On the Differentiability of parametrized families of Linear Operators and the Sensitivity of their stationary vectors

Heinz Weisshaupt
ZBSA University of Freiburg
Habsburgerstrasse 49, 79104 Freiburg, Germany
heinz@weiszhaupt.at
heinz.weisshaupt@zbsa.uni-freiburg.de

October 21, 2010

Abstract

We investigate the differentiability of functions of stationary vectors associated with operator valued functions as well as the differentiability of the operator valued functions themselves. We display formulas connecting the derivatives of the parametric families of operators and vectors. The results are applied to the case of stochastic kernels.

MSC 2000: 47A56, 28A15, 47N30

Keywords: differentiation, operator-valued functions, stationary vectors, stationary distributions, differentiation of operators, differentiation of Markov kernels, measure-valued differentiation

Introduction

Weak derivatives of measures and Markov kernels are important concepts in the optimization of stochastic models [9]. There exist various versions of the fundamental theorem connecting the derivative of a parametric family of ergodic Markov Kernels and the derivative of the corresponding parametric family of stationary vectors. One version is Theorem 3.8 (of this paper) others are [9] Lemma 3.73 and [5] Theorem 1. The versions differ in the involved test-function spaces and in the specification of differentiability.

Further we know by [6] Theorem 1 (please notice Remark A.7 of this article) that we can relax the definition of weak differentiability of parametric families of Markov kernels such that the definition itself does not involve the existence of a derivative object, but that the relaxed definition still implies the existence of such a derivative object as a nontrivial consequence.

These facts call for a unifying theory of weak differentiation that serves as a general framework for the discrete time results obtained so far and as a starting point for the investigation of continuous time Markov processes. It is possible to build such a framework in a functional analytic setting:

> By the interplay of weak and strong structures on normed spaces we obtain a functional analytic theory for the differentiation of parametric families of operators and their stationary vectors. The application of our functional analytic results to Markov kernels and stationary distributions is straightforward and outlined in this article.

Organization of the paper:

The article is essentially self contained, i.e., the stated results and their proofs do not rely on other work in the field of weak differentiation. The paper is especially independent of the results stated in [5], [6], [7] and [9].

In section one we investigate weak differentiability and derivative objects in our general functional analytic setting. First we introduce in 1.1-1.7 the most fundamental concepts used in the paper. These are mainly various weak structures concerning topology, convergence and differentiability. The topological (and convergence) structures are introduced in a way similar to the concept of weak differentiability. This provides a unified approach to weak concepts that is especially helpful for the reader not that familiar with topological vector spaces and weak topologies. In Lemma 1.13 we consider weak convergence of sequences/nets of continuous operators on normed spaces. These convergence results are used to establish Theorem 1.14. The theorem states that the derivative object of a weakly differentiable parametric family of continuous operators is again a parametric family of continuous operators. This result is specialized in the next section to Theorem 2.9.

The concern of section two is the application of Theorem 1.14 to kernel operators introduced by Definition 2.2. The concept of a stochastic kernel[1] that serves us as a functional analytic counterpart of the concept of a Markov kernel is introduced. Theorem 2.9 on the differentiability of kernel operators is closely related to a corrected version (please notice Remark A.7 of this article) of [6] Theorem 1. The purpose of Theorem 2.9 is to identify the derivative objects in the setting of kernel operators, i.e., Theorem 2.9 says that the derivative of a parametric family of kernel operators is again a parametric family of kernel operators.

Section three connects the sensitivities of continuous linear operators with the sensitivities of their stationary vectors/distributions. The basic result is Theorem 3.1. Its proof is based on a homeomorphism theorem for compact

[1]We decided to use the concept of stochastic kernels instead of the concept of Markov kernels in our result since it fits better within the functional analytic framework.

sets, i.e., it is based on a topological result. Thus Theorem 3.6 - that is just a modification of Theorem 3.1 by Lemma 3.5 - possesses a topological core. This contrasts the "algebraic" proof of [5] Theorem 1. Theorem 3.6 of this paper and [5] Theorem 1 are quite different results. While Theorem 3.6 is applicable to general test-function spaces, the test-function space involved in [5] Theorem 1 has to contain certain infinite sums of (the absolute values of) test-functions. Further the Lipschitz continuity hypothesis on the Markov kernels in [5] Theorem 1 involves the order structure of the test-function space and is thus much stronger than the Lipschitz continuity hypothesis on the operators in Theorem 3.6 and Theorem 3.1 of our article. On the contrary the theorems 3.6 and 3.1 assume Lipschitz continuity of the parametric family of stationary vectors whereas [5] Theorem 1 works without such a hypothesis.

Section four indicates the applicability of our theory to the Gibbs sampler. To obtain the results of section four we apply a version of the product rule for the differentiation of linear operators provided by Lemma A.5 in the appendix.

By Remark 1.8, 1.15, 2.10, 3.4, 3.9 and A.6 there exist modifications of our main results that are obtained by application of the uniform boundedness principle A.4.

We suppose that all topologies under consideration are Hausdorff.

1 Weak differentiability and Derivative Objects

Weak differentiability of families of continuous operators and its consequences are the concern of this section.

Definition 1.1 Given metric spaces (Θ, d) and $(\tilde{\Theta}, \tilde{d})$, we say that $f : \Theta \to \tilde{\Theta}$ is Lipschitz continuous at a fixed $\theta_0 \in \Theta$ if there exists ℓ_{θ_0} such that $\tilde{d}(f(\theta_0), f(\theta)) \leq \ell_{\theta_0} \cdot d(\theta_0, \theta)$ for all $\theta \in \Theta$. We say that $f : \Theta \to \tilde{\Theta}$ is pointwise Lipschitz continuous if f is Lipschitz continuous at any $\theta_0 \in \Theta$, i.e., if for any $\theta_0 \in \Theta$ there exists ℓ_{θ_0} such that $\tilde{d}(f(\theta_0), f(\theta)) \leq \ell_{\theta_0} \cdot d(\theta_0, \theta)$ for all $\theta \in \Theta$. We say that ℓ_{θ_0} is a Lipschitz constant for f at θ_0.

Remark 1.2 A pointwise Lipschitz continuous function is in general not Lipschitz continuous. A counterexample is provided by the function $f : (0, 1] \mapsto \mathbb{R}$ defined by $f(x) = \sin(\frac{1}{x})$.

Remark 1.3 Any differentiable[2] function $f : \Theta \mapsto \mathbb{R}$ defined on some compact set $\Theta \subset \mathbb{R}$ is pointwise Lipschitz continuous.

[2] Let $\Theta \subseteq \mathbb{R}$. We say that a function $f : \Theta \to \mathbb{R}$ is differentiable at $\theta_0 \in \Theta$ with derivative $D_{\theta_0} f$ if θ_0 is an accumulation point of Θ and for any sequence $(\theta_n)_{n \in \mathbb{N}}$ with $\lim_{n \to \infty} \theta_n = \theta_0$ we have $\lim_{n \to \infty} \frac{f(\theta_n) - f(\theta_0)}{\theta_n - \theta_0} = D_{\theta_0} f \in \mathbb{R}$ depending on f and θ_0 but independent of the choice of the sequence $(\theta_n)_{n \in \mathbb{N}}$. We say that $f : \Theta \to \mathbb{R}$ is differentiable if f is differentiable at any $\theta \in \Theta$.

Notation 1.4 By the term vector space we always denote a real vector space. Given a set Ψ we denote by $\mathbb{R}^\Psi$ the space of functions from Ψ to $\mathbb{R}$ endowed with the topology of pointwise convergence (also called the product topology). Topological properties of sequences/nets in $\mathbb{R}^\Psi$ or subsets of $\mathbb{R}^\Psi$ are always considered with respect to this topology. The term operator always denotes a linear operator.

Definition 1.5 Suppose that we are given two vector spaces Ψ, W and a bilinear mapping $\langle . \mid . \rangle : \Psi \times W \to \mathbb{R}$. Let $\Xi \subset W$. We say that Ξ is Ψ-separated if $\xi \mapsto (\langle \psi, \xi \rangle)_{\psi \in \Psi}$ is injective and that Ξ is Ψ-compact if $\{(\langle \psi, \xi \rangle)_{\psi \in \Psi} \mid \xi \in \Xi\}$ is a compact subset of $\mathbb{R}^\Psi$ with respect to the product topology, i.e., if Ξ is compact with respect to the weak (also called initial or projective) topology induced by the mappings $\xi \mapsto \langle \psi \mid \xi \rangle$.[3]

Definition 1.6 Given normed spaces $(V, \|.\|_V)$ and $(W, \|.\|_W)$, we denote their closed unit balls by $\mathbb{B}_V$ respectively $\mathbb{B}_W$. We denote by $(\mathcal{L}(V,W)\|.\|_{\mathcal{L}})$ (or for short $\mathcal{L}(V,W)$, $(\mathcal{L}, \|.\|_{\mathcal{L}})$ or $\mathcal{L}$) the space of bounded (=continuous) linear operators from V to W endowed with the operator norm $\|L\|_{\mathcal{L}} := \sup_{v \in \mathbb{B}_V} \|Lv\|_W$. The closed unit ball with respect to $\|.\|_{\mathcal{L}}$ is denoted by $\mathbb{B}_{\mathcal{L}}$. Further we let $\mathcal{L}(W) := \mathcal{L}(W,W)$

Beside Ψ-properties like Ψ-compactness (introduced in Definition 1.5), we distinguish $\|.\|_{\mathcal{L}}$-, V-$\|.\|_W$- and Ψ-Γ-properties.[4]

For example we say that a function $\theta \mapsto L_\theta \in \mathcal{L}$ is pointwise $\|.\|_{\mathcal{L}}$-Lipschitz continuous if it is pointwise Lipschitz continuous from some metric space Θ to the space $\mathcal{L}$ endowed with the operator norm $\|.\|_{\mathcal{L}}$. We say that a function $\theta \mapsto L_\theta \in \mathcal{L}$ is pointwise V-$\|.\|_W$-Lipschitz continuous if for any $v \in V$ the function $\theta \mapsto L_\theta v$ is pointwise $\|.\|_W$-Lipschitz continuous.

Let a bilinear mapping $\langle . \mid . \rangle : \Psi \times W \to \mathbb{R}$ and a set $\Gamma \subseteq V$ be given. We say that a function $\theta \mapsto L_\theta \in \mathcal{L}(V,W)$ is Ψ-Γ-differentiable if for any $\psi \in \Psi$ and for any $\gamma \in \Gamma$ the mapping $\theta \mapsto \langle \psi, L_\theta \gamma \rangle \in \mathbb{R}$ is differentiable. We denote the derivative of a function $f : \Theta \to \mathbb{R}$ with respect to θ by $\frac{df}{d\theta}$ or $D_\theta f(\theta)$, while we denote the derivative object of an operator-valued function $\theta \mapsto K_\theta$ at some $\theta_0 \in \Theta$ by K'_{θ_0}.

Definition 1.7 [5] Let Ψ and Ξ be vector spaces and let $\langle . \mid . \rangle : \Psi \times \Xi \to \mathbb{R}$ be a bilinear mapping such that Ξ is Ψ-separated.

---

[3] For the definition and fundamental properties of product and weak topologies see [12] Section 8; for projective topologies in a topological vector space context see [10] Chapter II Section 5.

[4] $\|.\|_{\mathcal{L}}$-properties are often called uniform properties and $\|.\|_W$-properties are often called strong properties in the literature. Ψ respectively Ψ-Γ-properties are usually called weak or weak-$*$ properties.

[5] Of course all the Ψ- and Ψ-V-concepts in Definition 1.7 are just the usual topological concepts with respect to the respective initial topologies. We state them explicitly for the convenience of the reader not that familiar with topological concepts.

Let $(\xi_n)_{n\in\mathcal{N}}$ be a net[6] of vectors $\xi_n \in \Xi$ such that $(\lim_{n\in\mathcal{N}}(\langle\psi \mid \xi_n\rangle)_{\psi\in\Psi}$ exists in $\mathbb{R}^\Psi$ and let $\xi \in \Xi$ be such that $(\langle\psi \mid \xi\rangle)_{\psi\in\Psi} = (\lim_{n\in\mathcal{N}}(\langle\psi \mid \xi_n\rangle)_{\psi\in\Psi}$. Then we say that the net $(\xi_n)_{n\in\mathcal{N}}$ Ψ-converges to its Ψ-limit ξ and write $\xi = {}^\Psi\lim \xi_n$.

We say that $L \in \mathcal{L}(V,\Xi)$ is a Ψ-V-limit of the net $(L_n)_{n\in\mathcal{N}}$ if $(\langle\psi \mid Lv\rangle)_{(\psi,v)\in\Psi\times V}$ is a limit point of the net $((\langle\psi \mid L_n v\rangle)_{(\psi,v)\in\Psi\times V})_{n\in\mathcal{N}}$ in $\mathbb{R}^{\Psi\times V}$.

Further we say that $\mathcal{K} \subset \mathcal{L}(V,\Xi)$ is Ψ-V-compact if

$$\{(\langle\psi \mid Kv\rangle)_{(\psi,v)\in\Psi\times V} \mid K \in \mathcal{K}\} \ \text{ is a compact subset of } \ \mathbb{R}^{\Psi\times V}.$$

We let ${}^\Psi\sum_{n\in\mathbb{N}} \xi_n :={}^\Psi \lim_{n\in\mathbb{N}} \sum_{m=1}^n \xi_m$. Analogously we define ${}^\Psi\int$ as the Ψ-limit of an appropriate sequence/net of finite sums. (We will actually need the Ψ-integral ${}^\Psi\int$ of a Ξ-valued function only in Example 2.12. The integrand under consideration will be continuous for the Ψ-topology on Ξ and the integral can be simply taken in the sense of Riemann.)

Remark 1.8 Before we start our investigations we remark that Proposition A.1 can be used to provide modifications of some results proved in this paper. For example it can be used in the case that $(V, \|.\|_V)$ is a Banach space to relax the Lipschitz continuity assumption on $\theta \mapsto L_\theta$ in Theorem 3.1 or Theorem 3.6 from pointwise $\|.\|_\mathcal{L}$-Lipschitz continuity to pointwise V-$\|.\|_W$-Lipschitz continuity. The advantage of pointwise V-$\|.\|_W$-Lipschitz continuity over pointwise $\|.\|_\mathcal{L}$-Lipschitz continuity is that it might be more easily verified in some concrete settings. However both kinds of Lipschitz continuity are by Proposition A.1 in the case that $(V, \|.\|_V)$ is a Banach space equivalent. By Lemma A.3 and Proposition 1.3 further relaxations of the Lipschitz continuity assumptions are possible (and outlined in the remarks 1.15 and 1.16).

Remark 1.9 The following rather trivial Proposition (1.10)[7] is used in the proof of Lemma 1.13. This lemma is crucial for the existence of derivative

---

[6] A net is a function from a directed set $\mathcal{N}$ to some topological space. For details concerning directed sets, nets and net-convergence consult [12] section 11. Note that $\mathbb{N}$ is a directed set and that a sequence is a special kind of a net.

[7] Proof of Proposition 1.10: The case $\psi = 0$ is trivial. Therefore we suppose that $\psi \neq 0$. We further suppose without loss of generality (by continuity of $\langle.|.\rangle : \Psi \times W \to \mathbb{R}$, rescaling the norms if necessary) that for $\psi \in \Psi$ and $w \in W$

$$\langle\psi \mid w\rangle \leq \|\psi\|_\Psi \cdot \|w\|_W.$$

To prove (3) we have thus to show that $\forall\varepsilon > 0$, $\forall\psi \in \Psi \setminus \{0\}$, $\forall v \in V$ there exists a neighborhood U_0 of $0 \in \widetilde{\Theta}$ such that

$$\forall\theta \in U_0 \quad \left|\langle\psi \mid [L_\theta - L_0]v\rangle\right| < \varepsilon$$

Thus let $\varepsilon > 0$, $\psi \in \Psi \setminus \{0\}$ and $v \in V$ be given. Choose $\gamma \in \Gamma$ such that $\|v - \gamma\| < \frac{\varepsilon}{3\|\psi\|}$ and let U_0 be such that

$$\forall\theta \in U_0 \quad \left|\langle\psi \mid [L_\theta - L_0]\gamma\rangle\right| < \frac{\varepsilon}{3}.$$

Then

$$\left|\langle\psi \mid [L_\theta - L_0]v\rangle\right| \leq$$
$$\left|\langle\psi \mid L_\theta[v-\gamma]\rangle\right| + \left|\langle\psi \mid [L_\theta - L_0]\gamma\rangle\right| + \left|\langle\psi \mid L_0[\gamma - v]\rangle\right| \leq$$

objects $K'_\theta \in \mathcal{L}(V,W)$ if we are given that $\theta \mapsto \langle\psi \mid K_\theta\,\gamma\rangle$ is differentiable for $\gamma \in \Gamma$ with Γ a $\|.\|_V$-dense subset of V. (Compare with Theorem 1.14, (7), (8) and (9)).

Proposition 1.10 *Let $(\Psi, \|.\|_\psi)$, $(W, \|.\|_W)$ and $(V, \|.\|_V)$ be normed spaces. Suppose that W is Ψ-separated via the $\|.\|_\Psi$-$\|.\|_W$-continuous bilinear mapping $\langle.|.\rangle : \Psi \times W \to \mathbb{R}$. Let Γ be a $\|.\|_V$-dense subset of $(V, \|.\|_V)$. Let $\widetilde{\Theta}$ be a topological space and let $0 \in \widetilde{\Theta}$ denote some distinguished point. Let $(L_\theta)_{\theta\in\widetilde{\Theta}} \in (\mathcal{L}(V,W))^{\widetilde{\Theta}}$ be such that:*

$$(\forall\theta \in \widetilde{\Theta})\ (L_\theta \in \mathbb{B}_{\mathcal{L}},\ \textit{i.e.},\ \|L_\theta\| \le 1), \tag{1}$$

$$(\forall\gamma \in \Gamma)(\forall\psi \in \Psi)\ \left(\lim_{\theta\to 0}\langle\psi \mid L_\theta\gamma\rangle = \langle\psi \mid L_0\gamma\rangle\right). \tag{2}$$

Then

$$(\forall v \in V)(\forall\psi \in \Psi)\ \left(\lim_{\theta\to 0}\langle\psi \mid L_\theta v\rangle = \langle\psi \mid L_0 v\rangle\right). \tag{3}$$

Remark 1.11 In Proposition 1.10 the bilinear mapping $\langle.\mid.\rangle : \Psi \times W \to \mathbb{R}$ is $\|.\|_\Psi$-$\|.\|_W$-continuous and W is Ψ separated. Thus we can identify W algebraically with a subspace of the topological dual[8] Ψ^* of Ψ via $\psi^*_w(\psi) := \langle\psi \mid w\rangle$.

The propositions 1.10 and 1.12, the lemmas 1.13 and A.5 as well as the theorems 1.14, 3.1 and 3.6 can be modified in accordance with this observation.[9]

But note that in general $(W, \|.\|_W)$ is not topologically equivalent to a subspace of the Banach-Dual $(\Psi^*, \|.\|_{\Psi*})$ of $(\Psi, \|.\|_\Psi)$.

Proposition 1.12 *Let $(\Psi, \|.\|_\psi)$, $(W, \|.\|_W)$ and $(V, \|.\|_V)$ be normed spaces. Suppose that W is Ψ-separated via the $\|.\|_\Psi$-$\|.\|_W$-continuous bilinear mapping*

$$\|\psi\|\cdot\|L_\theta\|\cdot\|v-\gamma\| + \left|\langle\psi \mid [L_\theta - L_0]\gamma\rangle\right| + \|\psi\|\cdot\|L_0\|\cdot\|v-\gamma\| <$$
$$\|\psi\|\cdot 1\cdot\frac{\varepsilon}{3\|\psi\|} + \frac{\varepsilon}{3} + \|\psi\|\cdot 1\cdot\frac{\varepsilon}{3\|\psi\|} = \varepsilon.\ \square$$

[8] By the term "topological dual of Ψ" we denote the vector space of continuous linear functionals on $(\Psi, \|.\|_\Psi)$.

[9] One has simply to replace the sentence:

"Suppose that W is Ψ-separated via the $\|.\|_\Psi$-$\|.\|_W$-continuous bilinear mapping $\langle.|.\rangle : \Psi \times W \to \mathbb{R}$."

in the propositions, lemmas and theorems by

"Suppose that W is (algebraically equivalent to) a subspace of the topological dual of Ψ and that $\langle.\mid.\rangle : \Psi \times W \to \mathbb{R}$ defined by $\langle\psi \mid w\rangle := w(\psi)$ is $\|.\|_\Psi$-$\|.\|_W$-continuous.",

respectively one has to replace the corresponding sentences that contain V, v in place of W, w.

$\langle .|. \rangle : \Psi \times W \to \mathbb{R}$. Suppose further that the unit ball $\mathbb{B}_W$ is Ψ-compact. Then $\mathbb{B}_{\mathcal{L}}$ is Ψ-V-compact.[10]

Lemma 1.13 *Let $(\Psi, \|.\|_\psi)$, $(W, \|.\|_W)$ and $(V, \|.\|_V)$ be normed spaces. Suppose that W is Ψ-separated via the $\|.\|_\Psi$-$\|.\|_W$-continuous bilinear mapping $\langle .|. \rangle : \Psi \times W \to \mathbb{R}$. Let Γ be a $\|.\|_V$-dense subset of $(V, \|.\|_V)$. Let $\widetilde{\Theta}$ be a topological space and let $0 \in \widetilde{\Theta}$ denote some distinguished point. Suppose that the unit ball $\mathbb{B}_W$ is Ψ-compact. Let the mapping $(\theta \mapsto L_\theta) \in (\mathcal{L}(V, W))^{\widetilde{\Theta} \setminus \{0\}}$ be such that*

$$(\forall \theta \in \widetilde{\Theta} \setminus \{0\}) \ (L_\theta \in \mathbb{B}_{\mathcal{L}}) \tag{4}$$

and that

$$(\forall \psi \in \Psi)(\forall \gamma \in \Gamma) \quad \lim_{h \to 0} \langle \psi \mid L_h \gamma \rangle \quad \text{exists.} \tag{5}$$

Then there exists a unique operator $L_0 \in \mathcal{L}(V, W)$ such that (3) holds.

Proof: By Proposition 1.12 the unit ball $\mathbb{B}_{\mathcal{L}}$ is Ψ-V-compact. Thus we conclude from (4) the existence of an operator L_0 such that[11]

$$\begin{aligned} L_0 \in \mathbb{B}_{\mathcal{L}} \text{ is a } \Psi\text{-}V\text{-limit of some ultranet } (L_{\theta_n})_{n \in \mathcal{N}} \\ \text{for that } \theta_n \to 0. \end{aligned} \tag{6}$$

Further (4) and (6) together imply (1) while (6) and (5) together imply (2). Application of Proposition 1.10 thus proves (3). Since W is Ψ-separated the operator L_0 is unique and the lemma has been proved. □

Theorem 1.14 *Let $(\Psi, \|.\|_\Psi)$, $(V, \|.\|_V)$ and $(W, \|.\|_W)$ be normed spaces. Suppose that W is Ψ-separated via the $\|.\|_\Psi$-$\|.\|_W$-continuous bilinear mapping $\langle .|. \rangle : \Psi \times W \to \mathbb{R}$. Let $\Gamma \subseteq V$ be a $\|.\|_V$-dense subset of V and let $\mathbb{B}_W$ be Ψ-compact.*

Let $(K_\theta)_{\theta \in \Theta \subseteq \mathbb{R}}$ be a parametric family of operators $K_\theta \in \mathcal{L}(V, W)$. Suppose that:

$$\theta \mapsto K_\theta \text{ is } \Psi\text{-}\Gamma\text{-differentiable and} \tag{7}$$

(A) $\theta \mapsto K_\theta$ *is pointwise $\|.\|_{\mathcal{L}}$-Lipschitz continuous.*

[10] Proof of Proposition 1.12: The proof of Proposition 1.12 proceeds along the lines of the proof of the Theorem of Alaoglu Bourbaki (see [2] 5.7.5 or [10] Chapter III, Section 4.3 Corollary): We note that Ψ separates W and thus all topologies under consideration are Hausdorff and the limits of ultranets are unique. By the product theorem of Tychonov $\mathcal{K} := \prod_{v \in V} \|v\| \cdot \mathbb{B}_W$ is compact. Further $\mathbb{B}_{\mathcal{L}} \subseteq \mathcal{K}$ and the topology under consideration on $\mathbb{B}_{\mathcal{L}}$ is the topology that $\mathbb{B}_{\mathcal{L}}$ inherits from $\mathcal{K}$ as a subspace of $\mathcal{K}$. It thus remains to be proved that $\mathbb{B}_{\mathcal{L}}$ is a closed subspace of $\mathcal{K}$. Let thus $(L_n)_{n \in \mathcal{N}}$ be an ultranet in $\mathbb{B}_{\mathcal{L}}$. Then by compactness of the Hausdorff space $\mathcal{K}$ there exists a unique limit L_0 of $(L_n)_{n \in \mathcal{N}}$ in $\mathcal{K}$. To prove that $L_0 \in \mathbb{B}_{\mathcal{L}}$ it suffices to show that L_0 is linear from V to W. But this is easily seen since linearity is preserved by limits with respect to Hausdorff vector space topologies. □

[11] By [12] Theorem 17.4 a space is compact if and only if every ultranet converges.

Then

$$(\forall v \in V)(\forall \psi \in \Psi)\ (D_\theta \langle \psi \mid K_\theta v \rangle \ \ \textit{exists}), \qquad i.e.,\ \theta \mapsto K_\theta \ \textit{is}\ \Psi\text{-}V\textit{-differentiable} \tag{8}$$

and there exists an operator $K'_\theta \in \mathcal{L}(V, W)$ *such that*

$$(\forall v \in V)(\forall \psi \in \Psi)\ (D_\theta \langle \psi \mid K_\theta v \rangle = \langle \psi \mid K'_\theta v \rangle), \qquad i.e.,\ \theta \mapsto K'_\theta \ \textit{is the}\ \Psi\text{-}V\textit{-derivative of}\ \theta \mapsto K_\theta. \tag{9}$$

(B) *Let* $\Omega \subseteq V$ *and let* $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ *be a* σ*-algebra. Suppose that for all* $\psi \in \Psi$ *and all* $\theta \in \Theta$ *the mappings*

$$\omega \mapsto \langle \psi \mid K_\theta \omega \rangle \ \ \textit{from}\ \ \Omega \ \ \textit{to}\ \ \mathbb{R}$$

are $\mathcal{A}$*-measurable. Then for all* $\psi \in \Psi$ *and all* $\theta \in \Theta$ *the mappings*

$$\omega \mapsto \langle \psi \mid K'_\theta \omega \rangle \ \textit{from}\ \Omega \ \textit{to}\ \mathbb{R} \ \textit{are also}\ \mathcal{A}\textit{-measurable.}$$

(C) *Suppose that we are given a constant* $c \in \mathbb{R}$ *and a* Ψ*-*V*-continuous linear functional*

$$F : \mathcal{L}(V, W) \to \mathbb{R} \ \ \textit{such that}\ \ (\forall \theta \in \Theta)\ (F(K_\theta) = c).$$

Then

$$F(K'_\theta) = 0.$$

Remark 1.15 If $(V, \|.\|_V)$ is a Banach space then - according to Remark 1.8 - hypothesis (A) of Theorem 1.14 can be relaxed to

(A') $\theta \mapsto K_\theta$ is pointwise V-$\|.\|_W$-Lipschitz continuous.

If in addition Ψ is a Banach space with respect to the norm $\|.\|_\Psi$ and

$$\|.\|_W \leq \sup_{\psi \in \mathbb{B}_\Psi} \langle \psi | w \rangle, \tag{10}$$

then hypothesis (A) of Theorem 1.14 can - by Lemma A.3 - be further relaxed to

(A") $\theta \mapsto K_\theta$ is pointwise Ψ-V-Lipschitz continuous.

Remark 1.16 If we suppose that $\Gamma = V$ and that Θ is (locally) compact, then the differentiability hypothesis (7) on $\theta \mapsto K_\theta$ implies by Remark 1.3 that (A") holds (in some compact neighborhood of θ_0). Thus we obtain, in the case that

$(V, \|.\|_V)$, $(\Psi, \|.\|_\Psi)$ are Banach spaces such that (10) is fulfilled, $\Gamma = V$ and Θ is locally compact, that Theorem 1.14 holds true without any Lipschitz condition on $\theta \mapsto K_\theta$.

Proof of Theorem 1.14: The mapping $\theta \mapsto K_\theta$ is by hypothesis (A) pointwise $\|.\|_{\mathcal{L}}$-Lipschitz continuous. Let $\theta \in \Theta$ be a fixed but arbitrarily chosen point and suppose without loss of generality that the Lipschitz constant of $\theta \mapsto K_\theta$ at θ equals 1. Let $\widetilde{\Theta} = \Theta - \theta := \{\vartheta - \theta \mid \vartheta \in \Theta\}$. For $0 \neq h \in \widetilde{\Theta}$ let $L_h := \frac{K_{\theta+h} - K_\theta}{h}$. Then $L_h \in \mathbb{B}_{\mathcal{L}}$ and, by hypothesis (7), for any $\psi \in \Psi$ and $\gamma \in \Gamma$ the limit $\lim_{h\to 0}\langle \psi \mid L_h \gamma\rangle$ exists. By an application of Lemma 1.13 we obtain that there exists a unique $L_0 \in \mathbb{B}_{\mathcal{L}}$ such that (3) holds, i.e.,

$$(\forall v \in V)(\forall \psi \in \Psi)\ \left(\lim_{\theta\to 0}\langle \psi \mid L_\theta v\rangle = \langle \psi \mid L_0 v\rangle\right). \tag{11}$$

If we let $K'_\theta = L_0$ then equation (11) becomes (9) and thus further (8) has been established.

The $\mathcal{A}$-measurability of

$$\omega \mapsto \langle \psi \mid K'_\theta \omega\rangle = \langle \psi \mid L_0 \omega\rangle$$

follows from the $\mathcal{A}$-measurability (applying hypothesis (B)) of

$$\omega \mapsto \left\langle \psi \middle| \frac{K_{\theta+\frac{1}{n}} - K_\theta}{\frac{1}{n}} \omega \right\rangle = \left\langle \psi \middle| L_{\frac{1}{n}} \omega \right\rangle$$

for $n \in \mathbb{N}$ and the fact that a pointwise sequential limit of $\mathcal{A}$-measurable functions is again $\mathcal{A}$-measurable.

That $F(K'_\theta) = 0$ follows from hypothesis (C), since by (C) and (11)

$$0 = \lim_{h\to 0} \frac{F(K_{\theta+h}) - F(K_\theta)}{h} = \lim_{h\to 0} F(L_h) = F({}^{\Psi\text{-}V}\lim_{h\to 0} L_h) = F(L_0) = F(K'_\theta).$$

□

2 Differentiation of kernel operators

This section is concerned with the differentiation of parametric families of kernel operators and the existence and structure of their derivative objects.

Definition 2.1 Let $(\mathcal{X}, \tau)$ be a locally compact second countable topological space. We denote by $(\mathcal{C}_c(\mathcal{X}), \|.\|_\infty)$ the space of continuous real valued functions with compact support on $\mathcal{X}$ endowed with the norm of uniform convergence. By $(C_0(\mathcal{X}), \|.\|_\infty)$ we denote the completion of $\mathcal{C}_c$ with respect to the $\|.\|_\infty$-norm, i.e., $\mathcal{C}_0$ denotes the space of functions that vanish at infinity. In the case that $\mathcal{X}$ is compact the spaces $\mathcal{C}_c(\mathcal{X})$ and $\mathcal{C}_0(\mathcal{X})$ coincide with the space $\mathcal{C}(\mathcal{X})$ of all continuous real valued functions on $\mathcal{X}$. Let $\mathcal{B}$ be the family of Baire-measurable subsets of $\mathcal{X}$, i.e., the σ-algebra generated by the compact sets. We denote by $(\mathcal{M}(\mathcal{X}), \|.\|_{Var})$ the space of signed σ-additive finite measures on $(\mathcal{X}, \mathcal{B})$ endowed with the total variation norm. We note that by the representation theorem of Riesz (see [2] 5.2.9) $\mathcal{M}(\mathcal{X}) = (\mathcal{C}_c(\mathcal{X}), \|.\|_\infty)^*$[12] Further $\mathcal{M}(\mathcal{X}) = (\mathcal{D}, \|.\|_\infty)^*$ for any $\|.\|_\infty$-dense linear subspace $\mathcal{D}$ of $\mathcal{C}_0$. Thus by the Theorem of Alaoglu Bourbaki (see [2] 5.7.5 or [10] Chapter III, Section 4.3 Corollary) the closed unit ball $\mathbb{B}_{Var}$ of $\mathcal{M}(\mathcal{X})$ is $\mathcal{D}$-compact. Further we let $\mathcal{M}_0(\mathcal{X}) := \{\mu \in \mathcal{M}(\mathcal{X}) \mid \mu(\mathcal{X}) = 0\}$. We denote by δ_x the Dirac measure at x given by $\delta_x(A) := \mathbb{1}_A(x)$.

Definition 2.2 We say that $K \in \mathcal{L}((\mathcal{M}(\mathcal{X}), \|.\|_{\mathcal{L}})$ is a kernel operator on $\mathcal{X}$ if

$$(\forall B \in \mathcal{B}) \quad (x \mapsto [K\delta_x](B)) \text{ is } \mathcal{B}\text{-measurable} \tag{12}$$

and

$$(\forall B \in \mathcal{B}) \quad (\forall \mu \in \mathcal{M}(\mathcal{X})) \quad [K\mu](B) = \int [K\delta_x](B) d\mu(x). \tag{13}$$

Remark 2.3 We say that a kernel operator is a stochastic kernel if it is positive (i.e., for μ non-negative, we also have $K\mu$ non-negative), and for all $\mu \in \mathcal{M}(\mathcal{X})$ we have $\mu(\mathcal{X}) = [K\mu](\mathcal{X})$. Note that there is a one-to-one correspondence between the space of stochastic kernels and the space of Markov-kernels, i.e., any stochastic kernel can be interpreted as a Markov kernel and vice versa.[13]

Lemma 2.4 *A second countable locally compact space $\mathcal{X}$ is a polish (completely metrizable and separable) space. Further a second countable locally compact space is σ-compact , i.e., it is the union of a countable family of compact sets.*[14]
□

Remark 2.5 Following Lemma 2.4, we endow $\mathcal{X}$ with a metric d such that $(\mathcal{X}, d)$ becomes a polish space whenever this seems convenient to us.

Definition 2.6 Let $\mathcal{D}$ be a family of $\|.\|_\infty$-bounded real valued functions over some measurable space $(\mathcal{X}, \mathcal{B})$. We denote by $\overline{\mathcal{D}}^{\text{seq}}$ the closure of $\mathcal{D}$ with respect to sequential pointwise convergence of $\|.\|_\infty$-bounded sequences.

Lemma 2.7 *Suppose that $\mathcal{D}$ is a ring of $\|.\|_\infty$-bounded real valued functions over some measurable space $(\mathcal{X}, \mathcal{B})$. Suppose that $\mathcal{G}$ generates the σ-algebra $\mathcal{B} \subseteq \mathcal{P}(\mathcal{X})$ and that $\{\mathbb{1}_{\mathcal{X}}\} \cup \{\mathbb{1}_G \mid G \in \mathcal{G}\} \subseteq \overline{\mathcal{D}}^{seq}$. Then $\{\mathbb{1}_B \mid B \in \mathcal{B}\} \subseteq \overline{\mathcal{D}}^{seq}$*

---

[12] $\mathcal{M}(\mathcal{X}) = (\mathcal{C}_c(\mathcal{X}), \|.\|_\infty)^*$ holds since locally compact second countable spaces are σ-compact. A locally compact (non σ-compact) counterexample is provided by any uncountable set $\mathcal{X}$ endowed with the discrete topology. In the case of such a $\mathcal{X}$ the space $\mathcal{M}(\mathcal{X})$ contains the measure μ that is 0 on countable sets and 1 on complements of countable sets. But the action of μ on $(\mathcal{C}_c(\mathcal{X}), \|.\|_\infty)$ is indistinguishable from the action of the 0-functional, i.e., the measure that equals 0 on all sets (countable and co-countable).

[13] Let K be a stochastic kernel, then a Markov kernel $\widetilde{K}$ is given by $(x, B) \mapsto \widetilde{K}(x, B) := [K\delta_x](B)$. If $\widetilde{K}$ is a Markov kernel, then $\mu \mapsto [K\mu](.) := \int \widetilde{K}(x, .)\, d\mu(x)$ is a stochastic kernel.

[14] See: [3] (Bourbaki, General Topology II), Chapter IX, §2, Section 9, the Corollary following Proposition 16 and further §6, Section 1, the Corollary following Proposition 2.

Proof: Since $\mathcal{D}$ is a ring $\overline{\mathcal{D}}^{seq}$ is again a ring. Note that the operation of complementation $B \mapsto \mathcal{X} \setminus B$ in the σ-algebra $\mathcal{B}$ corresponds to $\mathbb{1}_B \mapsto \mathbb{1}_{\mathcal{X}} - \mathbb{1}_B$ in the ring $\overline{\mathcal{D}}^{\mathrm{seq}}$ while the formation of countable intersections $(B_n)_{n\in\mathbb{N}} \mapsto \bigcap_{n\in\mathbb{N}} B_n$ corresponds to the formation of the limit of products $(\mathbb{1}_{B_n})_{n\in\mathbb{N}} \mapsto \lim_{n\in\mathbb{N}} \prod_{i=0}^{n} \mathbb{1}_{B_i}$ in the ring $\overline{\mathcal{D}}^{\mathrm{seq}}$. □

Lemma 2.8 *Let $\mathcal{D}$ be a $\|.\|_\infty$-dense subspace of $\mathcal{C}_0(\mathcal{X})$ for some second countable locally compact space $\mathcal{X}$ and let $\mathcal{B}$ denote the σ-algebra of Baire measurable subsets of $\mathcal{X}$. Then hypothesis (12) is equivalent to*

$$(\forall \psi \in \mathcal{D}) \quad x \mapsto \int \psi \; d[K\delta_x] \text{ is } \mathcal{B}\text{-measurable} \tag{14}$$

and hypothesis (13) is equivalent to

$$(\forall \psi \in \mathcal{D}) \quad (\forall \mu \in \mathcal{M}(\mathcal{X})) \quad \int \psi \; d[K\mu] = \int d\mu(x) \int \psi \; d[K\delta_x]. \tag{15}$$

Proof: That (12) implies (14) and that (13) implies (15) follows by an application of the dominated convergence Theorem since any $\psi \in \mathcal{C}_0$ is $\mathcal{B}$-measurable and $\|.\|_\infty$-bounded.

To prove the converse of both implications it suffices by the fact that $\mathcal{D}$ is a $\|.\|_\infty$-dense subspace of $\mathcal{C}_0(\mathcal{X})$ and the dominated convergence theorem to prove that for $B \in \mathcal{B}$ we have that $\mathbb{1}_B$ is an element of $\overline{\mathcal{C}_0}^{seq}$. Since $\mathcal{C}_0$ is a ring and the space $\mathcal{G}$ of compact subsets of $\mathcal{X}$ generates the σ-algebra $\mathcal{B}$ it suffices further by Lemma 2.7 to show that for any $G \in \mathcal{G}$ we have that $\mathbb{1}_G \in \overline{\mathcal{C}_0}^{seq}$ and $\mathbb{1}_{\mathcal{X}} \in \overline{\mathcal{C}_0}^{seq}$.

Let thus G be a compact subset of $\mathcal{X}$ and let $A_n := \{x \in \mathcal{X} \mid d(x, G) \geq \frac{1}{n}\}$. By Urysohn's Lemma ([12] 15.6) there exist continuous functions $\psi_n : \mathcal{X} \to [0, 1]$ such that $\psi_n \mid_{A_n} = 0$ and $\psi_n \mid_G = 1$. We thus obtained a $\|.\|_\infty$-bounded sequence $(\psi_n)_{n\in\mathbb{N}}$ of functions $\psi_n \in \mathcal{C}_0$ taking values in $[0, 1]$ such that

$$\lim_{n\in\mathbb{N}} \psi_n(x) = \mathbb{1}_G(x) \text{ and thus } \mathbb{1}_G \in \overline{\mathcal{C}_0}^{seq}.$$

Further there exists by σ-compactness of $\mathcal{X}$ (see Lemma 2.4) an increasing sequence $(G_n)_{n\in\mathbb{N}}$ of compact sets such that $\bigcup_{n\in\mathbb{N}} G_n = \mathcal{X}$. Thus

$$\mathbb{1}_{\mathcal{X}} = \lim_{n\in\mathbb{N}} \mathbb{1}_{G_n} \in \overline{\mathcal{C}_0}^{seq}. \ \square$$

Theorem 2.9 *Let $(K_\theta)_{\theta\in\Theta\subseteq\mathbb{R}}$ be a parametric family of kernel operators on a locally compact, second countable space $(\mathcal{X}, d)$. Let $\mathcal{D}$ be a $\|.\|_\infty$-dense subset of $\mathcal{C}_0$ and let Γ be a $\|.\|_{Var}$-dense subset of $\mathcal{M}(\mathcal{X})$. Suppose that $\theta \mapsto K_\theta$ is pointwise $\mathcal{M}(\mathcal{X})$-$\|.\|_{Var}$-Lipschitz continuous and $\mathcal{D}$-Γ-differentiable. Then there exist kernel operators $K'_\theta \in \mathcal{L}(\mathcal{M}(\mathcal{X}))$ such that $\int \psi \; d[K'_\theta\mu] = D_\theta \int \psi \; d[K_\theta\mu]$ for $\psi \in \mathcal{C}_0$ and $\mu \in \mathcal{M}(\mathcal{X})$.*

Remark 2.10 If $\mathcal{D} := \mathcal{C}_0$ then the Lipschitz continuity hypothesis can be relaxed to pointwise $\mathcal{C}_0$-$\mathcal{M}(\mathcal{X})$-Lipschitz continuity. This follows from Remark 1.15 and the fact that $(\mathcal{C}_0, \|.\|_\infty)$ is a Banach space and $(\mathcal{M}(\mathcal{X}), \|.\|_{Var})$ is its Banach dual. If in addition $\Gamma = \mathcal{M}(\mathcal{X})$ and Θ is locally compact, then by Remark 1.16 the Theorem 2.9 holds without any Lipschitz continuity assumption on $\theta \mapsto K_\theta$.

Proof: Taking Remark 1.15 into account the application of Theorem 1.14 with $(V, \|.\|_V) := (W, \|.\|_W) := (\mathcal{M}(\mathcal{X}), \|.\|_{Var})$, $\Psi := \mathcal{D}$ and $\langle \psi \mid \mu \rangle := \int \psi \, d\mu$ gives that there exist operators $K'_\theta \in \mathcal{L}(\mathcal{M}(\mathcal{X}))$ such that

$$(\forall \mu \in \mathcal{M}(\mathcal{X})) \quad (\forall \psi \in \mathcal{D}) \quad \left(D_\theta \int \psi \; d[K_\theta \mu] = \int \psi \; d[K'_\theta \mu] \right).$$

Thus to prove the lemma, we just have to show that the operators K'_θ are kernel operators[15], i.e., we have - by Lemma 2.8 - to show that (14) and (15) are fulfilled for $K = K'_\theta$.

(14) holds for $K = K_\theta$ and $K = K_{\theta+h}$, with $\theta, \theta + h \in \Theta$ arbitrary. Since the pointwise limit of a sequence of $\mathcal{B}$-measurable functions is again $\mathcal{B}$-measurable we obtain that

$$(\forall \psi \in \mathcal{D}) \quad \left(x \mapsto \int \psi \; d[K' \delta_x] \right) \equiv \left(x \mapsto \lim_{h \to 0} \frac{\int \psi \; d[K_{\theta+h} \delta_x] - \int \psi \; d[K_\theta \delta_x]}{h} \right)$$

is $\mathcal{B}$-measurable, i.e., (14) holds for $K = K'_\theta$.[16]

To prove (15) for $K = K'_\theta$ we let

$$C_n := \left\{ x \;\middle|\; \sup_{h \in \left[-\frac{1}{n}, 0\right) \cup \left(0, \frac{1}{n}\right]} \left| \frac{\int \psi \; d[K_{\theta+h} \delta_x] - \int \psi \; d[K_\theta \delta_x]}{h} \right| \leq n \right\}.$$

Note that $C_n \subseteq C_{n+1}$ and that $\bigcup_{n \in \mathbb{N}} C_n = \mathcal{X}$.

Letting $\mu_n(B) := \mu(B \cap C_n)$ for all $B \in \mathcal{B}$, we obtain by dominated convergence (since by our hypothesis on C_n the functions $\left| \frac{\int \psi \; d[K_{\theta+h} \delta_x] - \int \psi \; d[K_\theta \delta_x]}{h} \right|$

[15] The alternative would be that the K'_θ are more general operators that could be written in the form $K'_\theta = \kappa_{\theta,1} K_{\theta,1} - \kappa_{\theta,2} K_{\theta,2}$ with $K_{\theta,1}$, $K_{\theta,2}$ generalized stochastic operators as used in LeCam Theory and $\kappa_{\theta,1}, \kappa_{\theta,2} \geq 0$. (For the definition of these generalized stochastic operators See [4] and [11].)

[16] Alternatively we could have used Theorem 1.14 (B) to prove that (14) holds for $K = K'_\theta$.

are eventually, i.e., for small h, uniformly bounded by n on C_n) that

$$\underline{\int \psi \, d[K'_\theta \mu_n]} \;=\; \lim_{h\to 0} \frac{\int \psi \, d[K_{\theta+h}\mu_n] - \int \psi \, d[K_\theta \mu_n]}{h} \;=$$

$$= \lim_{h\to 0} \int d\mu_n(x) \, \frac{\int \psi \, d[K_{\theta+h}\delta_x] - \int \psi \, d[K_\theta \delta_x]}{h} \;=\; \underline{\int d\mu_n(x) \int \psi \, d[K'_\theta \delta_x]} \tag{16}$$

Taking into account that $K'_\theta \in \mathcal{L}(\mathcal{M}(\mathcal{X}))$, that μ_n converges with respect to $\|.\|_{Var}$ to μ and that $\left|\int \psi \, d[K'_\theta \delta_x]\right| \le \|\psi\|_\infty \cdot \|K'_\theta\|_{\mathcal{L}}$ is bounded, we obtain from (16) by letting $n \to \infty$ that

$$\underline{\int \psi \, d[K'_\theta \mu]} \;=\; \lim_{n\to\infty} \int \psi \, d[K'_\theta \mu_n]$$

$$= \lim_{n\to\infty} \int d\mu_n(x) \int \psi \, d[K'_\theta \delta_x] \;=\; \underline{\int d\mu(x) \int \psi \, d[K'_\theta \delta_x]} \,,$$

i.e., we obtain that (15) holds for $K = K'_\theta$. □

Remark 2.11 Even if the operators K_θ in Theorem 2.9 are stochastic kernels the kernel operators K'_θ are not necessarily elements of $\mathcal{L}(\mathcal{M}(\mathcal{X}), \mathcal{M}_0(\mathcal{X}))$ as example 2.12 below (that concerns the even more simple situation of the differentiation of pointwise $\|.\|_{Var}$-Lipschitz continuous probability-measure valued functions) shows. However, if $\mathcal{X}$ is compact then $\mathbb{1}_{\mathcal{X}} \in \mathcal{C}_0(\mathcal{X})$ and thus for stochastic K_θ and all $\mu \in \mathcal{M}(\mathcal{X})$

$$\int \mathbb{1}_{\mathcal{X}} \, d[K'_\theta \mu] = \lim_{h\to 0} \frac{\int \mathbb{1}_{\mathcal{X}} \, d[K_{\theta+h}\mu] - \int \mathbb{1}_{\mathcal{X}} \, d[K_\theta \mu]}{h} = \lim_{h\to 0} \frac{\mu(\mathcal{X}) - \mu(\mathcal{X})}{h} = 0,$$

i.e., $K'_\theta \in \mathcal{L}(\mathcal{M}(\mathcal{X}), \mathcal{M}_0(\mathcal{X}))$ holds.

Example 2.12 Define for $\theta \in \left[0, \frac{1}{2}\right)$ probability measures μ_θ on $[0, \infty)$ by

$$\mu_\theta := \frac{1}{1-\theta} \cdot \int_{s\in(\theta,1)} \mathbb{U}_{\left[0, \frac{1}{s}\right]} \, ds,$$

with $\mathbb{U}_{\left[0, \frac{1}{s}\right]}$ the uniform distribution on $\left[0, \frac{1}{s}\right]$. Then

$$\mu'_0 := {}^{\mathcal{C}_0}\lim_{h\to 0} \frac{\mu_h - \mu_0}{h} = \mu_0 \notin \mathcal{M}_0(\mathcal{X}).$$

Remark 2.13 The family $(\mu_\theta)_{\theta\in\Theta}$ is differentiable in a much stronger sense than just $\mathcal{C}_0$-differentiability. Thus we obtain $\mu'_0 = \mu_0$ since calculation gives for any $\alpha \in (0,\infty)$ that

$$\begin{aligned}\mu_0'[0,\alpha] &= \lim_{h\downarrow 0}\frac{\mu_h([0,\alpha])-\mu_0([0,\alpha])}{h}\\ &= \lim_{h\downarrow 0}\frac{1}{1-h}\int_{s\in(h,1)}\mathbb{U}_{[0,\frac{1}{s}]}([0,\alpha])\,ds \;-\; \int_{s\in(0,h]}\mathbb{U}_{[0,\frac{1}{s}]}([0,\alpha])\,ds\\ &= \int_{s\in(0,1)}\mathbb{U}_{[0,\frac{1}{s}]}([0,\alpha])\,ds \;-\; 0 \;=\; \mu([0,\alpha]).\end{aligned}$$

Remark 2.14 Example 2.12 shows that in general for locally compact $\mathcal{X}$ the unit ball $\mathbb{B}_0 := \mathbb{B}_{Var}\cap\mathcal{M}_0(\mathcal{X})$ of $(\mathcal{M}_0(\mathcal{X}),\|.\|_{Var})$ is not $\mathcal{C}_0$-compact. However, in the case that $\mathcal{X}$ is even compact, we obtain by the fact that $\mathcal{C}_0=\mathcal{C}$ that $\mathbb{B}_0$ is as the intersection of the $\mathcal{C}$-compact set $\mathbb{B}_{Var}$ with the $\mathcal{C}$-closed set $\mathcal{M}_0=\{\mu \mid \int \mathbb{I}_{\mathcal{X}}\,d\mu=0\}$ itself $\mathcal{C}$-compact (= $\mathcal{C}_0$-compact).

Example 2.15 Let μ_θ denote the measures defined on $[0,\infty)$ in Example 2.12. Let measures $\tilde{\mu}_\theta$ on $[0,\infty]$ be given by $\tilde{\mu}_\theta(A)=\mu_\theta(A\cap[0,\infty))$. Then according to Remark 2.11 $\tilde{\mu}_0'\in\mathcal{M}_0(\mathcal{C}([0,\infty]))$, i.e., $\tilde{\mu}_0'([0,\infty])=0$. Further

$$\tilde{\mu}_0'(A\cap[0,\infty))=\mu_0'(A)=\mu_0(A)=\tilde{\mu}_0(A\cap[0,\infty)).$$

Thus $\tilde{\mu}_0'(\{\infty\})=-1$ since the mass $\tilde{\mu}_0'([0,\infty))=1$ has to be compensated and the only set where this compensation can take place is the singleton $\{\infty\}$.

3 Sensitivity of stationary vectors

In this section we investigate the interdependence of the sensitivities of operators and their stationary vectors.

Theorem 3.1 *Let $(\Psi,\|.\|_\Psi)$ and $(V,\|.\|_V)$ be normed spaces. Suppose that V is Ψ-separated via the $\|.\|_\Psi$-$\|.\|_V$-continuous bilinear mapping $\langle .|.\rangle : \Psi\times V\to\mathbb{R}$. Let $W\subseteq V$ and let $\|.\|_W$ be the restriction of $\|.\|_V$ to W. Suppose that*

$$\mathbb{B}_W \text{ is } \Psi\text{-compact.} \tag{17}$$

Let $\Theta\subseteq\mathbb{R}$. Let $(L_\theta)_{\theta\in\Theta}$ be a parametrized pointwise $\|.\|_{\mathcal{L}}$-Lipschitz continuous family of operators in $\mathcal{L}(V,W)$ such that

$$\text{the restrictions } L_\theta^W \text{ of } L_\theta \text{ to } W \text{ are injective} \tag{18}$$

and

$$\text{the operators } L_\theta^W \text{ are } \Psi\text{-}\Psi\text{-continuous.} \tag{19}$$

Let $(\pi_\theta)_{\theta\in\Theta}$ be a pointwise $\|.\|_V$-Lipschitz-continuous family of vectors $\pi_\theta\in V$. Suppose that

$$(\forall\theta,\hat{\theta}\in\Theta)\ (\pi_\theta-\pi_{\hat{\theta}}\in W). \tag{20}$$

Suppose $\theta \mapsto L_\theta$ is Ψ-V-differentiable and that

$$(\forall \theta \in \Theta)\ (L_\theta \pi_\theta = 0) \tag{21}$$

holds. Then π_θ is Ψ-differentiable

$$L'_\theta \pi_\theta \in Im(L_\theta^W) \quad \text{and} \quad \pi'_\theta = -(L_\theta^W)^{-1} L'_\theta \pi_\theta.$$

Remark 3.2 Note that a linear operator $L_\theta^W : W \to W$ is Ψ-Ψ-continuous if and only if

$$(\forall \psi \in \Psi)\ (\exists \widetilde{\psi} \in \Psi) \quad \langle \psi | L_\theta^W . \rangle = \langle \widetilde{\psi} | . \rangle$$

Remark 3.3 Analogously to Theorem 1.14 we could have stated Theorem 3.1 for Ψ-Γ-differentiable $\theta \mapsto L_\theta$ with Γ a $\|.\|_V$-dense subset of V.

Remark 3.4 The Remarks 1.15 and 1.16 concerning Theorem 1.14 hold for Theorem 3.1 instead if we replace $\theta \mapsto K_\theta$ by $\theta \mapsto L_\theta$.

Under the hypothesis that $(\Psi, \|.\|_\Psi)$ is a Banach space we can also relax pointwise $\|.\|_V$-Lipschitz continuity of $\theta \mapsto \pi_\theta$ by Proposition A.2 to pointwise Ψ-Lipschitz continuity.

Proof of Theorem 3.1: Note that W is Ψ-separated since it is a subspace of the Ψ-separated space V. Note further that the hypotheses - and thus also the conclusions - of Theorem 1.14 are fulfilled if we let $L_\theta = K_\theta$ and $\Gamma = V$. Without loss of generality we suppose that $0 \in \Theta$ and prove the assertions of Theorem 3.1 just for $\theta = 0$. To do this we suppose without loss of generality that 1 is a $\|.\|_V$-Lipschitz constant for $\theta \mapsto \pi_\theta$ at $\theta = 0$. Let $\psi \in \Psi$ be arbitrary. By (21)

$$0 = \lim_{h \to 0} \left\langle \psi \,\middle|\, \frac{L_h \pi_h - L_0 \pi_0}{h} \right\rangle =$$

$$\lim_{h \to 0} \left\langle \psi \,\middle|\, \frac{[L_h - L_0]\pi_0}{h} \right\rangle + \lim_{h \to 0} \left\langle \psi \,\middle|\, \frac{[L_h - L_0][\pi_h - \pi_0]}{h} \right\rangle + \lim_{h \to 0} \left\langle \psi \,\middle|\, \frac{L_0[\pi_h - \pi_0]}{h} \right\rangle.$$

The second term vanishes by pointwise norm-Lipschitz continuity of π_θ and L_θ. The first term converges by the Ψ-V-differentiability hypothesis on L_θ and Theorem 1.14 to $\langle \psi \mid L'_0 \pi_0 \rangle$ with $L'_0 \in \mathcal{L}(V, W)$. Thus the third term converges to $-\langle \psi \mid L'_0 \pi_0 \rangle$ and thus we obtain by application of (20) that

$$\underline{\left(\lim_{h \to 0} \left\langle \psi \,\middle|\, \frac{L_0^W[\pi_h - \pi_0]}{h} \right\rangle \right)_{\psi \in \Psi}} = \left(\lim_{h \to 0} \left\langle \psi \,\middle|\, \frac{L_0[\pi_h - \pi_0]}{h} \right\rangle \right)_{\psi \in \Psi} =$$

$$\underline{\left(- \langle \psi \,| L'_0 \pi_0 \rangle \right)_{\psi \in \Psi}},$$

i.e., we obtain that

$$^{\Psi}\lim \left[L_0^W \frac{\pi_h - \pi_0}{h} \right] = -L'_0 \pi_0. \tag{22}$$

By the hypothesis that $\theta \mapsto \pi_\theta$ is $\|.\|_V$-Lipschitz continuous at 0 with Lipschitz-constant 1 and (20), we obtain that

$$\left\{ \frac{\pi_h - \pi_0}{h} \middle| h \in \Theta \setminus \{0\} \right\} \subseteq \mathbb{B}_W \tag{23}$$

By an application of the fact that

> an injective continuous mapping from a compact Hausdorff space to an arbitrary Hausdorff space is a homeomorphism onto its image (compare with [12] Theorem 17.4)

to the conjunction of (17), (18) and (19) we conclude that

L_0^W is a Ψ-Ψ-homeomorphism from $\mathbb{B}_W$ onto the image $L_0^W(\mathbb{B}_W)$,

i.e.,

$$\text{the inverse } (L_0^W)^{-1} \text{ of } L_0^W \text{ is } \Psi\text{-}\Psi\text{-continuous from } L_0^W(\mathbb{B}_W) \text{ to } \mathbb{B}_W. \tag{24}$$

We conclude from (22), (23) and (24) that

$$^\Psi\lim \frac{\pi_h - \pi_0}{h} = (L_0^W)^{-1} \left({}^\Psi\lim \left[L_0^W \frac{\pi_h - \pi_0}{h} \right] \right) = -(L_0^W)^{-1} L_0' \pi_0.$$

Thus

$$\pi_0' :=^\Psi \lim_{h \to 0} \frac{[\pi_h - \pi_0]}{h} \quad \text{exists,} \quad L_0'\pi_0 \in Im(L_0^W) \quad \text{and} \quad \pi_0' = -(L_0^W)^{-1} L_0' \pi_0.$$

□

Lemma 3.5 *Let Ψ and W be vector spaces and let $\langle . \mid . \rangle : \Psi \times W \to \mathbb{R}$ be a bilinear mapping, such that W is Ψ-separated.*

$$\textit{Let } K : W \to W \textit{ be a } \Psi\text{-}\Psi\textit{-continuous operator.} \tag{25}$$

Suppose that

$$(\forall w \in W) \quad {}^\Psi\sum_{i=0}^{\infty} K^i w \quad \textit{exists} . \tag{26}$$

Then $w \in W$ implies that

$$^\Psi\lim_{n \in N} K^n w = 0, \tag{27}$$

$$(id - K) : W \to W \textit{ is a bijection, and } {}^\Psi\sum_{i=0}^{\infty} K^i w = (id - K)^{-1} w. \tag{28}$$

Proof: (27) is an immediate consequence of (26). Further

$$(id - K)\sum_{i=0}^{n} K^i w = (id - K^{n+1})w \tag{29}$$

From (25), (26), (29) and (27) we conclude that

$$(id-K)\,{}^{\Psi}\!\sum_{i=0}^{\infty} K^i w ={}^{\Psi} \lim_{n\to\infty}(id-K)\sum_{i=0}^{n} K^i w ={}^{\Psi} \lim_{n\to\infty}(id-K^{n+1})w = w. \tag{30}$$

Analogously we conclude that

$${}^{\Psi}\!\sum_{i=0}^{\infty} K^i(id - K)w = w. \tag{31}$$

From (30) and (31) we obtain that $(id - K)$ is a bijection from W to W and that ${}^{\Psi}\sum_{i=0}^{\infty} K^i w = (id - K)^{-1}w$, i.e., (28) has been proved. □

Theorem 3.6 *Let the hypotheses of Theorem 3.1 with the exception of (18) be fulfilled. Let $K_\theta^W := id - L_\theta^W$ and let $K_\theta = id - L_\theta$. Suppose that*

$$(\forall w \in W) \qquad {}^{\Psi}\!\sum_{i=0}^{\infty} \left[K_\theta^W\right]^i w \quad \textit{exists.} \tag{32}$$

Then the Ψ-derivative π'_θ of $\theta \mapsto \pi_\theta$ exists and is given by

$$\pi'_\theta = (id - K_\theta^W)^{-1} K'_\theta \pi_\theta ={}^{\Psi} \sum_{i=0}^{\infty} K_\theta^i K'_\theta \pi_\theta. \tag{33}$$

Proof: Note that Ψ-Ψ-continuity of L_θ^W implies Ψ-Ψ-continuity of K_θ^W. Note further that for $K := K_\theta^W$ (32) becomes (26) and thus the hypotheses of Lemma 3.5 are fulfilled. Thus Lemma 3.5 (28) holds with $K = K_\theta^W$ and we conclude that $L_\theta^W = (id - K_\theta^W)$ is invertible and thus that (18) holds. Thus the hypotheses of Theorem 3.1 as well as the hypotheses of Lemma 3.5 with $K = K_\theta^W$ are fulfilled. We thus conclude (33) from the conjunction of the consequences of Theorem 3.1 and Lemma 3.5 with $K = K_\theta^W$, noting that $K'_\theta = -L'_\theta$. □

Remark 3.7 If $(V, \|.\|_V)$ and $(\Psi, \|.\|_\Psi)$ are Banach spaces and Θ is locally compact then by remarks 1.16 and 3.4 the Theorem 3.6 holds without any Lipschitz continuity hypothesis on $\theta \mapsto L_\theta$ resp. $\theta \mapsto K_\theta$.

Theorem 3.8 *Let $\Theta \subseteq \mathbb{R}$ be locally compact. Let $(K_\theta)_{\theta\in\Theta}$ be a family of stochastic kernels*[17] *on a compact second countable space $\mathcal{X}$ with stationary probability distributions $(\pi_\theta)_{\theta\in\Theta}$ (i.e., $\pi_\theta \in \mathcal{M}(\mathcal{X})$ is positive, $\pi_\theta(\mathcal{X}) = 1$ and*

[17] See Remark 2.3 for the definition of a stochastic kernel.

$K\pi_\theta = \pi_\theta$). Suppose that $\theta \mapsto K_\theta$ is $\mathcal{C}(\mathcal{X})$-$\mathcal{M}(\mathcal{X})$-differentiable and that $\theta \mapsto \pi_\theta$ is pointwise $\|.\|_{Var}$-Lipschitz continuous. Suppose further that the operators $K_\theta^{\mathcal{M}_0} := K_\theta \Big|_{\mathcal{M}_0(\mathcal{X})}$ are $\mathcal{C}(\mathcal{X})$-$\mathcal{C}(\mathcal{X})$-continuous and that

$$\forall \pi \in \mathcal{M}_0(\mathcal{X}) \quad {}^{\mathcal{C}}\sum_{n\in\mathbb{N}} [K_\theta^{\mathcal{M}_0}]^n \pi \quad \textit{exists.}$$

Then $\theta \mapsto \pi_\theta$ is $\mathcal{C}(\mathcal{X})$-differentiable and its derivative $\theta \mapsto \pi'_\theta$ fulfills

$$\pi'_\theta = \left(id - K_\theta^{\mathcal{M}_0}\right)^{-1} K'_\theta \pi_\theta =^{\mathcal{C}} \sum_{i=0}^{\infty} K_\theta^i K'_\theta \pi_\theta.$$

Proof: Let $(\Psi, \|.\|_\Psi) := (\mathcal{C}, \|.\|_\infty)$, let $(V, \|.\|_V) := (\mathcal{M}, \|.\|_{Var})$ and let $(W, \|.\|_W) := (\mathcal{M}_0, \|.\|_{Var})$. Since $(\mathcal{C}, \|.\|_\infty)$ and $(\mathcal{M}, \|.\|_{Var})$ are Banach spaces, Remark 3.7 implies that the hypotheses of Theorem 3.6 are fulfilled. □

Remark 3.9 The operators K'_θ in Theorem 3.8 are by Theorem 2.9 kernel operators. By Remark 3.4 (Proposition A.2) we can replace the Lipschitz continuity conditions on $\theta \mapsto \pi_\theta$ in Theorem 3.8 by pointwise $\mathcal{C}(\mathcal{X})$-Lipschitz continuity.

Remark 3.10 Note that for non-compact locally compact spaces $\mathcal{X}$ the proof of Theorem 3.8 does not work, since the unit ball $\mathbb{B}_0$ of $(\mathcal{M}_0, \|.\|_{Var})$ is by Remark 2.14 in general not $\mathcal{C}_0$-compact. Thus one has either to replace the space $(\mathcal{M}_0, \|.\|_{Var})$ by an appropriate space - depending on the parametric family $(K_\theta)_{\theta\in\Theta}$ under consideration - or one has to compactify $\mathcal{X}$ and work within the compactification, for example within the one point compactification. Corresponding to the use of the one point compactification one can replace $\mathcal{C}_0$ by the direct sum $\mathcal{C}_0 \oplus \mathbb{R}$ (of $\mathcal{C}_0$ and the one dimensional space of constant functions on the locally compact space $\mathcal{X}$ under consideration) and $\mathcal{M}_0(\mathcal{X})$ by the subspace $\{(\mu, c_\mu) \mid \mu(\mathcal{X}) + c_\mu = 0\}$ of $\mathcal{M}(\mathcal{X}) \times \mathbb{R}$.

4 Application

We apply the results obtained so far to the sensitivity of a discrete approximation of the truncated Brownian bridge with respect to the truncation parameter θ (see Example 4.5 and Remark 4.6). By (40) we see that in the case of Example 4.5 the sensitivity analysis of a parametric family π_θ of probabilities on an m-dimensional space boils down to the calculation of the sensitivities of one-dimensional distributions.

Theorem 4.1 *Let $\Theta \subseteq \mathbb{R}$ be locally compact. Suppose that $K_{\theta,i} \in \mathcal{L}(\mathcal{M}(X))$ with $i = 1, \ldots, m$ are parametrized families of stochastic kernels on a compact second countable space $\mathcal{X}$ with a joint family of stationary probability distributions $(\pi_\theta)_{\theta\in\Theta}$, i.e., such that $K_{\theta,i}\pi_\theta = \pi_\theta \in \mathcal{M}(X)$ for $i = 1, \ldots, m$. Suppose that the mappings $\theta \mapsto K_{\theta,i}$ are $\mathcal{C}(\mathcal{X})$-$\mathcal{M}(\mathcal{X})$-differentiable, that the operators*

$K_{\theta,i}$ are $\mathcal{C}(\mathcal{X})$-$\mathcal{C}(\mathcal{X})$-continuous and that $\theta \mapsto \pi_\theta$ is pointwise $\|.\|_{Var}$-Lipschitz continuous. Let further $K_\theta := K_{\theta,1}K_{\theta,2}\dots K_{\theta,m}$ and suppose that

$$\forall \pi \in \mathcal{M}_0(\mathcal{X}) \quad {}^{\mathcal{C}}\sum_{n\in\mathbb{N}} K_\theta^n \pi \quad \textit{exists.} \tag{34}$$

Then $\theta \mapsto \pi_\theta$ is $\mathcal{C}(\mathcal{X})$-differentiable and its derivative $\theta \mapsto \pi'_\theta$ fulfills

$$\pi'_\theta = \left(id - K_\theta^{\mathcal{M}_0}\right)^{-1} \hat{K}_\theta \pi_\theta =^{\mathcal{C}} \sum_{i=0}^{\infty} K_\theta^i \hat{K}_\theta \pi_\theta, \tag{35}$$

with $\hat{K}_\theta = \sum_{j=1}^m K_{\theta,1}K_{\theta,2}\dots K_{\theta,j-1}K'_{\theta,j}$

Proof: The theorem follows from Theorem 3.8 and Lemma A.5 taking $K_{\theta,i}\pi_\theta = \pi_\theta$ for $i = 1,\dots,m$ into account. □

Remark 4.2 Theorem 4.1 can be viewed as a result concerning the differentiation of a parametric family of systematic scan Gibbs samplers, with the operators $K_{\theta,i}$ providing versions of the respective conditional probabilities. (Consult [8] Section 6 for more information concerning the Gibbs sampler.)

Remark 4.3 The Lipschitz continuity hypothesis concerning $\theta \mapsto \pi_\theta$ in Theorem 4.1 can be weakened in analogy with Remark 3.9.

Proposition 4.4 *Let $m \geq 1$ be a fixed integer, and let $\rho \in (1,\infty)$ be fixed. Let ν_ρ be a probability measure on $[-\rho,\rho]^m$ having a density f (with respect to Lebesgue measure λ) such that $1/c \leq f \leq c$ for some constant $c \in (0,\infty)$. Let for $\theta \in \left(\frac{1}{\rho},\rho\right)$ probability measures ν_θ be defined by*

$$\nu_\theta(B) := \frac{\nu_\rho(B \cap [-\theta,\theta]^m)}{\nu_\rho([-\theta,\theta]^m)}. \tag{36}$$

Then $\theta \mapsto \nu_\theta$ is $\|.\|_{Var}$-Lipschitz continuous (if $\left(\frac{1}{\rho},\rho\right]$ is endowed with the usual euclidean distance induced by $|.|$). □

Example 4.5 We denote by $N(e,\sigma^2)(.)$ the probability distribution of the normal law on $(\mathbb{R},\mathcal{B})$ with expectation e and variance σ^2, i.e., we let

$$\phi(x) := \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}} \quad \text{and} \quad N(e,\sigma^2)(B) := \int_{y\in B} \frac{1}{\sigma} \cdot \phi\left(\frac{y-e}{\sigma}\right)\, dy.$$

Let $n \in \mathbb{N}$ be such that $n > 2$ but otherwise arbitrary. Let $\mathcal{X} = \{0\}\times[-\rho,\rho]^{n-1}\times\{0\}$ and let for $i = 1,\dots,n-1$ and $\theta \in \left(\frac{1}{\rho},\rho\right]$

$$K_{\theta,i}\delta_x := \delta_{x_0} \otimes \delta_{x_1} \otimes \cdots \otimes \delta_{x_{i-1}} \otimes \mu_{\theta,x_{i-1},x_{i+1}} \otimes \delta_{x_{i+1}} \otimes \cdots \otimes \delta_{x_{n-1}} \otimes \delta_{x_n} \tag{37}$$

with

$$\mu_{\theta,x_{i-1},x_{i+1}}(A) := \frac{N\left(\max\left(\min\left(\frac{x_{i-1}+x_{i+1}}{2},\theta\right),-\theta\right),\frac{1}{2n}\right)(A\cap[-\theta,\theta])}{N\left(\max\left(\min\left(\frac{x_{i-1}+x_{i+1}}{2},\theta\right),-\theta\right),\frac{1}{2n}\right)([-\theta,\theta])} \tag{38}$$

and $x_0 = x_n = 0$.

Let $K_\theta := K_{\theta,1}K_{\theta,2}\dots K_{\theta,n-1}$.

We are going to sketch that in this example the hypotheses of Theorem 4.1 hold and thus that the calculation of K'_θ can - by an application of (35) - be reduced to the calculation of one dimensional derivatives.

To do this we need first of all for any $\theta \in (0,\infty)$ the existence and uniqueness of a probability π_θ on $(\mathcal{X},\mathcal{B})$ such that $\pi_\theta = K_\theta\pi_\theta$. Existence and uniqueness of such a stationary probability π_θ is a consequence of the Doeblin minorization theorem ([1] Theorem 9.1) if we can show that there exists a measure κ_θ on $\mathcal{X}$, such that

$$(\forall x\in\mathcal{X})\ (\forall B\in\mathcal{B}(\mathcal{X}))\ \ (K_\theta\delta_x(A)\geq\kappa_\theta(A)).$$

We note that such a measure κ_θ is given by $c_\theta\cdot\delta_0\otimes\bigotimes_{i=1}^{n-1}\mathbb{U}[-\theta,\theta]\otimes\delta_0$ for some sufficiently small $c_\theta > 0$ and $\mathbb{U}[-\theta,\theta]$ the uniform distribution on $[-\theta,\theta]$. Further by [1] Theorem 9.1 we obtain that the rate of convergence is geometrically bounded by

$$\sup_{x\in\mathcal{X},B\in\mathcal{B}}|[K_\theta^m\delta_x](A)-\pi_\theta(A)|\leq(1-c_\theta)^m$$

and thus that the convergence rate is geometrically bounded with respect to the variation norm by

$$\sup_{x\in\mathcal{X}}\|K_\theta^m\delta_x-\pi_\theta\|_{Var}\leq 2\cdot(1-c_\theta)^m. \tag{39}$$

From (39) we obtain that (34) holds. $\mathcal{C}(\mathcal{X})$-$\mathcal{C}(\mathcal{X})$-continuity of the operators $K_{\theta,i}$, i.e., continuity of $x\mapsto\int\phi\, d[K_{\theta,i}\delta_x]$ for continuous ϕ, is easily checked.

Further it is not difficult to check that the $\mathcal{M}(\mathcal{X})$-$\mathcal{C}(\mathcal{X})$-derivative $K'_{\theta,i}$ of $\theta\mapsto K_{\theta,i}$ exists. It is given by

$$K'_{\theta,i}\delta_x := \delta_0\otimes\delta_{x_1}\otimes\cdots\otimes\delta_{x_{i-1}}\otimes\mu'_{\theta,x_{i-1},x_{i+1}}\otimes\delta_{x_{i+1}}\otimes\cdots\otimes\delta_{x_{n-1}}\otimes\delta_0$$

with

$$\mu'_{\theta,x_{i-1},x_{i+1}} \ = \ k_1\cdot\delta_\theta + k_2\cdot\delta_{-\theta} - (k_1+k_2)\cdot\mu_{\theta,x_{i-1},x_{i+1}} \tag{40}$$

for appropriate $k_1,k_2\in\mathbb{R}$ depending on x_{i-1}, x_{i+1} and θ.

It finally remains to check that $\theta\mapsto\pi_\theta$ is pointwise $\|.\|_{Var}$-Lipschitz continuous. This may be considered as the most difficult part in applying our theory. We outline a short sketch of the arguments needed in the case of our example:

Let measures ν_θ implicitly be defined by $\pi_\theta =: \delta_0\otimes\nu_\theta\otimes\delta_0$, with δ_0 the Dirac measure at 0 on $\mathbb{R}$ and let implicitly operators $\widetilde{K}_\theta$ be defined by

$$\delta_0\otimes\left[\widetilde{K}_\theta\nu\right]\otimes\delta_0 := K_\theta[\delta_0\otimes\nu\otimes\delta_0].$$

One next shows that (37) and (38) imply (36) and that (37) and (38) further imply that there exists a constant $\widetilde{c_\theta} \in \mathbb{R}$ such that $\frac{d[\widetilde{K_\theta}\nu]}{d\lambda} \leq \widetilde{c_\theta}$ for any probability ν and thus especially $\frac{d\nu_\theta}{d\lambda} = \frac{d[\widetilde{K_\theta}\nu_\theta]}{d\lambda} \leq \widetilde{c_\theta}$. Thus the hypotheses of Proposition 4.4 are fulfilled and we conclude that $\theta \mapsto \nu_\theta$ and thus further that $\theta \mapsto \pi_\theta$ is pointwise $\|.\|_{Var}$-Lipschitz continuous.

Remark 4.6 The stationary distributions π_θ in Example 4.5 can be considered as discrete approximations of truncations (the paths are uniformly bounded by θ) of the Brownian bridge on $[0, 1]$.

Remark 4.7 As we already mentioned in Example 4.5 it is in general difficult to check the Lipschitz continuity of the mapping $\theta \mapsto \pi_\theta$. It would therefore be interesting to know whether Lipschitz continuity of $\theta \mapsto \pi_\theta$ can be replaced by other - more easily checkable - hypotheses without restricting our results to the situation considered in [5].

Remark 4.8 It would be interesting to know whether the results of this section lead to sensitivity estimates for the Gibbs sampler that are applicable in computer simulation.

A Appendix

Proposition A.1 *Let $(V, \|.\|_V)$ be a Banach space and let $(W, \|.\|_W)$ be a normed space. Let (Θ, d) be a metric space. A function $\theta \mapsto L_\theta \in \mathcal{L}(V, W)$ is pointwise $\|.\|_{\mathcal{L}}$-Lipschitz continuous if and only if it is pointwise V-$\|.\|_W$-Lipschitz continuous.*

Proof: Of course pointwise $\|.\|_{\mathcal{L}}$-Lipschitz continuity implies pointwise V-$\|.\|_W$-Lipschitz continuity. So it only remains to prove the converse. Suppose that $\theta \mapsto L_\theta \in \mathcal{L}(V, W)$ is pointwise V-$\|.\|_W$-Lipschitz continuous. Let $\theta_0 \in \Theta$ be arbitrarily chosen but fixed. Then for any $\theta \in \Theta \setminus \{\theta_0\}$ and any $v \in V$ we have that $\left\| \frac{[L_\theta - L_{\theta_0}]v}{d(\theta, \theta_0)} \right\| \leq l_v$ for some constant l_v. Thus the family $(R_\theta)_{\theta \in \Theta \setminus \{\theta_0\}}$ of operators $R_\theta := \frac{L_\theta - L_{\theta_0}}{d(\theta, \theta_0)}$ is pointwise bounded. By the uniform boundedness principle (see Proposition A.4) it is therefore uniformly bounded, which proves the proposition. □

Proposition A.2 *Let $(\Psi, \|.\|_\Psi)$ be a Banach space and let $(W, \|.\|_W)$ be a normed space. Let $\langle . | . \rangle : \Psi \times W \to \mathbb{R}$ be a bilinear mapping such that*

$$\|w\|_W \leq \sup_{\psi \in \mathbb{B}_\Psi} \langle \psi \mid w \rangle. \tag{41}$$

Let (Θ, d) be a metric space and let $\theta \mapsto \pi_\theta$ be a mapping from (Θ, d) to $(W, \|.\|_W)$. Then pointwise Ψ-Lipschitz continuity of $\theta \mapsto \pi_\theta$ implies pointwise $\|.\|_W$-Lipschitz continuity of $\theta \mapsto \pi_\theta$.

Proof: Let $\theta_0 \in \Theta$ be arbitrarily chosen but fixed. Pointwise Ψ-Lipschitz continuity of $\theta \mapsto \pi_\theta$ implies that for any $\theta \in \Theta \setminus \{\theta_0\}$ and any $\psi \in \Psi$ we have that $\left|\left\langle \psi \,\middle|\, \frac{\pi_\theta - \pi_{\theta_0}}{d(\theta,\theta_0)} \right\rangle\right| \le l_\psi$ for some $l_\psi \in \mathbb{R}$. Thus the family $(r_\theta)_{\theta \in \Theta \setminus \{\theta_0\}}$ of linear functionals $r_\theta(.) := \left\langle . \,\middle|\, \frac{\pi_\theta - \pi_{\theta_0}}{d(\theta,\theta_0)} \right\rangle$ from Ψ to $\mathbb{R}$ is pointwise bounded. By the uniform boundedness principle (Proposition A.4) r_θ is uniformly bounded on $\mathbb{B}_\Psi$. Thus pointwise $\|.\|_W$-Lipschitz continuity of $\theta \mapsto \pi_\theta$ follows from the definition of r_θ and (41). □

Lemma A.3 *Let $(\Psi, \|.\|_\Psi)$ and $(V, \|.\|_V)$ be Banach spaces and let $(W, \|.\|_W)$ be a normed space. Let $\langle .|. \rangle : \Psi \times W \to \mathbb{R}$ be a bilinear mapping such that $\|w\|_W \le \sup_{\psi \in \mathbb{B}_\Psi} \langle \psi \mid w \rangle$. Let (Θ, d) be a metric space. Then Ψ-V-Lipschitz continuity of $\theta \mapsto L_\theta \in \mathcal{L}(V, W)$ implies $\|.\|_{\mathcal{L}}$-Lipschitz continuity of $\theta \mapsto L_\theta$.*

Proof: If we fix v and let $\pi_\theta := L_\theta v$, then $\theta \mapsto \pi_\theta$ is Ψ-Lipschitz continuous. By an application of Proposition A.2 we obtain from the Ψ-Lipschitz continuity of $\theta \mapsto \pi_\theta$ that $\theta \mapsto \pi_\theta$ is $\|.\|_W$-Lipschitz continuous. This implies that $\theta \mapsto L_\theta$ is V-$\|.\|_W$-Lipschitz continuous. Thus an application of Proposition A.1 proves the Lemma. □

The following proposition is a version of the well known uniform boundedness principle. It follows immediately from [10] Chapter III Section 4.2 Corollary.

Proposition A.4 *Let $(V, \|.\|_V)$ be a Banach space and let $(W, \|.\|_W)$ be a normed space. Let $\mathcal{R} \subset \mathcal{L}(V, W)$ be such that $(\forall v \in V)$ $(\sup\{\|Rv\|_W \mid R \in \mathcal{R}\} < \infty)$. Then $\sup\{\|R\|_{\mathcal{L}} \mid R \in \mathcal{R}\} < \infty$.* □

Lemma A.5 *(Product rule for weak differentiation of Operators.) Let $(\Psi, \|.\|_\Psi)$ and $(V, \|.\|_V)$ be normed spaces. Suppose that V is Ψ-separated via the $\|.\|_\Psi$-$\|.\|_V$-continuous bilinear mapping $\langle .|. \rangle : \Psi \times V \to \mathbb{R}$. Let $\Theta \subseteq \mathbb{R}$. Let $(K_{\theta,1})_{\theta \in \Theta}$ and $(K_{\theta,2})_{\theta \in \Theta}$ be parametrized pointwise $\|.\|_{\mathcal{L}}$-Lipschitz continuous, Ψ-V-differentiable families of Ψ-Ψ-continuous operators in $\mathcal{L}(V)$, both bounded in norm. Then $(K_\theta)_{\theta \in \Theta} := (K_{\theta,1} K_{\theta,2})_{\theta \in \Theta}$ is also a parametrized pointwise $\|.\|_{\mathcal{L}}$-Lipschitz continuous, Ψ-V-differentiable family of Ψ-Ψ-continuous operators and*

$$K'_\theta = K_{\theta,1} K'_{\theta,2} + K'_{\theta,1} K_{\theta,2}. \tag{42}$$

Proof: Pointwise Lipschitz continuity of $\theta \mapsto K_\theta$ as well as Ψ-Ψ-continuity of the operators K_θ for $\theta \in \Theta$ are easy consequences of the respective hypotheses on $(K_{\theta,1})_{\theta \in \Theta}$ and $(K_{\theta,2})_{\theta \in \Theta}$. Thus we have to show the Ψ-V-differentiability of $\theta \mapsto K_\theta$ and that equation (42) holds. This is done as follows:

We suppose without loss of generality (by continuity of $\langle .|. \rangle : \Psi \times V \to \mathbb{R}$, rescaling the norms if necessary) that $\langle \psi \mid v \rangle \le \|\psi\|_\Psi \cdot \|v\|_V$ for all $\psi \in \Psi$ and $v \in V$.

Choose $\psi \in \Psi$ and $v \in V$ arbitrary. By Ψ-Ψ-continuity of $K_{\theta,1}$ and Remark 3.2 there exists $\widetilde{\psi} \in \Psi$ such that $\langle \widetilde{\psi} \mid . \rangle = \langle \psi \mid K_{\theta,1} . \rangle$. Let further $\widetilde{v} := K_{\theta,2} v$.

By pointwise $\|.\|_{\mathcal{L}}$-Lipschitz continuity and Ψ-V-differentiability of $\theta \mapsto K_{\theta,1}$ and $\theta \mapsto K_{\theta,2}$ we obtain that for any $\varepsilon > 0$ there exists an $\eta_\varepsilon \in (0, \infty)$ such that $|h| \leq \eta_\varepsilon$ implies

$$\frac{1}{h} \|K_{\theta+h,1} - K_{\theta,1}\|_{\mathcal{L}} \cdot \|K_{\theta+h,2} - K_{\theta,2}\|_{\mathcal{L}} \leq \varepsilon. \tag{43}$$

and

$$\left| \left\langle \psi \middle| \left[\frac{K_{\theta+h,1} - K_{\theta,1}}{h} - K'_{\theta,1} \right] \widetilde{v} \right\rangle \right| \leq \varepsilon, \ \left| \left\langle \widetilde{\psi} \middle| \left[\frac{K_{\theta+h,2} - K_{\theta,2}}{h} - K'_{\theta,2} \right] v \right\rangle \right| \leq \varepsilon. \tag{44}$$

Thus let $\varepsilon > 0$ be arbitrarily chosen and let $|h| \leq \eta_\varepsilon$. Note that the first $\leq$-sign in the calculation below follows from (43) while the second is a consequence of (44). We calculate:

$$\frac{\langle \psi | [K_{\theta+h} - K_\theta] v \rangle}{h} = \frac{\langle \psi | K_{\theta+h,1} [K_{\theta+h,2} - K_{\theta,2}] v \rangle}{h} + \frac{\langle \psi | [K_{\theta+h,1} - K_{\theta,1}] K_{\theta,2} v \rangle}{h} =$$

$$\frac{\langle \psi | [K_{\theta+h,1} - K_{\theta,1} + K_{\theta,1}][K_{\theta+h,2} - K_{\theta,2}] v \rangle}{h} + \frac{\langle \psi | [K_{\theta+h,1} - K_{\theta,1}] K_{\theta,2} v \rangle}{h} \leq$$

$$\|\psi\|_\Psi \cdot \varepsilon \cdot \|v\|_V + \frac{\langle \psi | K_{\theta,1} [K_{\theta+h,2} - K_{\theta,2}] v \rangle}{h} + \frac{\langle \psi | [K_{\theta+h,1} - K_{\theta,1}] K_{\theta,2} v \rangle}{h} =$$

$$\|\psi\|_\Psi \cdot \varepsilon \cdot \|v\|_V + \left\langle \widetilde{\psi} \middle| \frac{K_{\theta+h,2} - K_{\theta,2}}{h} \, v \right\rangle + \left\langle \psi \middle| \frac{K_{\theta+h,1} - K_{\theta,1}}{h} \, \widetilde{v} \right\rangle \leq$$

$$\|\psi\|_\Psi \cdot \varepsilon \cdot \|v\|_V + \varepsilon + \varepsilon + \langle \widetilde{\psi} | K'_{\theta,2} v \rangle + \langle \psi | K'_{\theta,1} \widetilde{v} \rangle =$$

$$\|\psi\|_\Psi \cdot \varepsilon \cdot \|v\|_V + \varepsilon + \varepsilon + \langle \psi | K_{\theta,1} K'_{\theta,2} v \rangle + \langle \psi | K'_{\theta,1} K_{\theta,2} v \rangle.$$

Since $\varepsilon > 0$ has been chosen arbitrarily the calculation shows that

$$\limsup_{h \to 0} \frac{\langle \psi | [K_{\theta+h} - K_\theta] v \rangle}{h} \ \leq \ \langle \psi | K_{\theta,1} K'_{\theta,2} v \rangle + \langle \psi | K'_{\theta,1} K_{\theta,2} v \rangle. \tag{45}$$

An analogous calculation shows that

$$\liminf_{h \to 0} \frac{\langle \psi | [K_{\theta+h} - K_\theta] v \rangle}{h} \ \geq \ \langle \psi | K_{\theta,1} K'_{\theta,2} v \rangle + \langle \psi | K'_{\theta,1} K_{\theta,2} v \rangle. \tag{46}$$

From (45) and (46) we obtain that

$$\lim_{h \to 0} \frac{\langle \psi | [K_{\theta+h} - K_\theta] v \rangle}{h} \ = \ \langle \psi | K_{\theta,1} K'_{\theta,2} v \rangle + \langle \psi | K'_{\theta,1} K_{\theta,2} v \rangle,$$

i.e., $\theta \mapsto K_\theta$ is Ψ-V-differentiable and (42) holds. □

Remark A.6 (Compare with the product rule presented in [7].) If $(\Psi, \|.\|_\psi)$ and $(V, \|.\|_V)$ are Banach spaces such that $\|v\|_V \leq \sup_{\psi \in \mathbb{B}_\Psi} \langle \psi \mid v \rangle$ then Lemma A.5 holds without imposing a Lipschitz condition on $K_{\theta,1}$ and $K_{\theta,2}$. This is concluded along the lines of 1.15 and 1.16.

Remark A.7 Considering Remark 2.11, Example 2.12, Example 2.15 and Remark 3.10 of this article, we note that [6] theorems 1, 2 and 3 are literally incorrect. These theorems as well as their proofs become immediately correct if the spaces Y considered in [6] are replaced by their one point compactifications Y_∞, respectively [6] Theorem 1 becomes already correct if the space $\mathcal{C}_c^\bullet(Y) := \mathcal{C}_c(Y) \cup \mathbb{I}_Y$ in the statement of [6] Theorem 1 is replaced by $\mathcal{C}_c(Y)$.

Acknowledgement: I would like to thank the referee for his very accurate remarks that helped to improve the paper. Especially I would like to thank the referee for pointing out an error in a previous draft of this article. I would like to thank E. Makai for reading large parts of a nearly final version of this article.

References

[1] R. Bhattacharya, M. Majumdar *Random dynamical systems. Theory and applications.* Cambridge: Cambridge University Press. (2007).

[2] A. Bobrowski *Functional analysis for probability and stochastic processes.* Cambridge: Cambridge University Press. (2005).

[3] N. Bourbaki *Elements of mathematics. General topology. Chapters 5–10. Transl. from the French. Softcover edition of the 2nd printing 1989. (English)* Berlin: Springer. (1998).

[4] L. LeCam *Asymptotic methods in statistical decision theory.* Springer Series in Statistics. New York etc.: Springer-Verlag. (1986).

[5] B. Heidergott, A. Hordijk and H. Weisshaupt Measure-Valued Differentiation for Stationary Markov Chains. *Mathematics of Operations Research, Vol. 31, Number 1, 154-172 (2006).*

[6] B. Heidergott, A. Hordijk, and H. Weisshaupt, Derivatives of Markov kernels and and their Jordan decomposition. *J. Appl. Anal. 14, No. 1, 13-26 (2008).*

[7] B. Heidergott, F. Vazquez-Abad, G. Pflug and T. Farenhorst-Yuan Gradient estimation for discrete-event systems by measure-valued differentiation. *Transactions on Modeling and Computer Simulation (accepted January 2009).*

[8] J.S. Liu *Monte Carlo strategies in scientific computing.* Springer Series in Statistics. New York, NY: Springer.

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

[10] H.H. Schaefer, M.P. Wolff *Topological vector spaces. 2nd ed.* Graduate Texts in Mathematics. 3. New York, NY: Springer. (1999).

[11] H. Weisshaupt Local convexity, adapted experiments and the wave equation: Lecam's randomization criterion and related results. *Stud. Sci. Math. Hung. 43, No. 4, 443-486 (2006).*

[12] S. Willard *General topology.* Addison-Wesley Series in Mathematics. Reading, Mass. etc.: Addison-Wesley Publishing Company. (1970).