Geometric analysis for symmetric Fleming-Viot operators: Rademacher's theorem and exponential families

Alexander Schied
Institut für Mathematik, MA 7-4
Technische Universität Berlin
Strasse des 17. Juni 136
10623 Berlin, Germany
e-mail: schied@mathematik.hu-berlin.de

Revised version: August 17, 2001

Abstract

We use the natural geometry of a symmetric Fleming-Viot operator $\mathcal{L}$ to obtain analytical descriptions of the corresponding Dirichlet space $(\mathcal{E}, D(\mathcal{E}))$. In particular, we give a complete characterization of functions in $D(\mathcal{E})$ in terms of their differentiability properties along exponential families. Moreover, we prove a Rademacher theorem stating that any function which is Lipschitz continuous with respect to the Bhattacharya distance is contained in $D(\mathcal{E})$ and possesses a bounded gradient. A converse to this statement is also given. Thus, we relate the Bhattacharya distance to the potential theory of $\mathcal{L}$.

Key words: Rademacher's theorem, Fleming-Viot process, Bhattacharya distance, Dirichlet form, Dirichlet point process, weak differentiability.

MSC 2000 classification: 60G57, 31C25, 58J65, 58B20

1 Introduction

In this paper, we examine the generator $\mathcal{L}$ and the symmetric Dirichlet form $(\mathcal{E}, D(\mathcal{E}))$ associated with a Fleming-Viot process with parent-independent mutation. This process is a widely studied infinite-dimensional diffusion taking values in the space $\mathcal{M}_1 := \mathcal{M}_1(E)$ of probability measures over a Polish phase space E; see e.g. [6], [9],[8], [18]. Here, our main concern will be to characterize analytically the Dirichlet space $(\mathcal{E}, D(\mathcal{E}))$ in terms of the differentiability properties of its members. It is well-known that the form $(\mathcal{E}, D(\mathcal{E}))$ is of gradient type. But, as often in infinite dimensions, the gradient $Du(\mu)$ of a function $u \in D(\mathcal{E})$ does not depend on all values u takes in a small neighborhood of the point $\mu \in \mathcal{M}_1$, but only on those which lie in certain directions. The same phenomenon occurs on Wiener space, where only differentiability in directions of the Cameron-Martin space is relevant. In our case, however, the convex structure of $\mathcal{M}_1$ does not provide the natural geometry for $(\mathcal{E}, D(\mathcal{E}))$. Instead, the natural 'tangent space' on $\mathcal{M}_1$ varies from point to point and so does the corresponding inner product. As observed by Overbeck et al. [18], one should take as the tangent space in μ the set $L^2(\mu) := L^2(E, \mu)$ endowed with an inner product given by the covariance with respect to μ. Moreover, it was shown in a previous paper of the author [23] that the intrinsic metric of the Dirichlet form $(\mathcal{E}, D(\mathcal{E}))$ is given by the Bhattacharya distance ρ on $\mathcal{M}_1$. When ρ is restricted to the subspace of all measures which are absolutely continuous with respect to a given reference measure, one obtains the geometry of an infinite dimensional sphere segment. This aspect of ρ possesses applications in statistics (see, for instance, Amari [3]) and in differential geometry (see Friedrich [10]). For Fleming-Viot processes, ρ must be considered on the full set $\mathcal{M}_1$, and one obtains a hierarchical structure of infinite-dimensional sphere segments. Thus, the symmetric Fleming-Viot process provides an explicit case study of an infinite-dimensional process on a curved and non-flat state space.

Our first main result, Theorem 4, characterizes all functions in $D(\mathcal{E})$ by their differentiability properties along those curves in $\mathcal{M}_1$ which are formed by exponential families obtained from bounded measurable functions. It turns out that the arc length of an exponential family measured by the Bhattacharya distance relates exactly to the gradient of our form.

Another main result is an infinite-dimensional version of the celebrated theorem of Rademacher [19] stating that Lipschitz functions on $\mathbb{R}^d$ are differentiable almost everywhere and in the weak sense. In our Theorem 19, we show that a function on $\mathcal{M}_1$ which is Lipschitz continuous with respect to the Bhattacharya distance lies in $D(\mathcal{E})$ and possesses a bounded gradient. Also a converse of this theorem holds: a function in $D(\mathcal{E})$ with bounded gradient possesses a Lipschitz version.

Generalizations and extensions of Rademacher's theorem have recently been of interest also in other contexts: abstract Wiener space (and some of its generalizations) were considered by Kusuoka [13], Enchev and Stroock [7], Bogachev and Mayer-Wolf [4], and others. The paper [22] of Röckner and the author deals with the configuration space of infinite particle systems as another infinite dimensional and non-flat case study. Independently from these developments in probability theory, Cheeger [5] obtained general versions of

Rademacher's theorem on abstract metric measure spaces. However, Cheeger's assumptions include, for instance, a Vitali covering condition, so that his results and methods do not apply to the infinite-dimensional case studies mentioned before.

The paper is organized as follows. After the setup has been introduced, we give a characterization of the reversible distribution m of $\mathcal{L}$, the so-called Dirichlet point process, in terms of an integration by parts formula. This formula is one of the main tools in deriving Theorem 4, which connects the Dirichlet space $(\mathcal{E}, D(\mathcal{E}))$ with the differentiability properties of its members along exponential families. Theorem 4 is stated in Section 4. In the subsequent Section 5, we discuss the consequences of replacing the original σ-field on the phase space E by a smaller one. This technique is needed for the proof of our Rademacher-type theorem, which itself is presented in Section 7. For the proof, we need also some understanding of the geometric relations between exponential families, Fisher information, and Bhattacharya distance. The corresponding results are the subject of Section 6. In the final Section 8, we address some questions concerning the potential theory of $(\mathcal{E}, D(\mathcal{E}))$. For instance, we study the behavior of the Bhattacharya distance function of a given subset of $\mathcal{M}_1$ and show that m-zero sets have exceptional intrinsic interior.

2 Preliminaries and notation

Let $(E, \mathcal{B})$ be a standard Borel measurable space and $\mathcal{M}_1 := \mathcal{M}_1(E, \mathcal{B})$ the corresponding space of probability measures. $\mathcal{M}_1$ will be endowed with the σ-algebra $\mathcal{F}$ which is generated by the maps $\mu \mapsto \langle f, \mu \rangle := \int f \, d\mu$ with $f \in B_b(E)$, the space of bounded $\mathcal{B}$-measurable functions. A suitable set $\mathcal{F}C_b^\infty$ of "smooth" test functions on $\mathcal{M}_1$ is provided by functions u of the form

$$u(\mu) = F\big(\langle f_1, \mu \rangle, \ldots, \langle f_n, \mu \rangle\big) \tag{1}$$

where $n \in \mathbb{N}$, F is a C^∞-function on $\mathbb{R}^n$ and $f_i \in B_b(E)$. For $u \in \mathcal{F}C_b^\infty$ its gradient $Du : \mathcal{M}_1 \times E \to \mathbb{R}$ is defined by taking the derivative of u in the direction of the Dirac mass δ_x in a point $x \in E$:

$$Du(\mu, x) = \frac{d}{dt}\Big|_{t=0} u\big((1-t)\mu + t\delta_x\big)\,.$$

For u as in (1) one obtaines that

$$Du(\mu, x) = \sum_{i=1}^{n} \partial_i F\big(\langle f_1, \mu \rangle, \ldots, \langle f_n, \mu \rangle\big)\big(f_i(x) - \langle f_i, \mu \rangle\big)\,. \tag{2}$$

Here, $\partial_i F$ denotes the i^{th} partial derivative of F. Thinking of $Du(\mu)$ as a "tangent vector" on $\mathcal{M}_1$ suggests that the tangent space $T_\mu \mathcal{M}_1$ should consist of all those functions $f \in L^2(\mu)$ for which $\langle f, \mu \rangle = 0$. We endow this tangent space $T_\mu \mathcal{M}_1$ with the usual inner product $(\cdot,\cdot)_\mu$ on $L^2(\mu)$. To simplify notations, it will be useful to extend this inner product to all of $L^2(\mu)$ by combining it with the canonical projection onto $T_\mu\mathcal{M}_1$ so that

$$(f,g)_\mu := \mathrm{cov}_\mu(f,g)\,, \quad \text{and} \quad \|f\|_\mu^2 := (f,f)_\mu = \mathrm{var}_\mu(f)\,,$$

where cov_μ and var_μ denote, respectively, the covariance and the variance with respect to μ.

If A is the generator of a conservative Markov process on E, the (formal) generator of a Fleming-Viot process with mutation operator A is given by

$$\mathcal{L}u(\mu) = \tfrac{1}{2}\langle D^2u(\mu),\mu\rangle + \langle ADu(\mu),\mu\rangle\,, \tag{3}$$

where $D^2u(\mu,x) := D\big(Du(\cdot,x)\big)(\mu,x)$ and the operator A acts on the E-component of the function Du. For u as in (1) with $f_i \in D(A)$, $\mathcal{L}u(\mu)$ takes the following form:

$$\begin{aligned}\mathcal{L}u(\mu) \;=\; & \sum_{i,j=1}^n \partial_i\partial_j F\big(\langle f_1,\mu\rangle,\dots,\langle f_n,\mu\rangle\big)(f_i,f_j)_\mu \qquad (4)\\ & \qquad + \sum_{i=1}^n \partial_i F\big(\langle f_1,\mu\rangle,\dots,\langle f_n,\mu\rangle\big)\langle Af_i,\mu\rangle\,.\end{aligned}$$

For a large class of generators A it is possible to associate $\mathcal{L}$ with a measure-valued diffusion process; see, e.g., [6], [9], [8]. In the present paper, we will be interested in the case where $\mathcal{L}$ allows for a reversible stationary distribution m. At least under the assumption of the irreducibility of A, such a reversible stationary distribution can be found if and only if A is of uniform type, i.e.,

$$Af(x) = \frac{1}{2}\int \big(f(y) - f(x)\big)\,\nu(dy) \tag{5}$$

for some finite positive measure ν on $(E,\mathcal{B})$; see [17] for finite E and [14] in the general case. Thus, we will concentrate in the sequel on a mutation operator of the form (5) with ν fixed. Then $\mathcal{L}u$ is well-defined by (3) or (4) for all $u \in \mathcal{F}C_b^\infty$. Moreover, it is known that the unique stationary distribution m of $\mathcal{L}$ is given by the so-called Dirichlet point process with parameter ν. This random measure can be characterized by the following property. If $B_1,\dots,B_n$ is a partition of E into disjoint measurable sets, then the mapping $\mathcal{M}_1 \ni \mu \mapsto \big(\mu(B_1),\dots,\mu(B_n)\big)$ possesses under m a Dirichlet distribution

$$\frac{\Gamma(\nu(E))}{\Gamma(\nu(B_1))\cdots\Gamma(\nu(B_n))}x_1^{\nu(B_1)-1}\cdots x_n^{\nu(B_n)-1}\,\delta_{1-\sum_{k=1}^{n-1}x_k}(dx_n)\,dx_{n-1}\cdots dx_1 \tag{6}$$

on the $(n-1)$-dimensional simplex $\big\{(x_1,\dots,x_n)\;\big|\; x_i \geq 0,\; \sum_{k=1}^n x_k = 1\big\}$; see [8], [9]. With $\mathcal{L}$ and m we can associate a symmetric Dirichlet form $(\mathcal{E},D(\mathcal{E}))$ as the closure of

$$\mathcal{E}(u,v) := \frac{1}{2}\int (Du(\mu),Dv(\mu))_\mu\, m(d\mu)\,, \qquad \text{for } u,\, v \in \mathcal{F}C_b^\infty\,;$$

see [18]. Clearly, (Du,Dv) is the carré du champs operator associated with $\mathcal{L}$.

Remark 1 It is easy to see that any $u \in D(\mathcal{E})$ possesses a gradient Du, which is contained in the L^2-space with respect to the measure $\overline{m}(dx, d\mu) := \mu(dx)\, m(d\mu)$ on $\mathcal{M}_1 \times E$. Indeed, if $(u_n)_{n\in\mathbb{N}}$ is a sequence in $\mathcal{F}C_b^\infty$ approximating u in $D(\mathcal{E})$, then the corresponding gradients $(Du_n)_{n\in\mathbb{N}}$ form a Cauchy sequence in $L^2(\overline{m})$, and we can take Du as its limit point.

For $f \in B_b(E)$ define $S_f : \mathcal{M}_1 \to \mathcal{M}_1$ by

$$d(S_f\mu) := \frac{e^f}{\langle e^f, \mu\rangle}\, d\mu\,.$$

K. Handa [11] showed that m is quasi-invariant with respect to S_f, i.e., the image measure $m \circ S_f = m \circ (S_{-f})^{-1}$ is equivalent to m. Moreover, he calculated the corresponding density:

$$\frac{dm \circ S_f}{dm}(\mu) = \exp\left[\langle f, \nu\rangle - \nu(E)\log\langle e^f, \mu\rangle\right]. \tag{7}$$

See also [29] and [30] for related results. In his paper, Handa shows also that the Dirichlet point process m is the only stationary distribution for an operator of type (3) which is quasi-invariant with respect to S_f.

It was shown in [23] that the intrinsic metric of $\mathcal{E}$,

$$\rho(\mu, \lambda) := \sup\left\{u(\mu) - u(\lambda)\ \middle|\ u \in \mathcal{F}C_b^\infty,\ (Du, Du) \le 1\right\}, \tag{8}$$

is given by the Bhattacharya distance on $\mathcal{M}_1$:

$$\rho(\mu, \lambda) = 2\arccos\int\sqrt{\frac{d\mu}{d\eta}\cdot\frac{d\lambda}{d\eta}}\, d\eta$$

with η being any measure dominating both μ and λ. For instance, one can take $\eta := (\mu + \lambda)/2$.

3 An integration by parts formula

Recall that the operator A is fixed throughout the paper as $Af(x) = \frac{1}{2}\int\big(f(y) - f(x)\big)\,\nu(dy)$. A part of the following proposition is already implicitly contained in [9], but we give the full proof here, since the result will be crucial in the sequel.

Proposition 2 *For a given measure m on $\mathcal{M}_1(E)$ the following conditions are equivalent.*

1. *m is a Dirichlet point process with parameter ν.*
2. *For u, $v \in \mathcal{F}C_b^\infty$ and $f \in B_b(E)$,*

$$\int (Du, f)v\, dm = -\int u(Dv, f)\, dm - \int uv\langle Af, \cdot\rangle\, dm\,. \tag{9}$$

Moreover, if either condition 1 or condition 2 holds, then (9) is true for all $u, v \in D(\mathcal{E})$.

Proof: First we show that 1 implies 2. For $u \in \mathcal{F}C_b^\infty$ it is easy to see that $\frac{d}{dt}u(S_{tf}\mu) = (Du(S_{tf}\mu), f)_{S_{tf}\mu}$. Thus

$$\frac{d}{dt}\bigg|_{t=0} \int (u \circ S_{tf}) v \, dm = \int (Du, f) v \, dm \,. \tag{10}$$

On the other hand, Handa's result [11] on quasi-invariance of m implies that

$$\int (u \circ S_{tf}) v \, dm = \int u (v \circ S_{-tf}) \frac{dm \circ S_{tf}^{-1}}{dm} \, dm \,. \tag{11}$$

Moreover, by (7),

$$\frac{d}{dt}\bigg|_{t=0} \frac{dm \circ S_{tf}^{-1}}{dm}(\mu) = -\langle f, \nu \rangle + \nu(E)\langle f, \mu \rangle = -\langle Af, \mu \rangle \,.$$

Since all terms are bounded, we may interchange differentiation and integration when differentiating the right hand side of (11);

$$\frac{d}{dt}\bigg|_{t=0} \int u (v \circ S_{-tf}) \frac{dm \circ S_{tf}^{-1}}{dm} \, dm = -\int u (Dv, f) \, dm - \int uv \langle Af, \cdot \rangle \, dm \,.$$

Combining this identity with (10) yields Assertion 2.

Next, if (9) holds for all $u, v \in \mathcal{F}C_b^\infty$, then it follows from (2) that

$$-\int \mathcal{L}u \cdot v \, dm = \int (Du, Dv) \, dm = -\int u \cdot \mathcal{L}v \, dm \,.$$

Therefore, m is a reversible distribution for the operator $\mathcal{L}$ from which it follows that m is a Dirichlet point process with parameter ν; see e.g. Theorem 8.1 of [8].

Finally, the extension of the integration by parts formula to all functions in $D(\mathcal{E})$ follows by approximation and Remark 1. ☐

Lemma 3 *Suppose that $f \in B_b(E)$, $v \in D(\mathcal{E})$, and u is an arbitrary bounded measurable function $\mathcal{M}_1$. Then*

$$\int (u \circ S_f - u) v \, dm = -\int_0^1 \int u \circ S_{tf} \big[(Dv, f) + v \langle Af, \cdot \rangle \big] \, dm \, dt \,.$$

Proof: By a monotone class theorem and by approximation arguments it suffices to prove the assertion for $u, v \in \mathcal{F}C_b^\infty$. Then also $u \circ S_{tf} \in \mathcal{F}C_b^\infty$. Since $S_{(s+t)f} = S_{sf} \circ S_{tf}$, it follows that

$$(D(u \circ S_t), f) = \frac{d}{ds}\bigg|_{s=t} (u \circ S_{sf}) \,.$$

Therefore,

$$\int (u \circ S_f - u) v \, dm = \int_0^1 \int \frac{d}{dt}(u \circ S_{tf}) \, v \, dm \, dt = \int_0^1 \int (D(u \circ S_t), f) \, v \, dm \, dt \,.$$

An application of the integration by parts formula of Proposition 2 thus yields the assertion. ☐

4 L^2-theory for $\mathcal{E}$ and $\mathcal{L}$

Recall that $\overline{m}$ denotes the measure $\overline{m}(dx,d\mu) = \mu(dx)\,m(d\mu)$. Here is our first main result.

Theorem 4 *Suppose that $u \in L^2(m)$ and a function $\widetilde{D}u \in L^2(\overline{m})$ are given. Then the following assertions are equivalent.*

1. *$u \in D(\mathcal{E})$ and $Du = \widetilde{D}u$.*
2. *As $t \to 0$, $\frac{1}{t}\big(u \circ S_{tf} - u\big) \to (\widetilde{D}u, f)$ in $L^2(m)$ for all $f \in B_b(E)$.*
3. *As $t \to 0$, $\frac{1}{t}\int \big(u \circ S_{tf} - u\big)v\,dm \to \int(\widetilde{D}u,f)v\,dm$ for all $v \in \mathcal{F}C_b^\infty$ and $f \in B_b(E)$.*
4. *For all $f \in B_b(E)$ and $v \in \mathcal{F}C_b^\infty$,*

$$\int(\widetilde{D}u,f)v\,dm = -\int u\big[(Dv,f) + \langle Af,\cdot\rangle\big]\,dm\,.$$

The proof of Theorem 4 will require some preparations. For instance, we will have to show that the operator $\mathcal{L}$ is essentially self-adjoint on $\mathcal{F}C_b^\infty$. In fact, we will show even a slightly more general result. It implies, in particular, the following proposition, which is of independent interest if E is endowed with an a-priori topology. It will be proved subsequently to Proposition 8 below.

Proposition 5 *Suppose that E carries a Polish topology τ such that $\mathcal{B}$ coincides with the Borel σ-field of E. Let $\mathcal{F}C_b^\infty(\tau)$ denote the subset of functions in $\mathcal{F}C_b^\infty$ of the form (1) which are based on continuous functions f_i. Then the closure of the pre-Dirichlet form $(\mathcal{E},\mathcal{F}C_b^\infty(\tau))$ coincides with $(\mathcal{E},D(\mathcal{E}))$.*

Remark 6 For the application of Theorem 4 in Section 5 below it would be convenient to relax our assumption that $(E,\mathcal{B})$ is a standard Borel space. And, indeed, this property is used neither in Handa's proof [11] of the quasi-invariance (7) of m, nor during Section 3. Since these are the only auxiliary results needed, we will prove Theorem 4 under the milder assumption that $\mathcal{B}$ is separable, i.e., there exists a countable generating sequence $B_1, B_2, \dots$ in $\mathcal{B}$. Clearly, if $(E,\mathcal{B})$ is a standard Borel space, then $\mathcal{B}$ is separable.

Now we prepare for the proof of Theorem 4. Let Δ denote a partition of E into pairwise disjoint measurable sets $B_1,\dots,B_n$, and let $\mathcal{P}_\Delta$ denote the algebra of functions generated by the mappings $\mu \mapsto \mu(B_i)$ with $i = 1,\dots,n$. Every $u \in \mathcal{P}_\Delta$ is a polynomial of the form

$$u(\mu) = \sum_{m=1}^{N} c_m \prod_{i=1}^{n} \mu(B_i)^{k_{mi}} \tag{12}$$

with $k_{mi} \geq 0$ and $c_m \in \mathbb{R}$. Clearly, $\mathcal{P}_\Delta \subset \mathcal{F}C_b^\infty$. For a polynomial $u \in \mathcal{P}_\Delta$ we introduce the quantities

$$d(u) := \max_{m=1,\dots,N} \sum_{i=1}^{n} k_{mi} \qquad \text{and} \qquad c(u) := \sum_{m=1}^{N} |c_m|\,.$$

Here, we assume tacitly that the particular representation (12) of u cannot be further reduced.

Lemma 7 *The operator $\mathcal{L}$ maps $\mathcal{P}_\Delta$ into itself: $\mathcal{L}\mathcal{P}_\Delta \subset \mathcal{P}_\Delta$. Moreover, for any $u \in \mathcal{P}_\Delta$,*

$$\begin{aligned} d(\mathcal{L}u) &\leq d(u)\,, \qquad (13)\\ c(\mathcal{L}u) &\leq d(u)\big(d(u) + 2\nu(E)\big)\, c(u) \qquad (14)\end{aligned}$$

Proof: Write $u \in \mathcal{P}_\Delta$ as $u = \sum_{m=1}^N c_m u_m$ with $u_m(\mu) = \prod_{i=1}^n \mu(B_i)^{k_{mi}}$. For $i = 1, \dots, n$ define $u_{m,i}$ by

$$u_{m,i}(\mu) := \begin{cases} u_m(\mu)\mu(B_i)^{-1} & \text{if } k_{mi} \geq 1, \\ 0 & \text{otherwise.}\end{cases}$$

A straightforward calculation shows that

$$\begin{aligned} \frac{1}{2}\langle D^2 u(\mu), \mu\rangle &= -\frac{1}{2}\sum_{m=1}^N c_m \sum_{i,j=1}^n k_{mi}(k_{mj} - \delta_{ij})\big[u_m(\mu) - \delta_{ij}u_{m,i}(\mu)\big] \\ \langle ADu(\mu), \mu\rangle &= -\sum_{m=1}^N c_m \sum_{i=1}^n k_{mi}\big[\nu(E)u_m(\mu) - \nu(B_i)u_{m,i}(\mu)\big]\,, \end{aligned}$$

where δ_{ij} is Kronecker's symbol. Since $u_{m,i} \in \mathcal{P}_\Delta$ and $d(u_{m,i}) \leq d(u_m)$, it follows that $\mathcal{L}u \in \mathcal{P}_\Delta$ and $d(\mathcal{L}u) \leq d(u)$. Moreover,

$$\begin{aligned} c(\mathcal{L}u) &\leq \sum_{m=1}^N |c_m|\Big[\frac{1}{2}\sum_{i,j=1}^n k_{mi}(k_{mj} - \delta_{ij}) + \sum_{i=1}^n k_{mi}\big[\nu(E) + \nu(B_i) + k_{mi} - 1\big]\Big] \\ &\leq \sum_{m=1}^N |c_m|\big[d(u)^2 + 2\nu(E)d(u)\big]\,. \end{aligned}$$

This completes the proof of the lemma. ☐

Now let $\mathbb{D} := \{\Delta_1, \Delta_2, \dots\}$ be a sequence of finite partitions of E into measurable sets such that Δ_{n+1} is finer than Δ_n, for all n. Then $\mathcal{P}(\mathbb{D})$ defined by $\mathcal{P}(\mathbb{D}) := \bigcup_n \mathcal{P}_{\Delta_n}$ is an algebra of functions in $\mathcal{F}C_b^\infty$.

Proposition 8 *Suppose that $\mathbb{D} := \{\Delta_1, \Delta_2, \dots\}$ satisfies $\mathcal{B} = \sigma\big(\bigcup_n \Delta_n\big)$. Then $(\mathcal{L}, \mathcal{P}(\mathbb{D}))$ is an essentially self-adjoint operator on $L^2(m)$. In particular, $(\mathcal{L}, \mathcal{F}C_b^\infty)$ is essentially self-adjoint.*

Proof: It follows from Lemma 7 that $\mathcal{L}\mathcal{P}(\mathbb{D}) \subset \mathcal{P}(\mathbb{D})$ and that, for every $u \in \mathcal{P}(\mathbb{D})$, $d(\mathcal{L}^n u) \leq d(u)$ and

$$c(\mathcal{L}^n u) \leq d(u)\big(d(u) + 2\nu(E)\big)c(\mathcal{L}^{n-1}u) \leq \big[d(u)\big(d(u) + 2\nu(E)\big)\big]^n c(u)\,.$$

Since $\|\mathcal{L}^n u\|_{L^2(m)} \leq \|\mathcal{L}^n u\|_{L^\infty(m)} \leq c(\mathcal{L}^n u)$, it follows that every $u \in \mathcal{P}(\mathbb{D})$ is analytic in the sense that

$$\sum_{n=0}^{\infty} \frac{\|\mathcal{L}^n u\|_{L^2(m)}}{n!} < \infty\,.$$

Moreover, $\mathcal{P}(\mathbb{D})$ is dense in $L^2(m)$ because it is an algebra of bounded functions containing the constants and generating the canonical σ-algebra $\mathcal{F}$ on $\mathcal{M}_1(E)$. Thus, essential self-adjointness of $(\mathcal{L}, \mathcal{P}(\mathbb{D}))$ follows from Nelson's analytic vector theorem; see Corollary 2 of Section X.6 in [20]. Finally, since $(E, \mathcal{B})$ is a separable measurable space, we can always find some $\mathbb{D}$ satisfying our assumptions: just take Δ_n as the set of atoms of $\sigma(B_1, \dots, B_n)$ if $B_1, B_2, \dots$ is sequence of sets generating $\mathcal{B}$. ☐

Remark 9 As an alternative to the above proof, one can used the fact that the semigroup generated by $\mathcal{L}$ leaves $\mathcal{P}(\mathbb{D})$ invariant. Then one can apply Theorem X.49 of [20] instead of Nelson's analytic vector theorem. The writer thanks M. Röckner for pointing out this fact.

Proof of Proposition 5: As above, we can find an increasing sequence $\mathbb{D} = \{\Delta_1, \Delta_2, \dots\}$ consisting of partitions of E satisfying the assumptions of Proposition 8. In addition, we may demand that ν does not charge the boundary ∂B of any set $B \in \bigcup_n \Delta_n$, i.e., $\nu(\partial B) = 0$. Under this condition, the indicator I_B can ν-a.e. be approximated by continuous functions. Now recall that

$$\int \mu(B)\, m(d\mu) = \frac{\Gamma(\nu(E))\Gamma(\nu(B)+1)}{\Gamma(\nu(B))\Gamma(\nu(E)+1)} = \frac{\nu(B)}{\nu(E)}\,. \tag{15}$$

Thus it is easy to see that any function in $\mathcal{P}(\mathbb{D})$ can be approximated in $D(\mathcal{E})$ by functions in $\mathcal{F}C_b^\infty(\tau)$. Finally, Proposition 8 implies that $\mathcal{P}(\mathbb{D})$ is dense in $D(\mathcal{E})$. ☐

Lemma 10 *Let $f \in B_b(E)$ be given. Then S_f is a continuous linear operator from $L^2(m)$ to $L^2(m)$ and from $D(\mathcal{E})$ to $D(\mathcal{E})$. Moreover, if $u \in D(\mathcal{E})$ and $u_f := u \circ S_f$, then*

$$\begin{aligned} Du_f(\mu, x) &= Du(S_f\mu, x)\frac{e^{f(x)}}{\langle e^f, \mu\rangle} \\ \mathcal{E}(u_f, u_f) &\leq \exp\big[(2 + 2\nu(E))\|f\|_\infty\big]\mathcal{E}(u,u) \end{aligned}$$

with $\|f\|_\infty$ denoting the usual supremum norm: $\|f\|_\infty := \sup_{x\in E}|f(x)|$.

Proof: That S_f is continuous on $L^2(m)$ is obvious from (7). Moreover, S_f maps $\mathcal{F}C_b^\infty$ to $\mathcal{F}C_b^\infty$ so that it suffices to prove the two remaining assertions for $u \in \mathcal{F}C_b^\infty$. In the latter case, the formula for Du_f follows from a straightforward calculation. In order to obtain the estimate for $\mathcal{E}(u_f, u_f)$, note that

$$\begin{aligned}\mathcal{E}(u_f, u_f) &= \int\int Du(S_f\mu, x)^2 \frac{e^{2f(x)}}{\langle e^f, \mu\rangle^2}\, \mu(dx)\, m(d\mu) \\ &= \int\int Du(\mu, x)^2 e^{f(x)} \langle e^{-f}, \mu\rangle\, \mu(dx)\, (m \circ S_f^{-1})(d\mu)\,,\end{aligned}$$

and plug in formula (7). $\square$

Lemma 11 *Suppose that $u \in L^2(m)$ and $f \in B_b(E)$ are given. Then $t \mapsto u \circ S_{tf} =: u_{tf}$ is a continuous curve in $L^2(m)$. If $u \in D(\mathcal{E})$, then $\mapsto u_{tf}$ is a continuously differentiable curve in $L^2(m)$ in the sense that it can be written as a vector valued Riemann integral:*

$$u_{tf} - u = \int_0^t (Du, f) \circ S_{sf}\, ds = \int_0^t (Du_{sf}, f)\, ds$$

with a continuous integrand.

Proof: First note that the form (7) of the density of $m \circ S_{tf}$ with respect to m implies that $t \mapsto \|u_{tf}\|_{L^2(m)}$ is continuous. Next, it follows from Lemma 3 that

$$\int (u_{tf} - u)v\, dm = -\int_0^t \int u_{sf}\big[(Dv, f) + v\langle Af, \cdot\rangle\big]\, dm\, ds \to 0$$

as $t \to 0$ for all $v \in D(\mathcal{E})$. Since $\|u_{tf}\|_{L^2(m)}$ is locally bounded in t and $D(\mathcal{E})$ is dense in $L^2(m)$, we obtain that $u_{tf} \rightharpoonup u$ weakly in $L^2(m)$ as $t \to 0$. Strong continuity thus follows from the continuity of the norm.

Now suppose that $u \in D(\mathcal{E})$. Then $(Du_{tf}(\mu), f)_\mu = (Du, f) \circ S_{tf}\mu$, which as an L^2-random variable is continuous in t by the first part of this proof. Hence, the two Riemann integrals in the assertion exist and coincide; see Theorem 3.3.2 of [12]. Integrating by parts twice yields that

$$\int (u_{tf} - u)v\, dm = \int v \int_0^t (Du_{sf}, f)\, ds\, dm$$

for all $v \in D(\mathcal{E})$. This concludes the proof. $\square$

Proof of Theorem 4: First note that "1 ⇒ 2" follows from Lemma 11, and that "2 ⇒ 3" is trivial. As for the implication "3 ⇒ 4", note that it follows from Lemma 3 that

$$\int (\widetilde{D}u, f)v\,dm = -\lim_{t\to 0}\frac{1}{t}\int_0^t \int u\circ S_{sf}\big[(Dv,f) + v\langle Af,\cdot\rangle\big]\,dm\,.$$

But this implies Assertion 4, because $s \mapsto u \circ S_{sf}$ is a continuous mapping into $L^2(m)$ by Lemma 11.

Finally, we show "4 $\iff$ 1". To this end, let $\mathcal{V}$ denote the linear hull of all "smooth vector fields" of the form $v(\mu)f(x)$ with $v \in \mathcal{F}C_b^\infty$ and $f \in B_b(E)$. For $V = \sum_{i=1}^n v_i f_i \in \mathcal{V}$ define its divergence by

$$\mathrm{div}V(\mu) := \sum_{i=1}^n \Big[(Dv_i(\mu), f_i)_\mu + v_i(\mu)\langle Af_i, \mu\rangle\Big]$$

Let $T\mathcal{M}_1$ denote the tangent bundle $\bigcup_\mu T_\mu\mathcal{M}_1$. Then $(\mathrm{div}, \mathcal{V})$ is a densely defined operator from $L^2(\mathcal{M}_1 \to T\mathcal{M}_1, m)$ to $L^2(m)$. Thus its adjoint $(d, W^{1,2})$ is closed, and so is the bilinear form $\widehat{\mathcal{E}}(w,w) := \int (dw, dw)\,dm$, $w \in W^{1,2}$. Clearly $\mathcal{F}C_b^\infty \subset W^{1,2}$ and $d = D$ on $\mathcal{F}C_b^\infty$, because for $w \in \mathcal{F}C_b^\infty$

$$\int (Dw, V)\,dm = -\int w\,\mathrm{div}V\,dm\,, \qquad \text{for all } V \in \mathcal{V}$$

by Proposition 2. Therefore the closed form $(\widehat{\mathcal{E}}, W^{1,2})$ extends $(\mathcal{E}, D(\mathcal{E}))$, which, in view of Proposition 8, can only be true if the two forms coincide. ☐

5 Projecting onto smaller spaces

In this section, we consider the situation where we are given a sub-σ-field $\widetilde{\mathcal{B}}$ of $\mathcal{B}$. As explained in Remark 6 it is not necessary to require that $(E, \widetilde{\mathcal{B}})$ is itself again a standard Borel space. We assume only that $\widetilde{\mathcal{B}}$ is separable. The set $\widetilde{\mathcal{M}}_1 = \mathcal{M}_1(E, \widetilde{\mathcal{B}})$ will be identified with the larger space $\mathcal{M}_1 = \mathcal{M}_1(E, \mathcal{B})$ endowed with the σ-field $\widetilde{\mathcal{F}}$ which is generated by $\mu \mapsto \mu(B)$ for $B \in \widetilde{\mathcal{B}}$. Let us use the tilde to denote objects based on $(E, \widetilde{\mathcal{B}})$ and $(\mathcal{M}_1, \widetilde{\mathcal{F}})$ instead of $(E, \mathcal{B})$ and $(\mathcal{M}_1, \mathcal{F})$. For instance, $\widetilde{\mathcal{F}C}_b^\infty$ denotes the corresponding set of "smooth" functions based on bounded $\widetilde{\mathcal{B}}$-measurable f_i's. Clearly, $\widetilde{\mathcal{F}C}_b^\infty \subset \mathcal{F}C_b^\infty$, and $\widetilde{D}$ coincides with D on $\widetilde{\mathcal{F}C}_b^\infty$. Hence $(\mathcal{E}, D(\mathcal{E}))$ extends the closure $(\widetilde{\mathcal{E}}, D(\widetilde{\mathcal{E}}))$ of the form $(\widetilde{\mathcal{E}}, \widetilde{\mathcal{F}C}_b^\infty)$. In particular, $D(\widetilde{\mathcal{E}}) \subset D(\mathcal{E})$, and $\widetilde{D} = D$ on $D(\widetilde{\mathcal{E}})$.

Now take an arbitrary function $u \in D(\mathcal{E})$ and let $\widetilde{u} := E_m[\,u \,|\, \widetilde{\mathcal{F}}\,]$ denote the expectation of u conditional on $\widetilde{\mathcal{F}}$ and with respect to the measure m. Note that for a bounded $\widetilde{\mathcal{B}}$-measurable function f,

$$\widetilde{u}\circ S_f = E_m[\,u\circ S_f\,|\widetilde{\mathcal{F}}\,]\,. \tag{16}$$

Indeed, if v is bounded and $\widetilde{\mathcal{F}}$-measurable, then

$$\int (u \circ S_f)\, v\, dm = \int u\, (v \circ S_{-f}) \frac{dm \circ S_f^{-1}}{dm}\, dm = \int (\widetilde{u} \circ S_f)\, v\, dm\,.$$

In view of (16), Theorem 4 implies that for $v \in \widetilde{\mathcal{F}}C_b^\infty$ and a bounded $\widetilde{\mathcal{B}}$-measurable function f

$$\frac{1}{t}\int (\widetilde{u} \circ S_{tf} - \widetilde{u})\, v\, dm = \frac{1}{t}\int (u \circ S_{tf} - u)\, v\, dm \to \int Du(\mu, x)\, f(x)\, v(\mu)\, \overline{m}(dx, d\mu),$$

as $t \to 0$. Thus the following result is implied by Theorem 4 together with the fact that $D(\widetilde{\mathcal{E}}) \subset D(\mathcal{E})$ and $\widetilde{D}\widetilde{u} = D\widetilde{u}$.

Corollary 12 *If $u \in D(\mathcal{E})$, then also $\widetilde{u} := E_m[\, u \,|\, \widetilde{\mathcal{F}}\,] \in D(\mathcal{E})$, and $D\widetilde{u}$ is obtained by taking the conditional expectation of Du with respect to the product σ-field $\widetilde{\mathcal{F}} \otimes \widetilde{\mathcal{B}}$ under the measure $\overline{m}(dx, d\mu) = \mu(dx)\, m(d\mu)$.*

Next, we consider the particular situation where $\widetilde{\mathcal{B}}$ is generated by a finite partition $\Delta = \{A_1, \dots, A_n\}$ of E into measurable sets. Then it is not difficult to see that the projected situation is isomorphic to considering the finite-dimensional model with $E_\Delta := \{1, \dots, n\}$ and $\nu_\Delta := \sum_{i=1}^n \nu(A_i)\delta_i$. Subsequently, we will use the subscript Δ to indicate the quantities of this finite-dimensional model. In particular, any function $u \in L^1(m)$ induces a function $u_\Delta \in L^1(m_\Delta)$ by the relation

$$u_\Delta(\pi_\Delta \mu) = E_m[\, u \,|\, \widetilde{\mathcal{F}}\,](\mu)\,, \qquad \text{where } \pi_\Delta \mu := \sum_{i=1}^n \mu(A_i)\delta_i.$$

By the above, we obtain the following result.

Corollary 13 *If $u \in D(\mathcal{E})$, then $u_\Delta \in D(\mathcal{E}_\Delta)$ and*

$$(D_\Delta u_\Delta, D_\Delta u_\Delta) \circ \pi_\Delta \le E_m\big[\, (Du, Du) \,\big|\, \widetilde{\mathcal{F}}\,\big] \qquad m\text{-a.e.}$$

6 Exponential families, Fisher information, and the Bhattacharya distance

Due to the quasi-invariance of m under exponential shifts, a prominent role is played in our analysis by the exponential families $t \mapsto S_{tf}\mu$. In this section, we start investigating the geometric properties of these curves when $\mathcal{M}_1$ is endowed with the Bhattacharya distance ρ, the intrinsic metric of the Dirichlet form $(\mathcal{E}, D(\mathcal{E}))$. The topology generated by ρ on $\mathcal{M}_1$ is equivalent to the one generated by the total variation distance $\|\cdot\|_{\text{var}}$. This follows from the estimates

$$\|\mu - \lambda\|_{\text{var}} \le \rho(\mu, \lambda) \le \frac{\pi}{\sqrt{2}}\|\mu - \lambda\|_{\text{var}}^{1/2}\,; \tag{17}$$

cf. Equation (2.15) of [23].

Let ω be a curve from $[a,b]$ into $\mathcal{M}_1$. Its *energy* with respect to ρ is defined as

$$\mathbf{E}_{a,b}(\omega) = \sup\Big\{\frac{1}{2}\sum_{i=1}^{n}\frac{\rho(\omega(t_i),\omega(t_{i-1}))^2}{t_i - t_{i-1}}\ \Big|\ a \le t_0 < t_1 < \cdots < t_n \le b,\ n \in \mathbb{N}\Big\},$$

and the *arc length* of ω is given by

$$\mathbf{L}_{a,b}(\omega) = \sup\Big\{\sum_{i=1}^{n}\rho(\omega(t_i),\omega(t_{i-1}))\ \Big|\ a \le t_0 < t_1 < \cdots < t_n \le b,\ n \in \mathbb{N}\Big\}.$$

It is proved in [23] that $\mathbf{E}_{a,b}(\omega)$ admits the following representation formula as integral of the Fisher information along the curve ω:

$$\mathbf{E}_{a,b}(\omega) = \begin{cases} \frac{1}{2}\int_a^b \left\|\frac{d\dot\omega(t)}{d\omega(t)}\right\|_{\omega(t)}^2 dt & \text{if } \omega \in \mathbf{H}_{a,b}, \\ \infty & \text{otherwise,}\end{cases} \tag{18}$$

where $\mathbf{H}_{a,b}$ is the space of those $\eta : [a,b] \longrightarrow \mathcal{M}_1$ that take of the form of a Bochner integral

$$\eta(t) = \eta(0) + \int_0^t \dot\eta(s)\,ds\,, \tag{19}$$

for some finite signed measures $\dot\eta(s)$ being absolutely continuous with respect to $\eta(s)$, for almost every s, and whose Radon-Nikodym derivative $d\dot\eta(s)/d\eta(s)$ satisfies

$$\int_a^b \left\|\frac{d\dot\eta(t)}{d\eta(t)}\right\|_{\eta(t)}^2 dt < \infty\,.$$

Moreover, it was shown in [23] that for $\eta \in \mathbf{H}_{a,b}$

$$\mathbf{L}_{a,b}(\omega) = \int_a^b \left\|\frac{d\dot\omega(t)}{d\omega(t)}\right\|_{\omega(t)} dt\,. \tag{20}$$

Lemma 14 *Suppose that $f \in B_b(E)$ and $\mu \in \mathcal{M}_1(E)$. Then the curve $\omega(t) := S_{tf}\mu$ has the length*

$$\mathbf{L}_{a,b}(\omega) = \int_a^b \|f\|_{\omega(t)}\,dt \tag{21}$$

and possesses the energy

$$\mathbf{E}_{a,b}(\omega) = \frac{1}{2}\int_a^b \|f\|_{\omega(t)}^2\,dt = \frac{1}{2}\Big(\langle f, S_{bf}\mu\rangle - \langle f, S_{af}\mu\rangle\Big)\,.$$

Proof: It suffices to note that $\omega(t) = S_{tf}\mu$ is indeed of the form (19) with

$$d\dot{\omega}(t) = \big(f - \langle f, \omega(t)\rangle\big)\, d\omega(t)$$

in order to obtain the formula for the length and the first identity for the energy. Finally, note that

$$\|f\|^2_{\omega(t)} = \frac{d}{dt}\langle f, S_{tf}\mu\rangle\,. \tag{22}$$

Thus, the assertion follows. $\square$

Remark 15 1. Recall from Theorem 4 that for $u \in D(\mathcal{E})$ and $f \in B_b(E)$

$$\frac{d}{dt}\Big|_{t=0} u \circ S_{tf} = (Du, f)\,.$$

This shows that the arc length (21) of an exponential family for f relates to the concept of differentiation given by D.

2. One can deduce from the preceding lemma that

$$\rho(S_g\mu, S_f\mu) \leq 2\|f - g\|_\infty\,,$$

uniformly in μ and for all f, $g \in B_b(E)$. To prove this estimate, one connects $S_f\mu$ and $S_g\mu$ by the curve $\omega(t) := S_{t(g-f)}S_f\mu$, for which $\mathbf{L}_{0,1}(\omega) \leq 2\|f - g\|_\infty$.

Lemma 16 *Suppose that $\omega \in \mathbf{H}_{a,b}$ and u is a ρ-Lipschitz continuous function on $\mathcal{M}_1$. Then $t \mapsto u(\omega(t))$ is absolutely continuous and*

$$\Big|\frac{d}{dt}u(\omega(t))\Big| \leq \mathrm{Lip}(u) \cdot \Big\|\frac{d\dot{\omega}(t)}{d\omega(t)}\Big\|_{\omega(t)} \qquad \textit{for almost every } t.$$

In particular, if $\omega(t) = S_{tf}\mu$ is as in Lemma 14, then

$$\Big|\frac{d}{dt}u(\omega(t))\Big| \leq \mathrm{Lip}(u) \cdot \|f\|_{\omega(t)} \qquad \textit{for almost every } t.$$

Proof: If $\{t_0, t_1, \ldots, t_n\}$ is any ordered partition of some finite interval $[a, b]$, then

$$\sum_{i=1}^n \frac{\big(u(\omega(t_i)) - u(\omega(t_{i-1}))\big)^2}{t_i - t_{i-1}} \leq 2\mathrm{Lip}(u)^2\mathbf{E}_{a,b}(\omega) = \mathrm{Lip}(u)^2 \int_a^b \Big\|\frac{d\dot{\omega}(t)}{d\omega(t)}\Big\|^2_{\omega(t)}\, dt\,.$$

The lemma in Chapter II, No. 36 of [21] states that this condition implies that $t \mapsto u(\omega(t))$ is absolutely continuous and that

$$\int_a^b \Big(\frac{d}{dt}u(\omega(t))\Big)^2\, dt \leq \mathrm{Lip}(u)^2 \int_a^b \Big\|\frac{d\dot{\omega}(t)}{d\omega(t)}\Big\|^2_{\omega(t)}\, dt\,.$$

Since a and b were arbitrary, the lemma follows. $\square$

Next we want to compare the energy $\mathbf{E}_{0,t}(\omega) = \frac{1}{2}(\langle f, S_{tf}\mu\rangle - \langle f, \mu\rangle)$ of the curve $\omega(t) = S_{tf}\mu$ to the the expression $\frac{1}{2t}\rho(\mu, S_{tf}\mu)^2$, which is in fact the energy of an energy minimizing geodesic from μ to $S_{tf}\mu$; see [23]. To this end, it will be convenient to write $\rho(\mu, S_{tf}\mu) = 2\arccos q(t)$, where

$$q(t) := \frac{\langle e^{tf/2}, \mu\rangle}{\sqrt{\langle e^{tf}, \mu\rangle}} = \exp\Big[-\frac{1}{2}\int_0^t \langle f, S_{sf}\mu\rangle - \langle f, S_{sf/2}\mu\rangle\, ds\Big].$$

Here, the second identity follows from the fact that $\langle f, S_{tf}\mu\rangle = \frac{d}{dt}\log\langle e^{tf}, \mu\rangle$.

Lemma 17 *Let ω be a curve of the form $\omega(t) = S_{tf}\mu$. Then*

$$0 \leq \mathbf{E}_{0,t}(\omega) - \frac{\rho(\mu, S_{tf}\mu)^2}{2t} \leq t^2 4\|f\|_\infty^3 + t^3 4\|f\|_\infty^4 .$$

Proof: The first inequality is obvious. As for the second one, let

$$d_{\mathrm{KH}}(\mu, \lambda) := 2\bigg(\int \bigg(\sqrt{\frac{d\mu}{d\eta}} - \sqrt{\frac{d\lambda}{d\eta}}\bigg)^2 d\eta\bigg)^{1/2} \qquad \text{for } \eta \gg \mu,\, \lambda$$

denote the Kakutani-Hellinger distance of μ and λ (times a multiplicative factor). Equation (2.9) of [23] states that $\rho(\mu, \lambda) \geq d_{\mathrm{KH}}(\mu, \lambda)$. Therefore

$$\rho(\mu, S_{tf}\mu)^2 \geq d_{\mathrm{KH}}(\mu, S_{tf}\mu)^2 = 8 - 8q(t) \geq 4q(t)\int_0^t \langle f, S_{sf}\mu\rangle - \langle f, S_{sf/2}\mu\rangle\, ds .$$

With the notation $g(t) := \langle f, S_{tf}\mu\rangle - \langle f, \mu\rangle$ it thus follows from Lemma 14 that

$$\delta := \mathbf{E}_{0,t}(\omega) - \frac{\rho(\mu, S_{tf}\mu)^2}{2t} \leq \frac{1}{t}\Big[\frac{t}{2}g(t) - 2q(t)\int_0^t g(s) - g(s/2)\, ds\Big].$$

Applying the mean value theorem to the term in parentheses, we obtain the existence of some $\xi \in [0, t]$ such that

$$\delta \leq \frac{1}{2}g(\xi) + \frac{\xi}{2}g'(\xi) - 2q(\xi)\big[g(\xi) - g(\xi/2)\big] - q(\xi)\big[g(\xi) - g(\xi/2)\big]\int_0^\xi g(s) - g(s/2)\, ds .$$

The identity (22) implies that g is increasing. Hence, we may drop the last term above. Using that $1 - q(\xi) \leq \frac{1}{2}\int_0^\xi g(s) - g(s/2)\, ds$, it follows that

$$\delta \leq \frac{\xi}{2}\big[g'(\xi_1) + g'(\xi)\big] - \xi g'(\xi_2) + \frac{\xi^3}{2}g'(\xi_1)g'(\xi_2),$$

for certain $\xi_1 \in [0, \xi]$ and $\xi_2 \in [\xi/2, \xi]$. Thus

$$\delta \leq \frac{t^2}{2}\sup_{0\leq s\leq t}|g''(s)| + \frac{t^3}{2}\sup_{0\leq s\leq t} g'(s)^2,$$

from which an easy calculation concludes the proof. ☐

The following proposition states that energy-minimizing ρ-geodesics can be approximated by curves which are piecewise exponential families.

Proposition 18 *Suppose that $\mu, \lambda \in \mathcal{M}_1$ satisfy $\lambda \ll \mu$, and $\varepsilon > 0$ is given. Then there are $f_1, \ldots, f_n \in B_b(E)$ such that the curve $\omega(t)$ defined for $0 \le t \le 1$ by*

$$\omega(t) := S_{(t-\frac{k}{n})f_{k+1}} S_{\frac{1}{n}f_k} \cdots S_{\frac{1}{n}f_1}\mu \qquad \text{for } \frac{k}{n} \le t \le \frac{k+1}{n} \tag{23}$$

satisfies $\rho(\omega(1), \lambda) < \varepsilon$ and $\mathbf{E}_{0,1}(\omega) \le \frac{1}{2}\rho(\mu,\lambda)^2 + \varepsilon$. Moreover, the same is true if we replace the latter condition by $\mathbf{L}_{0,1}(\omega) \le \rho(\mu,\lambda) + \varepsilon$.

Proof: First note that we may assume without loss of generality that $d\lambda/d\mu$ is bounded above since we can approximate λ in ρ-distance by measures which do have a bounded Radon-Nikodym derivative with respect to μ. Next, let $r := \rho(\mu,\lambda)/2$ and define a curve $\gamma : [0,1] \to \mathcal{M}_1$ by

$$\frac{d\gamma(t)}{d\mu} := \left(\cos rt - \cot r \, \sin rt + \frac{\sin rt}{\sin r} \cdot \sqrt{\frac{d\lambda}{d\mu}}\right)^2 .$$

Then γ is the unique energy-minimizing geodesic from $\mu = \gamma(0)$ to $\lambda = \gamma(1)$ in the sense that $\rho(\gamma(s), \gamma(t)) = |t-s|\rho(\mu,\lambda)$ for s, $t \in [0,1]$; see [23]. Consequently, $\mathbf{E}_{0,1}(\gamma) = \rho(\mu,\lambda)^2/2$. Moreover, we can assume without loss of generality that $d\lambda/d\mu$ is bounded away from 0, because otherwise we can replace λ by $\gamma(t)$ for t close to 1.

There exists a constant c such that

$$-c|t-s| \le \log \frac{d\gamma(t)}{d\gamma(s)} \le c|t-s| \, .$$

So, if we let

$$f_i := n \log \frac{d\gamma(\frac{i}{n})}{d\gamma(\frac{i-1}{n})} ,$$

then $\|f_i\|_\infty \le c$ and $\gamma(\frac{i}{n}) = S_{\frac{1}{n}f_i}\gamma(\frac{i-1}{n})$. Moreover, it follows from Lemma 17 that, for ω defined as in the assertion with this choice of $f_1, \ldots, f_n$,

$$\mathbf{E}_{0,1}(\omega) \le \frac{\rho(\mu,\lambda)^2}{2} + \frac{4c^3(1+c/n)}{n} .$$

Making n large proves the first part of the proposition. In order to replace the energy functional by the arc length, just note that $\mathbf{L}_{0,1}^2 \le 2\mathbf{E}_{0,1}$ by (18) and (20). $\square$

7 A Rademacher theorem

Theorem 19 *Suppose $u \in L^2(m)$. Then the following conditions are equivalent.*

1. *$u \in D(\mathcal{E})$ with $\|Du(\mu)\|_\mu \leq L$ for m-a.e. μ.*
2. *u possesses a ρ-Lipschitz continuous m-version with Lipschitz constant at most L.*
3. *There exists a measurable subset Ω^* of $\mathcal{M}_1$ having full m-measure, an m-version $\widetilde{u}$ of u, and a measurable function $\widetilde{D}u : \Omega^* \times E \to \mathbb{R}$ such that the following holds.*
    - *For all $\mu \in \Omega^*$, $\|\widetilde{D}u(\mu)\|_\mu \leq L$.*
    - *If $f \in B_b(E)$ then*

      $$\frac{\widetilde{u}(S_{tf}\mu) - \widetilde{u}(\mu)}{t} \longrightarrow (\widetilde{D}u(\mu), f)_\mu \qquad \text{as } t \to 0,$$

      for all $\mu \in \Omega^*$ and in $L^2(m)$.

    In fact, one can take for $\widetilde{u}$ any ρ-Lipschitz continuous m-version of u, and the function $\widetilde{D}u$ is a version of the gradient of u.

Proof of Theorem 19: Condition 3 implies 1 by Theorem 4, so let us start by showing that 3 follows from 2. Let $u \in L^2(m)$ itself be ρ-Lipschitz continuous with $L = \mathrm{Lip}(u)$. For $f \in B_b(E)$ consider the set

$$\Omega_f := \left\{ \mu \in \mathcal{M}_1 \;\middle|\; \frac{u(S_{tf}\mu) - u(\mu)}{t} \to G_f(\mu) \text{ as } t \to 0, \text{ and } |G_f(\mu)| \leq L \cdot \|f\|_\mu \right\}.$$

Then Ω_f is measurable, because $t \mapsto u(S_{tf}\mu)$ is Lipschitz continuous by Lemma 16, and so the existence of $\lim_{t\to 0} \frac{1}{t}\big(u(S_{tf}\mu) - u(\mu)\big)$ is equivalent to the existence of the limit of $\frac{1}{r}\big(u(S_{rf}\mu) - u(\mu)\big)$ for $r \to 0$, $r \in \mathbb{Q}$. Moreover, we claim that Ω_f has full m-measure. Indeed, if μ is fixed, then it follows from Lemma 16 that the set of all $s \in \mathbb{R}$ for which $S_{sf}\mu \in \Omega_f$ has full Lebesgue measure. Thus

$$0 = \int \int_0^1 \mathrm{I}_{\Omega_f^C}(S_{sf}\mu)\, ds\, m(d\mu) = \int_0^1 \int_{\Omega_f^C} \frac{dm \circ S_{sf}^{-1}}{dm}\, dm\, ds\,.$$

But $dm \circ S_{sf}^{-1}/dm$ is m-a.s. strictly positive, and hence $m(\Omega_f^C) = 0$.

For the next step, note first that it follows from Lemma 16 that

$$\sup_{-1\leq t\leq 1} \left|\frac{u(S_{tf}\mu) - u(\mu)}{t}\right| \leq \sup_{-1\leq t\leq 1} \frac{L}{t} \int_0^t \|f\|_{S_{sf}\mu}\, ds \leq 2L \cdot \|f\|_\infty\,,$$

where $\|\cdot\|_\infty$ is the usual supremum norm on $B_b(E)$. Thus, dominated convergence implies that, for all $v \in \mathcal{F}C_b^\infty$,

$$\int v\, G_f\, dm = \lim_{t\to 0} \frac{1}{t} \int v\,(u \circ S_{tf} - u)\, dm\,.$$

Applying Lemma 3 to the integral on the right hand side and using the continuity of $t \mapsto u(S_{tf}\mu)$ yields the identity

$$\int v\, G_f\, dm = -\int \big[(Dv, f) + \langle Af, \cdot\rangle\big]\, u\, dm\,.$$

So, if f can be written as $\alpha_1 f_1 + \cdots + \alpha_k f_k$ with $\alpha_i \in \mathbb{R}$ and $f_i \in B_b(E)$, then

$$\int v\, G_f\, dm = -\sum_{i=1}^{k} \alpha_i \int \big[(Dv, f_i) + \langle Af_i, \cdot\rangle\big]\, u\, dm = \int v \sum_{i=1}^{k} \alpha_i G_{f_i}\, dm\,.$$

Since $v \in \mathcal{F}C_b^\infty$ was arbitrary, it follows that

$$G_f = \sum_{i=1}^{k} \alpha_i G_{f_i} \qquad m\text{-a.s.}$$

Next, endow E for the moment with a compact metric topology and take some countable $\mathbb{Q}$-vector space $\mathbb{F}$ which is dense in $C(E)$ with respect to uniform convergence and which is stable under the mappings $f \mapsto (-n) \vee f \wedge n$ for $n \in \mathbb{N}$. In particular, $\mathbb{F}$ will be dense in $L^2(\mu)$ for any measure $\mu \in \mathcal{M}_1$. Take Ω^* to be the set of all $\mu \in \bigcap_{f\in\mathbb{F}} \Omega_f$ for which $f \mapsto G_f(\mu)$ is a $\mathbb{Q}$-linear mapping on $\mathbb{F}$. Clearly $m(\Omega^*) = 1$. Since $|G_f(\mu)| \leq L \cdot \|f\|_\mu$, we can extend $G_\cdot(\mu)$ to a continuous linear functional (again denoted by $G_\cdot(\mu)$) on the Hilbert space $T_\mu\mathcal{M}_1$. Therefore, we obtain the existence of a function $\widetilde{D}u(\mu) \in T_\mu\mathcal{M}_1$ such that $G_f(\mu) = (\widetilde{D}u(\mu), f)_\mu$ and $\|\widetilde{D}u(\mu)\|_\mu \leq L$. This yields the first part of assertion 3.

The second part of 3 is already clear if $f \in \mathbb{F}$. To pass to the general case, take f, $g \in B_b(E)$ and consider the curve $\omega(s) := S_{s(g-f)}S_{tf}\mu$. Then $\omega(0) = S_{tf}\mu$ and $\omega(t) = S_{tg}\mu$. Therefore, Lemma 16 yields that

$$|u(S_{tf}\mu) - u(S_{tg}\mu)| \;\; \leq \;\; \operatorname{sgn}(t) L \int_0^t \|g - f\|_{\omega(s)}\, ds\,.$$

Thus, if $g \in \mathbb{F}$, then

$$\begin{aligned} &\Big|\frac{u(S_{tf}\mu) - u(\mu)}{t} - (\widetilde{D}u(\mu), f)_\mu\Big| \leq \\ &\leq \Big|\frac{u(S_{tg}\mu) - u(\mu)}{t} - (\widetilde{D}u(\mu), g)_\mu\Big| + \frac{L}{t}\int_0^t \|g - f\|_{\omega(s)}\, ds + \|\widetilde{D}u(\mu)\|_\mu \|g - f\|_\mu\,. \end{aligned}$$

Since $g \in \mathbb{F}$, the first term on the right hand side tends to 0 as $t \to 0$. As for the integral term, note first that

$$\frac{d\omega(s)}{d\mu} = \frac{e^{s(g-f)+tf}}{\langle e^{s(g-f)+tf}, \mu\rangle} \leq \exp\Big[2|s| \cdot \|g\|_\infty + 2|t - s| \cdot \|f\|_\infty\Big]\,.$$

Since our assumptions on $\mathbb{F}$ allow for taking g such that $\|g\|_\infty \leq 1 + \|f\|_\infty$, we can assume that there exists a constant c depending only on f such that $d\omega(s)/d\mu \leq c$, uniformly in $|s| \leq |t| \leq 1$. It follows that, for $t \to 0$,

$$\left| \frac{u(S_{tf}\mu) - u(\mu)}{t} - (\widetilde{D}u(\mu), f)_\mu \right| \leq o(1) + (cL + L)\|g - f\|_\mu ,$$

which can be made arbitrarily small by making $\|g - f\|_\mu$ small. This proves assertion 3 of Theorem 19.

Next, we will prove that any $u \in D(\mathcal{E})$ with $\|Du(\mu)\|_\mu \leq L$ for m-a.e. μ possesses a ρ-Lipschitz version $\hat{u}$ with $\mathrm{Lip}(\hat{u}) \leq L$. To this end, let us first consider the case in which E is finite. By (15), we may assume without loss of generality that $\nu(\{x\}) > 0$ for every $x \in E$. Then it follows from (6) that also $\mu(\{x\}) > 0$ for m-a.e. μ and each $x \in E$.

Lemma 20 *Suppose that $(v_n)_{n\in\mathbb{N}}$ is a sequence of functions converging to 0 in $L^2(m)$, and $f \in B_b(E)$ is given. Then $\int_r^t |v_n \circ S_{sf}|\,ds \to 0$ in m-probability, for all r, $t \in \mathbb{R}$.*

Proof: This follows immediately from

$$\int |v_n|^2\, dm = \frac{1}{t - r} \int \int_r^t |v_n \circ S_{sf}| \frac{dm \circ S_{-sf}^{-1}}{dm}\, ds\, dm$$

and the uniform boundedness of the occurring densities. □

Now suppose that $\widetilde{D}u$ is an $\overline{m}$-version of Du such that $\|\widetilde{D}u(\mu)\|_\mu \leq L$ for all $\mu \in \mathcal{M}_1(E)$. Choose a sequence $(u_n)_{n\in\mathbb{N}} \subset \mathcal{F}C_b^\infty$ approximating u in the Dirichlet norm, i.e., $u_n \to u$ in $L^2(m)$ and $Du_n \to \widetilde{D}u$ in $L^2(\overline{m})$. Take $f \in B_b(E)$, and apply Lemma 20 to the functions $v_n(\mu) = \|Du_n(\mu) - \widetilde{D}u(\mu)\|_\mu$. We obtain the existence of a subsequence $(u_{n_k})_{k\in\mathbb{N}}$, depending on f, such that for m-a.e. $\mu \in \mathcal{M}_1$

$$\begin{gathered} u_{n_k}(\mu) \text{ converges to a finite limit } u_\infty(\mu), \text{ and} \\ \int_r^t \|Du_{n_k}(S_{sf}\mu) - \widetilde{D}u(S_{sf}\mu)\|_{S_{sf}\mu}\, ds \longrightarrow 0 \qquad \text{for all } r \text{ and } t. \end{gathered} \tag{24}$$

By a diagonalization argument, we can find a subsequence $(u_{n_k})_{k\in\mathbb{N}}$ such that (24) holds for m-a.e. μ and simultaneously for all $f \in \mathbb{F}$, where $\mathbb{F}$ is as in the first part of proof. Denote by Ω_0 the set of all $\mu \in \mathcal{M}_1$ for which (24) is true. Clearly, u_∞ is a m-version of u on Ω_0. Suppose we had already shown that

$$u_\infty \text{ is } \rho\text{-Lipschitz continuous on } \Omega_0 \text{ with } \mathrm{Lip}(u_\infty) \leq L. \tag{25}$$

Then we could define $\hat{u}$ as the McShane extension of u_∞:

$$\hat{u}(\mu) := \sup_{\lambda\in\Omega_0} \Big[u_\infty(\lambda) - L\rho(\mu, \lambda) \Big] , \tag{26}$$

which coincides with u_∞ on Ω_0 and is ρ-Lipschitz continuous with $\mathrm{Lip}(\hat{u}) \le L$; see [16]. Thus $\hat{u}$ would be the desired ρ-Lipschitz version of u.

But it remains to prove (25). To this end, let us note first that by (24)

$$\begin{aligned} u_{n_k}(S_{tf}\mu) &= u_{n_k}(\mu) + \int_0^t (Du_{n_k}(S_{sf}\mu), f)_{S_{sf}\mu}\, ds \\ &\longrightarrow u_\infty(\mu) + \int_0^t (\widetilde{D}u(S_{sf}\mu), f)_{S_{sf}\mu}\, ds\,, \end{aligned}$$

for all $\mu \in \Omega_0$, $t \in \mathbb{R}$, and $f \in \mathbb{F}$. Therefore, it follows that for all $t \in \mathbb{R}$ and $f \in \mathbb{F}$

$$\begin{aligned} & S_{tf}\Omega_0 \subseteq \Omega_0 \quad \text{and} \\ & u_\infty(S_{tf}\mu) = u_\infty(\mu) + \int_0^t (\widetilde{D}u(S_{sf}\mu), f)_{S_{tf}\mu}\, ds \qquad \text{for } \mu \in \Omega_0. \end{aligned}$$

Since it was assumed that $\|\widetilde{D}u(\cdot)\| \le L$, it follows by Lemma 14 that

$$|u_\infty(S_{tf}\mu) - u_\infty(\mu)| \le L \int_0^t \|f\|_{S_{sf}\mu}\, ds = L \cdot \mathbf{L}_{0,t}(S_{\cdot f}\mu)\,.$$

Thus we can conclude (25) from Remark 15 and Proposition 18, and the assertion is proved in case where E is finite.

In order to treat the case of general E, we will borrow from [7] the idea of a finite-dimensional approximation, for which we have prepared in Section 5. To this end, choose an increasing sequence $\Delta_1, \Delta_2, \ldots$ of partitions of E with $\mathcal{B} = \sigma\big(\bigcup_n \Delta_n\big)$. With the notations and results of Section 5 we find that $(D_{\Delta_n}u_{\Delta_n}, D_{\Delta_n}u_{\Delta_n}) \le L\ m_{\Delta_n}$-a.e. Therefore, u_{Δ_n} possesses a modification $\hat{u}_{\Delta_n}$ which is ρ_{Δ_n}-Lipschitz continuous with $\mathrm{Lip}(\hat{u}_{\Delta_n}) \le L$. Let $u_\infty := \lim_n \hat{u}_{\Delta_n} \circ \pi_{\Delta_n}$ on the set Ω_0 where this limit exists. Clearly $m(\Omega_0) = 1$ by martingale convergence. Also, u_∞ is ρ-Lipschitz continuous on Ω_0 as $\rho_{\Delta_n}(\pi_{\Delta_n}\mu, \pi_{\Delta_n}\lambda) \to \rho(\mu, \lambda)$ for all μ and λ. So we can again use the McShane extension (26) of u_∞ to obtain the desired function $\hat{u}$. ☐

8 The Bhattacharya distance and potential theory

For the moment denote by $\langle u \rangle$ the mean $\int u\, dm$ of a function $u \in L^1(m)$. Since the metric ρ is bounded above by π, Theorem 19 yields that for $u \in D(\mathcal{E})$,

$$\big\|u - \langle u \rangle\big\|_{L^\infty(m)} \le \pi \big\|\, \|Du\|_\mu \,\big\|_{L^\infty(m)}\,. \tag{27}$$

Equation (8) implies that π is in fact the optimal constant. An L^2-version of this estimate was proved by W. Stannat [25]. In the L^2-case, the optimal constant is different from π.

At first sight, Equation (27) might suggest that we are close to a finite-dimensional situation, and that one can expect results which hold on (locally) compact spaces like Riemannian manifolds. K.-T. Sturm obtained such kind of results in the setup of regular Dirichlet forms; see e.g. [28], [27]. This analysis, however, relies strongly on the behavior of the intrinsic distance function $\rho(\mu,\cdot)$ to fixed points μ and on the growth of the volume $m(B_r^\rho(\mu))$ of ρ-balls $B_r^\rho(\mu)$ of radius r around μ. But if E is uncountable and the intensity measure ν is diffuse, we can obtain that for any μ the function $\rho(\mu,\cdot)$ is m-a.s. identical to the constant π. Hence, we find a singular volume growth behavior:

$$m(B_r^\rho(\mu)) = \begin{cases} 0 & \text{if } r < \pi, \\ 1 & \text{if } r = \pi. \end{cases}$$

Indeed, to see that $\rho(\mu,\lambda) = \pi$ for m-a.e. λ, just note that a typical λ is concentrated on an i.i.d. sample of countably many points of distribution $\nu/|\nu|$. Therefore, λ will almost surely be singular to any given measure μ. But if λ and μ are singular, then $\rho(\mu,\lambda) = \pi$.

This volume growth behavior can be seen as a truly infinite dimensional aspect of the situation encountered here, and it reflects the non-separability of $\mathcal{M}_1$ in the ρ-topology (which is same as the topology induced by the total variation distance, by (17)).

So, instead of looking at the ρ-distance to a fixed point in $\mathcal{M}_1$ we turn now to the study of the function

$$\rho_K(\mu) := \inf_{\lambda\in K} \rho(\lambda,\mu)$$

for subsets K of $\mathcal{M}_1$. For simplicity, we assume throughout this section that E is endowed with a Polish topology and that $\mathcal{M}_1$ carries the corresponding weak topology. The reader is referred to [15] for the notions associated with the capacity of a Dirichlet form.

Lemma 21

1. *There exists a countable set $\{u_1, u_2, \dots\}$ of continuous functions in $\mathcal{F}C_b^\infty$ with $(Du_k, Du_k) \le 1$ such that $\rho(\lambda,\mu) = \sup_{k\in\mathbb{N}} \big[u_k(\lambda) - u_k(\mu)\big]$ for all λ, $\mu \in \mathcal{M}_1$.*
2. *$\rho : \mathcal{M}_1 \times \mathcal{M}_1 \to [0,\pi]$ is lower semicontinuous.*
3. *If K is compact, then $\rho_K : \mathcal{M}_1 \to [0,\pi]$ is lower semicontinuous and, hence, measurable.*
4. *If E is a finite set, then ρ_K is continuous for any K.*

Proof: 1. Take a sequence $f_1, f_2, \dots$ of bounded continuous functions on E generating the Borel field $\mathcal{B}$. Let $\mathbf{D}$ denote the $\mathbb{Q}$-algebra generated by these functions and the constant 1, and suppose in addition that $\mathbf{D}$ is stable under the mapping $f \mapsto 0 \vee f \wedge 1$. Next let $\mathcal{C}_n$ denote some countable set of functions which is dense in the set of compactly supported C^1-functions on $\mathbb{R}^n$ with respect to the topology of uniform convergence on compacts. Then

$$\mathbf{F} := \bigcup_{n=1}^\infty \Big\{ F(\langle f_1,\mu\rangle, \dots, \langle f_n,\mu\rangle) \,\Big|\, F \in \mathcal{C}_n,\ f_1,\dots,f_n \in \mathbf{D} \Big\}$$

is a countable subset of $\mathcal{F}C_b^\infty$. Moreover, Proposition 1.3 of [23] and its proof show that $\mathbf{F}$ can replace $\mathcal{F}C_b^\infty$ in (8).

2. Let

$$\rho_n(\lambda,\mu) := \sup_{k\le n}\big[u_k(\lambda) - u_k(\mu)\big] \tag{28}$$

where the u_k are as in 1. Then ρ_n is continuous on $\mathcal{M}_1 \times \mathcal{M}_1$ and $\rho_n(\lambda,\mu) \nearrow \rho(\lambda,\mu)$. Hence, ρ is lower semicontinuous.

3. Suppose that (μ_n) is a sequence converging to μ in $\mathcal{M}_1$, and take a subsequence (μ_{n_k}) such that $\rho_K(\mu_{n_k}) \to \liminf_n \rho_K(\mu_n)$. By compactness of K, there are $\lambda_n \in K$ such that $\rho_K(\mu_n) = \rho(\lambda_n,\mu_n)$. Passing to a subsequence if necessary, we may assume that λ_{n_k} converges to some $\lambda \in K$. Thus, by assertion 2,

$$\rho_K(\mu) \le \rho(\lambda,\mu) \le \liminf_{k\uparrow\infty} \rho(\lambda_{n_k},\mu_{n_k}) = \liminf_{n\uparrow\infty} \rho_K(\mu_n)\,.$$

4. If E is finite, then ρ is continuous. Thus, ρ_K is upper semicontinuous as an infimum of continuous functions. Moreover, ρ_K coincides with $\rho_{\overline{K}}$, where $\overline{K}$ is the closure of K. As $\overline{K}$ is compact, ρ_K is also lower semicontinuous by 3. $\square$

Proposition 22 *If K is compact, then ρ_K is an $\mathcal{E}$-quasi continuous function in $D(\mathcal{E})$ and ρ_K satisfies $(D\rho_K, D\rho_K) \le 1$ m-a.e.*

Proof: Let $\rho_{K,n}(\mu) := \inf_{\lambda\in K}\rho_n(\lambda,\mu)$, where ρ_n is as in (28). Clearly, $\rho_{K,n}$ is increasing in n and less than ρ_K. Next, we will show that in fact $\rho_{K,n} \nearrow \rho_K(\mu)$ for all $\mu \in \mathcal{M}_1$. To this end, take some $\alpha < \rho_K(\mu)$ and let $U_n := \{\lambda \,|\, \rho_n(\lambda,\mu) > \alpha\}$. Then U_n is open and for each $\lambda \in K$ there is some $n_\lambda \in \mathbb{N}$ such that $\lambda \in U_{n_\lambda}$. In other words, $\{U_n \,|\, n \in \mathbb{N}\}$ covers K, and we can hence find some n_0 with $K \subset U_{n_0}$. But this means that $\rho_{K,n}(\mu) > \alpha$ for all $n \ge n_0$, which proves that $\rho_{K,n}(\mu) \nearrow \rho_K(\mu)$.

As in the second half of the proof of Lemma 21, one shows that $\rho_{K,n}$ is both upper and lower semicontinuous and, thus, continuous. Moreover, it is easily seen that $\rho_{K,n}$ is ρ-Lipschitz continuous. Hence, Theorem 19 implies that $\rho_{K,n} \in D(\mathcal{E})$ with $(D\rho_{K,n}, D\rho_{K,n}) \le 1$ m-a.e. Therefore, ρ_K is the pointwise limit of continuous functions with bounded Dirichlet norm, which implies the assertion by Lemma I.2.12 and Proposition III.3.5 of [15]. $\square$

Corollary 23 *Suppose that $M \subset \mathcal{M}_1$ satisfies $m(M) = 1$. Then the complement of its closure in ρ-distance (or, equivalently, in total variation norm),*

$$\big\{\rho_M > 0\big\} = \big\{\lambda \in \mathcal{M}_1 \,\big|\, \inf_{\mu\in M} \|\lambda-\mu\|_{\text{var}} > 0\,\big\}\,,$$

is an exceptional set for $(\mathcal{E}, D(\mathcal{E}))$.

Proof: By inner regularity, there are compact sets $K_1 \subset K_2 \subset \cdots \subset M$ such that $m(K_n) \nearrow 1$. By Proposition 22, $\rho_{K_n} \in D(\mathcal{E})$ and $\mathcal{E}(\rho_{K_n}, \rho_{K_n}) \leq 1$. Moreover, ρ_{K_n} is $\mathcal{E}$-quasi continuous and vanishes on K_n. Let u denote the pointwise limit of the functions ρ_{K_n} as $n \uparrow \infty$. By the same arguments as in the proof of Proposition 22, u is $\mathcal{E}$-quasi continuous. Since $m(u = 0) = 1$, we have $u = 0$ $\mathcal{E}$-quasi everywhere by Proposition III.3.9 of [15]. But $\{u = 0\} \subset \{\rho_M > 0\}$. That ρ and $\|\cdot\|_{\text{var}}$ generate the same topology on $\mathcal{M}_1$ follows from (17). ☐

Acknowledgment: The writer thanks H. Föllmer, M. Röckner, B. Schmuland, and W. Stannat for discussions and helpful remarks.

References

[1] Albeverio, S., Kondratiev, Yu.G., Röckner, M. *Analysis and geometry on configuration spaces.* J. Funct. Anal. 154, No.2, 444-500, (1998).

[2] Albeverio, S., Kondratiev, Yu.G., Röckner, M. *Analysis and geometry on configuration spaces: The Gibbsian case.* J. Funct. Anal. 157, No.1, 242-291, (1998).

[3] Amari, S. *Differential-geometrical methods in statistics.* Lecture Notes in Statistics, 28. Berlin etc.: Springer-Verlag (1985).

[4] Bogachev, V.I., Mayer-Wolf, E. *Some remarks on Rademacher's theorem in infinite dimensions.* Potential Anal. 5, 23–30 (1996).

[5] Cheeger, J. *Differentiability of Lipschitz functions on metric measure spaces.* GAFA, Geom. Funct. Anal. 9, 418-517 (1999).

[6] Dawson, D. A. *Measure-valued Markov processes.* In: Ecole d'Eté de Probabilités de Saint-Flour XXI, (Lect. Notes Math. vol. 1541, pp. 1-260) Berlin etc.: Springer (1993).

[7] Enchev, O., Stroock, D. *Rademacher's theorem for Wiener functionals.* Ann. Probab. 21, 25-33 (1993).

[8] Ethier, S., Kurtz, T. *Fleming-Viot processes in population genetics.* SIAM J. Control Optimization 31, No.2, 345-386 (1993).

[9] Ethier, S., Kurtz, T. *Markov processes. Characterization and convergence.* Wiley Series in Probability and Mathematical Statistics. New York etc.: Wiley (1986).

[10] Friedrich, T. *Die Fisher-Information und symplektische Strukturen.* Math. Nachr. 153, 273-296 (1991).

[11] Handa, K. *Quasi-invariance and reversibility in the Fleming-Viot process.* Preprint Saga University (1999); to appear in: Probab. Theory Rel. Fields.

[12] Hille, E., Phillips, R. *Functional analysis and semi-groups.* Providence: American Mathematical Society (1957).

[13] Kusuoka, S. *Dirichlet forms and diffusion processes on Banach spaces.* J. Fac. Sci., Univ. Tokyo, Sect. I A 29, 79-95 (1982).

[14] Li, Z., Shiga, T., Yao, L. *A reversibility problem for Fleming-Viot processes.* Elect. Comm. in Probab. 4, 65-76 (1999).

[15] Ma, Z. M., Röckner, M. *Introduction to the theory of (nonsymmetric) Dirichlet forms.* Springer-Verlag, Berlin (1992).

[16] McShane, E. *Extension of range of functions.* Bull. AMS 40, 837-842 (1934).

[17] Overbeck, L., Röckner, M. *Geometric aspects of finite and infinite-dimensional Fleming-Viot processes.* Random Oper. Stoch. Equ. 5, No.1, 35-58 (1997).

[18] Overbeck, L., Röckner, M., Schmuland, B. *An analytic approach to the Fleming-Viot processes with interactive selection.* Ann. Probab. 23, No.1, 1-36 (1995).

[19] Rademacher, H. *Über partielle und totale Differenzierbarkeit von Funktionen mehrerer Variabeln und über die Transformation der Doppelintegrale.* Math. Ann. 79, 340-359 (1919).

[20] Reed, M., Simon, B. *Methods of modern mathematical physics II: Fourier analysis, self-adjointness.* San Diego, Academic Press (1975).

[21] Riesz, F., Nagy, S. B. *Vorlesungen über Funktionalanalysis.* Berlin, VEB Deutscher Verlag der Wissenschaften (1956).

[22] Röckner, M., Schied, A. *Rademacher's theorem on configuration spaces and applications.* J. Funct. Ana. 169, No. 2, 325-356 (1999).

[23] Schied, A. *Geometric aspects of Fleming-Viot and Dawson-Watanabe processes.* Ann. Probab. 25, No.3, 1160-1179 (1997).

[24] Shiga, T. *A stochastic equation based on a Poisson system for a class of measure-valued diffusion processes.* J. Math. Kyoto Univ. 30, 245-279 (1990).

[25] Stannat, W. *On the validity of the log-Sobolev inequality for symmetric Fleming-Viot operators.* Ann. Probab., Vol. 28, No. 2, 667-684, (2000).

[26] Sturm, K.-T. *The geometric aspect of Dirichlet forms.* In: Jost, J. et al., New directions in Dirichlet forms. Providence, RI: American Mathematical Society. AMS/IP Stud. Adv. Math. 8, 233-277 (1998).

[27] Sturm, K.-T. *On the geometry defined by Dirichlet forms.* In: Bolthausen, E. et al., Seminar on stochastic analysis, random fields and applications. Proceedings of a seminar held at the Centro Stefano Franscini, Ascona, Switzerland, June 7-12, 1993. Basel: Birkhäuser. Prog. Probab. 36, 231-242 (1995).

[28] Sturm, K.-T. *Analysis on local Dirichlet spaces I. Recurrence, conservativeness and L^p-Liouville properties.* J. reine angew. Math. 456, 173-106 (1994).

[29] Tsilevich, N., Vershik, A. *Quasi-invariance of the gamma process and multiplicative properties of the Poisson-Dirichlet measures.* C. R. Acad. Sci., Paris, Ser. I, Math. 329, No.2, 163-168 (1999).

[30] Tsilevich, N., Vershik, A., Yor, M. *Distinguished properties of the gamma process and related topics.* Preprint, Sankt Petersburg (2000).