closed) sets which generates the topology. Throughout the paper we let $Y$ always denote a locally compact second countable Hausdorff space. We denote by $\mathcal{Y}$ the $\sigma$–field of Baire measurable subsets of $Y$, i.e., the $\sigma$–field generated by the compact subsets of $Y$.

Remark 1 On a second countable locally compact space the Borel–field (the $\sigma$–field generated by the open or closed sets) and the Baire–field coincide. (This holds true since any open set in a second countable locally compact space is a countable union of compact sets.) Thus, $\mathcal{Y}$ is the $\sigma$–field generated by the family of open sets in $Y$.

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

Remark 2 Notice that a metrizable space is second countable if and only if it is separable (see [8] Theorem 16.11). Conversely, a locally compact or even a compact space may be separable but not second countable. An example of
A separable compact space that fails to be second countable is provided by the Stone-Cech compactification of the natural numbers.

Let $X$ be an arbitrary set and let $\mathcal{X}$ be an arbitrary $\sigma$–field on $X$. Let $\mathcal{B}_b(Y)$ be the family of real–valued bounded $\mathcal{Y}$–measurable functions on $Y$, let $\mathcal{C}_c$ 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 call a Baire measurable function, say $g$, simple if and only 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

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

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

We note that $\mathcal{C}_c(Y) \subset \mathcal{B}_b(Y)$ and define the supremum norm $\| \cdot \|$ on $\mathcal{B}_b(Y)$ by

$$\|g\| := \sup_{y \in Y} |g(y)|.$$  

We call a set $G \subset \mathcal{B}_b(Y)$ uniformly bounded or sup–norm bounded if

$$\sup_{g \in G} \|g\| < \infty.$$  

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 bounded on uniformly bounded subsets of $\mathcal{C}_c(Y)$ (such functionals may also be called sup-norm bounded). We say that a linear functional $\tilde{J} : \mathcal{B}_b(Y) \to \mathbb{R}$ is an extended integral if it is bounded on uniformly bounded subsets $G$ of $\mathcal{B}_b(Y)$.

We say that a sequence $(f_n)_{n \in \mathbb{N}}$ of functions $f_n$ from some set $S$ to a Hausdorff space $V$ converges point–wise if $\lim_{n \to \infty} f_n(s)$ exists for any $s \in S$.

**Definition 1** A kernel $P(\cdot, \cdot)$ from $X$ to $Y$ is a function $P : X \times Y \to \mathbb{R}$ 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 $\mathcal{P}(X, Y)$.

**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
\[ I(x, \cdot) \text{ is an integral (i.e. a linear functional on } \mathcal{C}_c(Y) \text{ which is sup-norm bounded)} \text{ and} \]

\[ x \mapsto I(x, f) \text{ is for any } f \in \mathcal{C}_c(Y) \text{ a } X \text{- measurable function on } X. \]

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

**Definition 3** Let \( Z \) denote an arbitrary Hausdorff space. We say that a function \( F : \mathcal{B}_b(Y) \to Z \) 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 \( S \subseteq \mathcal{F} \) is point-wise sequentially closed if \( S \) contains all the limits which 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 \([5]\) Section 3.2.)

**Proposition 1** Let \( K \subseteq Y \) be compact and let \( O \subseteq Y \) be open with compact closure such that \( K \subseteq O \). 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 \([8]\) 15.6) to \( K \) and \( Y \setminus O \cup \{\infty\} \) in the one-point compactification (see \([8]\) 19.2 and 19A) \( Y \cup \{\infty\} \) of \( Y \), since any compact space is normal (see \([8]\) 17.10).

\( \square \)

**Lemma 1** It holds that:

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

(b) The function-space \( \mathcal{B}_\text{simp}(Y) \) 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 re-formulation 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 \([5]\).)

(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 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 function \( 1_A \) with \( A \in \mathcal{Y} \) belongs to the sequential closure of \( \mathcal{C}_c(Y) \). So we obtain (c) from (b).

\[ \square \]

**Lemma 2** Any conditional integral \( I \in \mathcal{I}(X,Y) \) extends uniquely to a conditional integral \( \tilde{I} : X \times \mathcal{B}_b(Y) \mapsto \mathbb{R} \) such that for any \( x \in X \) the function \( \tilde{I}(x,\cdot) \) 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 \( G : \mathcal{P}(X,Y) \mapsto \mathcal{I}(X,Y) \) given by

\[ [G(P)](x,f) = \int f(y) \ P(x,dy) \text{ for all } f \in \mathcal{C}_c(Y), \tag{5} \]

or, if we prefer to consider the extensions \( \tilde{I} \) of the conditional integrals \( I \), by

\[ [\tilde{G}(P)](x,g) = \int g(y) \ P(x,dy), \]

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

We call the above extension \( \tilde{I} \) of a conditional integral \( I \) the extended conditional integral. By Lemma 1 there is a one–one correspondence between conditional integrals \( I \) and their extensions \( \tilde{I} \).

**Proof of Lemma 2:** The proof consists of 3 steps. First we show that for a given conditional integral \( I \in \mathcal{I}(X,Y) \) there exists for any \( x \in X \) a unique measure \( P(x,\cdot) \) on \( (Y,\mathcal{Y}) \). Then we show that the integrals \( I(x,\cdot) \) on \( \mathcal{C}_c(Y) \) extend for arbitrary \( x \in X \) uniquely to extended integrals \( \tilde{I}(x,\cdot) \) on \( \mathcal{B}_b(Y) \).

**Step 1:** Let \( I \) be a given conditional integral. According to the Riesz representation theorem, there exists for any \( x \in X \) a unique measure \( P(x,\cdot) \) on \( (Y,\mathcal{Y}) \), such that

\[ I(x,f) = \int f(y) \ P(x,dy) \text{ for all } f \in \mathcal{C}_c(Y). \tag{6} \]
Thus, there exists for any $x \in X$ a unique extended integral $\tilde{I}(x, \cdot)$ such that
\[
\tilde{I}(x, g) = \int g(y) \, P(x, dy) \quad \text{for all } g \in \mathcal{B}_b(Y).
\] (7)

Note that, by the dominated convergence theorem, $\tilde{I}(x, \cdot)$ is sequentially point-wise continuous on uniformly bounded sets. $\tilde{I}(x, \cdot)$ is also the unique extension of $I(x, \cdot)$ from $\mathcal{C}_c(Y)$ to $\mathcal{B}_b(Y)$ which is sequentially point-wise continuous on uniformly bounded sets, since $\{f \in \mathcal{C}_c(Y) \mid -1 \leq f \leq 1\}$ is point-wise sequentially dense in $\{g \in \mathcal{B}_b(Y) \mid -1 \leq g \leq 1\}$ (The fact that $\{f \in \mathcal{C}_c(Y) \mid -1 \leq f \leq 1\}$ is point-wise sequentially dense in $\{g \in \mathcal{B}_b(Y) \mid -1 \leq g \leq 1\}$ is proved completely analogous as we proved (c) in Lemma 1.)

**Step 2:** In the second step we show that the functions $x \mapsto \tilde{I}(x, g)$ are $\mathcal{X}$-measurable, for $g \in \mathcal{B}_b(Y)$ arbitrary, i.e., we show that $\tilde{I}$ is a conditional integral. Further we show that the unique corresponding function $P : X \times \mathcal{Y}$, 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
\[
[T(g)](x) = \tilde{I}(x, g).
\]

The fact that, for arbitrary $x \in X$, the integral $\tilde{I}(x, \cdot)$ is point-wise sequentially continuous on uniformly bounded sets of $\mathcal{B}_b(Y)$ (where we take $M = \mathcal{B}_b(Y)$ and $V = \mathbb{R}$ in Definition 3) implies that $T$ is also point-wise sequentially continuous (where we take $M = \mathcal{B}_b(Y)$ and $V = \mathbb{R}^X$ in Definition 3).

Further, $f \in \mathcal{C}_c(Y)$ implies by definition of $T$ and the fact that $I \in \mathcal{I}(X,Y)$ that
\[
T(f) = \left[ x \mapsto I(x, f) \right] \in \mathcal{B}(X),
\] (8)
i.e., we have that $T(\mathcal{C}_c(Y)) \subseteq \mathcal{B}(X)$.

By (8) 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 $g \in \mathcal{B}$ implies that $x \mapsto \tilde{I}(x, g)$ is $\mathcal{X}$-measurable. The fact that $x \mapsto \tilde{I}(x, g)$ 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 the Riesz representation theorem unique.

In the first two steps we have shown that to an integral $I \in \mathcal{I}(X,Y)$ there corresponds a unique kernel $P \in \mathcal{P}(X,Y)$ and a unique extended integral $\tilde{I}$. Further we know by equation (6) and (5) that this correspondence is given by $G^{-1}$. In the third step we show that to any $P \in \mathcal{P}(X,Y)$ there corresponds a unique $I = G(P) \in \mathcal{I}(X,Y)$.

7
Step 3: We show now that any kernel $P$ corresponds to an unique integral $I$. That any kernel $P$ gives us by formula (7) 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 $g = \sum_{i=1}^{n} \gamma_{i} 1_{B_{i}} \in B_{simp}$ we have:

$$\tilde{I}(x, g) = \sum_{i} \gamma_{i} P(x, B_{i}).$$

So for $g \in B_{simp}$ the function $x \mapsto \tilde{I}(x, g)$ is a finite sum of $\mathcal{X}$-measurable functions and thus itself $\mathcal{X}$-measurable. It remains to be shown that $x \mapsto \tilde{I}(x, g)$ is for any $g \in B(Y)$ 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, $f \in B_{simp}(Y)$ implies (by definition of $T$ and the fact that for $g \in B_{simp}(Y)$ the function $x \mapsto \tilde{I}(x, g)$ is $\mathcal{X}$-measurable) that:

$$T(f) = [x \mapsto \tilde{I}(x, f)] \in B(X),$$

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

By (9) together with Lemma 1 (b) and point-wise sequential continuity of $T$, we obtain that $T(B_{b}(Y)) = B(X)$. In other words, we obtain that $g \in B$ implies that $x \mapsto \tilde{I}(x, g)$ is $\mathcal{X}$-measurable.

Now we define weak differentiability of conditional integrals and kernels.

**Definition 4** Let $\Theta$ be an open interval in $\mathbb{R}$ and let $\vartheta \mapsto I_{\vartheta}$ be a path in (mapping from $\Theta$ 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 $\vartheta \mapsto I_{\vartheta}$ is weakly differentiable then we say that it is bounded weakly differentiable if

$$\sup_{f \in \mathcal{C}_{c}(Y)} \left| \frac{dI_{\vartheta}(x,f)}{d\vartheta} \right| < \infty,$$

for any $x \in X$.

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

\[
\int f(y)P'_\vartheta(x, dy) = \frac{d}{d\vartheta} \int f(y)P_\vartheta(x, dy).
\]

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

\[
I'_\vartheta(x, f) = \frac{d}{d\vartheta} I_\vartheta(x, f).
\]

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

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

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. Furthermore, \( I'(x, \cdot) \) is by the condition of boundedness in the definition of bounded weakly differentiable for any \( x\in X \) norm-bounded; i.e., \( I'(x, \cdot) \) is bounded on uniformly bounded subsets of \( \mathcal{C}_c(Y) \). 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 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:

\[
|P|(x, B) = \sup_{\substack{A\subseteq\mathcal{Y} \\text{ s.t. } \mathcal{X}_A \subseteq B}} 2 \cdot P(x, A) - P(x, B), \quad x \in X, \ B \in \mathcal{Y}.
\]

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

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

\[
x \mapsto |P|(x, B)
\]
is $\mathcal{X}$–measurable.

Let $\mathcal{A}$ be the set–field generated by a countable basis of the topology of $Y$. Then, $\mathcal{A}$ is countable and generates the $\sigma$–field $\mathcal{Y}$. For any set $B \in \mathcal{Y}$ and any measure $\mu$ on $(Y, \mathcal{Y})$ there exists a sequence $(A_n)_{n \in \mathbb{N}}$ of sets $A_n \in \mathcal{A}$ such that $\lim \mu(A_n \triangle B) = 0$ (see [6] Lemma A.24). Thus, the function

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

is the point–wise supremum over the countable family

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

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

**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, $P^+(x, B), P^-(x, B) \geq 0$ and $P^+(x, \cdot), P^-(x, \cdot)$ are measures, and $x \mapsto P^+(x, B)$ as well as $x \mapsto P^+(x, B)$ are $\mathcal{X}$- measurable functions on $X$. 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 $\theta$ the kernel $P_\theta$ is Markov. Then there exist for any $\vartheta$ Markov kernels $Q^+_{\vartheta}$ and $Q^-_{\vartheta}$ from $X$ to $Y$ and a $\mathcal{X}$-measurable function $c_\vartheta : X \rightarrow \mathbb{R}$ such that the weak derivative $P'_\vartheta$ of $P_\vartheta$ decomposes in the form

$$P_\vartheta(x, B) = c_\vartheta(x) \left( Q^+_{\vartheta}(x, B) - Q^-_{\vartheta}(x, B) \right) \quad \forall (x, B) \in X \times \mathcal{Y}.$$
Proof: By Theorem 1, the weak derivative $P_\vartheta'$ is for any $\vartheta$ a kernel and by the Corollary 1, $P_\vartheta'$ possesses a Jordan decomposition $(P_\vartheta^+, P_\vartheta^-)$, i.e., $P_\vartheta' = P_\vartheta^+ - P_\vartheta^-$ and $P_\vartheta^+, P_\vartheta^-$ are positive kernels. Since the $P_\vartheta$ are Markov kernels we have $P_\vartheta^+(x, Y) = P_\vartheta^-(x, Y)$. Let $c_\vartheta : X \to \mathbb{R}$ be defined by

$$c_\vartheta(x) := P_\vartheta^+(x, Y) = P_\vartheta^-(x, Y).$$

Since $P_\vartheta^+$ is a kernel, the function $c(\cdot)$ is $X$-measurable. Let

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

$$Q_\vartheta^-(x, B) := \frac{1}{c(x)}P_\vartheta^-(x, B) \text{ for all } x \text{ with } c(x) > 0$$

and let for an arbitrary fixed probability measure $\mu$, arbitrary $x$ with $c_\vartheta(x) = 0$ and arbitrary $B \in \mathcal{Y}$

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

Then $Q_\vartheta^+$ as well as $Q_\vartheta^-$ are Markov kernels. □

Remark 3 This specific decomposition $(c_\vartheta(\cdot), Q_\vartheta^+, Q_\vartheta^-)$ is only possible because the kernels $P_\vartheta'$ stem from weak differentiation of a Markov kernel valued function $\vartheta \mapsto P_\vartheta$.

References


