The Annals of Probability
1997, Vol. 25, No. 3, 1160–1179

GEOMETRIC ASPECTS OF FLEMING–VIOT AND DAWSON–WATANABE PROCESSES

BY ALEXANDER SCHIED

Humboldt-Universität zu Berlin

This paper is concerned with the intrinsic metrics of the two main classes of superprocesses. For the Fleming–Viot process, we identify it as the Bhattacharya distance, and for Dawson–Watanabe processes, we find the Kakutani–Hellinger metric. The corresponding geometries are studied in some detail. In particular, representation formulas for geodesics and arc length functionals are obtained. The relations between the two metrics yield a geometric interpretation of the identification of the Fleming–Viot process as a Dawson–Watanabe superprocess conditioned to have total mass 1. As an application, a functional limit theorem for super-Brownian motion conditioned on local extinction is proved.

0. Introduction. Let A be the generator of a diffusion process. One can use the corresponding *carré du champs operator*:

$$\Gamma(u, v) := A(uv) - uAv - vAu$$

to define an *intrinsic metric* ρ associated with the diffusion by setting

$$\rho(x, y) = \sup\{u(x) - u(y) \mid \Gamma(u, u)(z) \le 1\, \forall z\}. \tag{0.1}$$

If, for example, A is an elliptic second-order differential operator with smooth coefficients on $\mathbb{R}^d$, then it is well known that ρ is the Riemannian distance function associated with the Riemannian inner product obtained by inverting the diffusion matrix of A. For more general diffusions, the intrinsic metric (0.1) can replace this Riemannian distance in heat kernel estimates and large deviation theorems. See, for instance, Carlen, Kusuoka and Strook (1987) and Davies (1989). For other probabilistic applications of (0.1) see Sturm (1994, 1995) and Kuwae and Uemura (1995).

The aim of this paper is to calculate the intrinsic metrics associated with both the Fleming–Viot and the Dawson–Watanabe superprocesses. In particular, this complements the results of Overbeck and Röckner (1996), where the Fleming–Viot process has been analyzed by differential geometric methods, but without identification of the intrinsic metric.

Our formulas for the intrinsic metrics will be presented in the next section. In Section 2 we will examine the corresponding geometries in terms of their geodesics and arc length functionals. Also we will provide a geometric interpretation of the results in Etheridge and March (1991) and Perkins (1991),

Received February 1996; revised October 1996.
AMS 1991 *subject classifications.* Primary 60J60, 60G57; secondary 58G32, 60J80.
Key words and phrases. Intrinsic metric, Fleming–Viot process, Dawson–Watanabe superprocess, Kakutani–Hellinger distance, Bhattacharya metric.

where the Fleming–Viot superprocess is identified as a Dawson–Watanabe process conditioned to have total mass 1, and we will rediscover some results of Overbeck and Röckner (1996). As an application, a functional limit theorem for super-Brownian motion conditioned on local extinction is proved in Section 3. The final Section 4 contains proofs of our results.

1. The intrinsic metrics for the Fleming–Viot and the Dawson–Watanabe superprocesses. Let $M = M(E)$ denote the space of positive finite measures on some standard Borel space (E, B) and let M_1 will denote the subset of probability distributions. The variation norm of a signed measure μ on (E, B) is defined as usual by

$$\|\mu\|_{\mathrm{var}} = \sup\{\langle f, \mu\rangle \mid f \text{ is measurable and } |f(x)| \le 1\ \forall\, x \in E\}.$$

Here we wrote $\langle f, \mu\rangle$ for the integral of f with respect to μ. The space $M^{\pm}$ of all signed measured on (E, B) endowed with $\|\cdot\|_{\mathrm{var}}$ is a Banach space [see Dunford and Schwartz (1967), III.7.4]. However, it is only separable if E is countable.

A *curve* into M (or M_1) will be mapping $\omega: [0, 1] \to M$ (or $\omega: [0, 1] \to M_1$) that is continuous in variation. We will say that a curve ω is differentiable if it can be differentiated in $M^{\pm}$ with respect to $\|\cdot\|_{\mathrm{var}}$ and has a derivative $\dot\omega(t) \in M^{\pm}$ $(0 \le t \le 1)$. With $C^1(M)$ we will denote the set of all continuous functions $u: M \to \mathbb{R}$ such that, for any differentiable curve ω, the mapping $t \mapsto u(\omega(t))$ is differentiable with derivative $(d/dt)u(\omega(t)) = \langle Du(\omega(t)), \dot\omega(t)\rangle$, with a bounded and measurable function $Du: M \times E \to \mathbb{R}$. If $u \in C^1(M)$ then Du can be obtained as follows:

$$Du(\mu, x) = \frac{d}{dt}\bigg|_{t=0} u(\mu + t\delta_x), \qquad \mu \in M,\ x \in E. \tag{1.0}$$

Here δ_x denotes the unit point mass in $x \in E$. By considering only functions on and curves into M_1, we can define a space $C^1(M_1)$. However, in this case it is more appropriate to consider

$$\overline{D}u(\mu, x) = \frac{d}{dt}\bigg|_{t=0} u((1-t)\mu + t\delta_x), \qquad \mu \in M_1,\ x \in E, \tag{1.1}$$

instead of (1.0). Clearly $(d/dt)u(\omega(t)) = \langle \overline{D}u(\omega(t)), \dot\omega(t)\rangle = \langle Du(\omega(t)), \dot\omega(t)\rangle$ holds for all differentiable curves into M_1.

Now let $(A, D(A))$ denote the generator of some conservative Feller proces with respect to some Polish topology on E. For a function $u \in C^1(M_1)$ of the form $u(\mu) = \phi(\langle f_1, \mu\rangle, \dots, \langle f_n, \mu\rangle)$, with $n \in \mathbb{N}$, a bounded and smooth function ϕ on $\mathbb{R}^n$ having bounded derivatives and $f_1, \dots, f_n \in D(A)$, the infinitesimal generator of a *Fleming–Viot process with mutation operator* A takes the form

$$\overline{L}u(\mu) = \tfrac{1}{2}\int \overline{D}^2 u(\mu, x)\,\mu(dx) + \int A\overline{D}u(\mu, x)\,\mu(dx), \qquad \mu \in M_1, \tag{1.2}$$

where $\overline{D}^2 u(\mu, x) := \overline{D}(\overline{D}u(\cdot, x))(\mu, x)$ and $A\overline{D}u(\mu, x) := A(\overline{D}u(\mu, \cdot))(x)$. We refer the reader to Ethier and Kurtz (1986) and Dawson (1993) for surveys on Fleming–Viot processes. It follows easily from (1.2) that, for u as above, the *carré du champs* operator $\overline{\Gamma}$ associated with $\overline{L}$ takes the form

$$\overline{\Gamma}(u, u)(\mu) = \int \left(\overline{D}u(\mu, x)\right)^2 \mu(dx), \qquad \mu \in M_1$$

and hence can be extended to the whole of $C^1(M_1)$. We now can state our result on the intrinsic metric of a Fleming–Viot process.

THEOREM 1.1. *For $\nu, \mu \in M_1$ and any $\eta \in M$ such that both ν and μ are absolutely continuous with respect to η, define*

$$\delta(\nu, \mu) = \arccos \int \sqrt{\frac{d\nu}{d\eta}\frac{d\mu}{d\eta}}\, d\eta. \tag{1.3}$$

Then

$$\sup\left\{u(\mu) - u(\nu) \mid \overline{\Gamma}(u, u) \le 1, u \in C^1(M_1)\right\} = 2\,\delta(\mu, \nu). \tag{1.4}$$

It can be seen easily that the right-hand side of (1.3) does not depend on the particular choice of η. According to Amari (1985), $\delta(\nu, \mu)$ is called the *Bhattacharya distance* of μ and ν.

The Dawson–Watanabe superprocesses form another important class of measure-valued diffusions. We refer to Dawson (1993) for a survey. The infinitesimal generator L of the kind of processes we are interested in here can be described as in (1.2) by simply omitting bars. That is, for $(A, D(A))$, and $u = \phi(\langle f_1, \cdot \rangle, \dots, \langle f_n, \cdot \rangle) \in C^1(M)$ as above,

$$Lu(\mu) = \tfrac{1}{2}\int D^2 u(\mu, x)\mu(dx) + \int ADu(\mu, x)\mu(dx), \qquad \mu \in M, \tag{1.5}$$

where $D^2 u(\mu, x) := D(Du(\cdot, x))(\mu, x)$ and $ADu(\mu, x) := A(Du(\mu, \cdot))(x)$. If, for example, A is the Laplace operator on $E = \mathbb{R}^d$, L generates the so-called super-Brownian motion. For u as above, L has a *carré du champs* operator Γ given by

$$\Gamma(u, u) = \int (Du(\mu, x))^2 \mu(dx), \qquad \mu \in M.$$

Again Γ can be extended to the whole of $C^1(M)$. Our next result is the following theorem.

THEOREM 1.2. *For $\nu, \mu \in M$ and any $\eta \in M$ such that both ν and μ are absolutely continuous with respect to η, let*

$$d(\nu, \mu) = \left(\int \left(\sqrt{\frac{d\mu}{d\eta}} - \sqrt{\frac{d\nu}{d\eta}}\right)^2 d\eta\right)^{1/2}, \tag{1.6}$$

denote the Kakutani–Hellinger distance of ν and μ. Then

$$\sup\left\{u(\mu) - u(\nu) \mid \Gamma(u, u) \le 1, u \in C^1(M)\right\} = 2\,d(\nu, \mu). \tag{1.7}$$

Note that the right-hand side of (1.6) again is independent of η and that d differs here by a factor $\sqrt{2}$ from the Kakutani–Hellinger distance as defined in Jacod and Shiryaev (1987). The metric d appears also in the rate function for certain large deviation principles for super-Brownian motion, and has hence a clear probabilistic significance. See Schied (1996). Other consequences of our results will be discussed in the next section after we have studied the geometries of (M_1, δ) and (M, d) and their relations.

One might ask if our results would be affected by choosing other domains for the *carré du champs* operators instead of the C^1-spaces. Indeed, our formulas (1.2) and (1.5) for the infinitesimal generators $\overline{L}$ and L were a priori only valid on function spaces of the form

$$FC_b^\infty(\mathbf{D}) := \{u(\mu) = \phi(\langle f_1, \mu\rangle, \dots, \langle f_n, \mu\rangle) \mid f_1, \dots, f_n \in \mathbf{D},\ \phi \in C_b^\infty(\mathbb{R}^n),\ n \in \mathbb{N}\},$$

where $\mathbf{D}$ is a certain set of measurable and bounded functions on E and $C_b^\infty(\mathbb{R}^n)$ denotes the space of all bounded smooth functions on $\mathbb{R}^n$ with bounded derivatives.

PROPOSITION 1.3. *Suppose* $\mathbf{D}$ *contains all constant functions and an algebra generating the* σ*-field* B. *Then the conclusions of Theorems* 1.1 *and* 1.2 *remain true, if* $C^1(M_1)$ *and* $C^1(M)$ *in* (1.4) *and* (1.7) *respectively are replaced by* $FC_b^\infty(\mathbf{D})$.

2. The Kakutani–Hellinger geometry. Let d denote the Kakutani–Hellinger distance as in (1.6). We define *energy* and *length* of a curve into M as follows:

$$E(\omega) = \sup\left\{\frac{1}{2}\sum_{i=1}^n \frac{d(\omega(t_i), \omega(t_{i-1}))^2}{t_i - t_{i-1}} \,\middle|\, 0 \le t_0 < t_1 < \cdots < t_n \le 1,\ n \in \mathbb{N}\right\},$$

$$L(\omega) = \sup\left\{\sum_{i=1}^n d(\omega(t_i), \omega(t_{i-1})) \,\middle|\, 0 \le t_0 < t_1 < \cdots < t_n \le 1,\ n \in \mathbb{N}\right\}.$$

It has been proved in Schied (1996) that E admits the following integral representation formula:

$$E(\omega) = \begin{cases} \dfrac{1}{8}\displaystyle\int_0^1 \left\| \frac{d\dot\omega(t)}{d\omega(t)} \right\|^2_{L^2(\omega(t))} dt, & \text{if } \omega \in \mathbf{H}, \\ \infty, & \text{otherwise}, \end{cases} \tag{2.0}$$

where $\mathbf{H}$ is the space of those $\omega: [0, 1] \to M$ that are of the form

$$\omega(t) = \omega(0) + \int_0^t \dot\omega(s)\, ds, \qquad 0 \le t \le 1, \tag{2.1}$$

for some locally finite signed measures $\dot\omega(s)$ which are absolutely continuous with respect to $\omega(s)$, for almost every s, and whose Radon–Nikodym derivative $d\dot{\omega}(x)/d\omega(s)$ satisfies

$$\int_0^1 \left\| \frac{d\dot{\omega}(t)}{d\omega(t)} \right\|^2_{L^2(\omega(t))} dt < \infty. \tag{2.2}$$

THEOREM 2.1. *If $\omega \in \mathbf{H}$, then $L(\omega)$ is given by*

$$L(\omega) = \frac{1}{2}\int_0^1 \left\| \frac{d\dot{\omega}(t)}{d\omega(t)} \right\|_{L^2(\omega(t))} dt. \tag{2.3}$$

Fix $\mu, \nu \in M$, choose any $\eta \in M$ such that $\mu, \nu \ll \eta$, and define a curve γ into M by

$$\frac{d\gamma(t)}{d\eta} = \left((1-t)\sqrt{\frac{d\nu}{d\eta}} + t\sqrt{\frac{d\mu}{d\eta}} \right)^2, \qquad 0 \le t \le 1. \tag{2.4}$$

Then, if ω is any curve such that $\omega(0) = \nu$ and $\omega(1) = \mu$, we have that

$$L(\omega) \ge L(\gamma) = d(\nu, \mu) \tag{2.5}$$

and

$$E(\omega) \ge E(\gamma) = \tfrac{1}{2} d(\nu, \mu)^2. \tag{2.6}$$

Moreover, equality in (2.6) *implies $\omega \equiv \gamma$, and equality in* (2.5) *implies that ω and γ coincide modulo reparameterization.*

Note that Jensen's inequality together with (2.0) and (2.3) implies that

$$L(\omega)^2 \le 2E(\omega), \tag{2.7}$$

for any curve ω.

A curve γ as in (2.5) will be called *minimal geodesic*. Of course, γ can be extended to a mapping $\gamma: \mathbb{R} \to M$, but in general it will not minimize the arc length outside the interval $[0,1]$. Indeed, suppose that $\nu \not\ll \mu$ and that $\gamma(t)$ is defined as in (2.4). Then, for $t > 1$,

$$\left| \sqrt{\frac{d\nu}{d\eta}} - \sqrt{\frac{d\gamma(t)}{d\eta}} \right| = |1 - |1-t||\sqrt{\frac{d\nu}{d\eta}} < t\sqrt{\frac{d\nu}{d\eta}} \quad \text{on } \left\{ \frac{d\mu}{d\eta} = 0 \right\}.$$

Since $\eta(d\mu/d\eta = 0) > 0$ by assumption, this implies that

$$d(\gamma(0), \gamma(t)) < t d(\nu, \mu) \quad \text{if } t > 1. \tag{2.8}$$

Let us now consider the angular distance δ defined in (1.3). For a curve ω into M_1 define as above energy and length with respect to δ:

$$\bar{E}(\omega) = \sup\left\{ \frac{1}{2}\sum_{i=1}^n \frac{\delta(\omega(t_i), \omega(t_{i-1}))^2}{t_i - t_{i-1}} \,\middle|\, 0 \le t_0 < t_1 < \cdots < t_n \le 1,\ n \in \mathbb{N} \right\},$$

$$\bar{L}(\omega) = \sup\left\{ \sum_{i=1}^n \delta(\omega(t_i), \omega(t_{i-1})) \,\middle|\, 0 \le t_0 < t_1 < \cdots < t_n \le 1,\ n \in \mathbb{N} \right\}.$$

THEOREM 2.2. *The angular distance* δ *defines a metric on* M_1 *and, for two probability measures* $\nu \neq \mu$,

$$d(\nu, \mu) < \delta(\nu, \mu) \leq \frac{\pi}{\sqrt{8}} d(\nu, \mu). \tag{2.9}$$

Moreover, $\overline{E}(\omega) = E(\omega)$ *and* $\overline{L}(\omega) = L(\omega)$ *holds for any curve* ω *into* M_1. *Fix* $\mu, \nu \in M_1$, *and define a curve* $\overline{\gamma}$ *into* M_1 *as follows by its Radon–Nikodym derivatives with respect to any* $\eta \gg \nu, \mu$:

$$\frac{d\overline{\gamma}(t)}{d\eta} = \left(\sqrt{\frac{d\nu}{d\eta}} (\cos \theta t - \cot \theta \sin \theta t) + \sqrt{\frac{d\mu}{d\eta}} \frac{\sin \theta t}{\sin \theta} \right)^2, \quad 0 \leq t \leq 1, \tag{2.10}$$

where $\theta := \delta(\nu, \mu)$. *Then, if* ω *is any other curve into* M_1 *such that* $\omega(0) = \nu$ *and* $\omega(1) = \mu$, *we have that*

$$\overline{L}(\omega) = L(\omega) \geq L(\gamma) = \delta(\nu, \mu) \tag{2.11}$$

and

$$\overline{E}(\omega) = E(\omega) \geq E(\gamma) = \tfrac{1}{2}\delta(\nu, \mu)^2. \tag{2.12}$$

Amari (1985) also studied some geometric aspects of d and δ, but only when they are restricted to the set of probability measures that are absolutely continuous with respect to some fixed reference measure η.

Let us now state a few consequences of the results we obtained so far. Equation (2.11) implies that

$$\delta(\nu, \mu) = \inf\{ L(\omega) \mid \omega(0) = \nu, \omega(1) = \mu \text{ and } \omega(t) \in M_1 \ \forall\, t \}. \tag{2.13}$$

That is, δ is induced by d and the inclusion $M_1 \subset M$. This seems to be the geometric interpretation of the results in Etheridge and March (1991) and Perkins (1991), where the Fleming–Viot process is identified as a Dawson–Watanabe superprocess conditioned to have total mass 1.

One might be tempted to transfer concepts from Riemannian geometry to our setting. Indeed, (2.3) suggests that the "tangent space" of (M, d) in some measure ν should be $L^2(\nu)$ with its usual inner product. This works well if E is countable and ν has full support. But in the general situation we can always find some nontrivial $\lambda \in M$ which is singular with respect to ν. Define η by $\eta = \nu + \lambda$ and let γ denote the geodesic (2.4) from ν to μ, where μ is defined by $d\mu/d\eta = ((d\lambda/d\eta)^{1/2} + (d\nu/d\eta)^{1/2})^2$. Then we have that $\| d\dot{\gamma}(0)/d\nu \|_{L^2(\nu)} = 0$. Thus geodesics are no longer determined by their starting point and an initial vector. In particular, there is no reasonable analogue of an exponential map.

Formally, (2.11) or the form of $\overline{\Gamma}$ suggest that the "tangent space" of (M_1, δ) in some $\nu \in M_1$ should be the orthogonal complement in $L^2(\nu)$ of the constant functions, and that it should be endowed with the $L^2(\nu)$ inner product. This means that (M_1, δ) carries the structure of a "Riemannian submanifold" of (M, d), and this gives another formal explanation of (2.13).

Proposition V.4.4 of Jacod and Shiryaev (1987) asserts that

$$\tfrac{1}{2}\|\nu - \mu\|_{\mathrm{var}} \le d(\nu, \mu) \le \|\nu - \mu\|_{\mathrm{var}}^{1/2}, \qquad \nu, \mu \in M_1, \tag{2.14}$$

where the upper bound also holds for $\nu, \mu \in M$. Combining (2.14) with (2.9) yields

$$\frac{1}{2}\|\nu - \mu\|_{\mathrm{var}} < \delta(\nu, \mu) \le \frac{\pi}{\sqrt{8}}\|\nu - \mu\|_{\mathrm{var}}^{1/2}, \qquad \nu, \mu \in M_1,\ \nu \ne \mu. \tag{2.15}$$

These estimates are sharp, and together with Theorem 1.1 they yield an improvement of Theorem 5.1 of Overbeck and Röckner (1996), which states a nonsharp estimate like (2.15) for the intrinsic metric of the Fleming–Viot process. Equations (2.14) and (2.15) show that convergence with respect to d or δ is a very strong concept. But in the case where the operator A in (1.2) and (1.5) is the generator of a Markov chain with bounded jump intensities, Shiga (1990) has proved that the corresponding measure-valued diffusions are almost surely continuous in variation, and hence with respect to their intrinsic metrics. Konno and Shiga (1988) proved a similar result, when A is the Laplace operator on $\mathbb{R}$.

There is a natural notion of curvature in geodesic spaces. It is called curvatue in the sense of CAT (Comparison, Alexandrov, Toponogov) and it is obtained by comparing geodesic triangles with comparison triangles in simply connected two-dimensional Riemannian manifolds of constant curvature. See Ballmann (1990) or Sturm (1994). In our case, it is clear from (2.10) that every geodesic triangle in (M_1, δ) matches another one in the Euclidean two-sphere of radius one. Thus (M_1, δ) has constant curvature $+1$ in the sense of CAT. According to Remark 3.3 in Overbeck and Röckner (1996), this fact has been independently noticed by B. Driver by calculating the curvature tensor that corresponds to the coefficient matrix of a Fleming–Viot process in the case where E is a finite set with $n + 1$ elements. In this situation, our equation (2.10) now shows in addition that (M_1, δ) is just a Euclidean n-sphere where two points are identified when they can be obtained from each other by reflections with respect to the hyperplanes $\{(x_1, \dots, x_{n+1}) \in \mathbb{R}^{n+1} \mid x_i = 0\}$ $(i = 1, \dots, n + 1)$. But this is just a sphere segment with boundary. These aspects of the Bhattacharya metric with finite E have been studied before in Amari (1985).

The sphere segment above can be parameterized on

$$K := \left\{(x_1, \dots, x_n) \in \mathbb{R}^n \mid x_i \ge 0,\ i = 1, \dots, n,\ \sum_{i=1}^{n} x_i \le 1\right\},$$

by the mapping

$$K \ni (x_1, \dots, x_n) \mapsto \Phi(x_1, \dots, x_n) := \left(x_1^{1/2}, \dots, x_n^{1/2}, \left(1 - \sum_{i=1}^{n} x_i\right)^{1/2}\right).$$

The canonical volume element of the n-sphere reads in these coordinates as

$$m(dx_1, \ldots, dx_n) = \left(x_1 \cdots x_n \left(1 - \sum_{i=1}^{n} x_i \right) \right)^{-1/2} dx_1 \cdots dx_n,$$

and this yields the invariant distribution of a Fleming–Viot process with parent independent mutation [i.e., the case $A = 0$ in (1.2)] when mapped onto K. Theorem V.4.6 in Ikeda and Watanabe (1989) can now be used to characterize all mutation operators A that yield symmetrizable drifts and to calculate their symmetrizing measures. This has been carried out in Theorem 2.2 of Overbeck and Röckner (1996) by direct means.

3. Small-time asymptotics for super-Brownian motion conditioned on local extinction. Suppose $p > n$ and let $M_p(\mathbb{R}^n)$ denote the space of all positive σ-finite Borel measures μ on $\mathbb{R}^n$ such that $\langle \phi_p, \mu \rangle < \infty$, where $\phi_p(x) = (1 + |x|^2)^{p/2}$ $(x \in \mathbb{R}^n)$. The space $M_p(\mathbb{R}^n)$ becomes Polish when endowed with the topology generated by the maps

$$M_p(\mathbb{R}^n) \ni \mu \mapsto \langle f, \mu \rangle, \qquad f \in \{\phi_p\} \cup C_c(\mathbb{R}^n).$$

Clearly one can extend the Kakutani–Hellinger distance d to $M_p(\mathbb{R}^n)$, but note that now the distance between two measures might be infinite. However, one can show easily that Theorem 2.1 with appropriate modifications remains true on this space.

Let $C([0,1]: M_p(\mathbb{R}^n))$ denote the space of all continuous paths from $[0,1]$ into $M_p(\mathbb{R}^n)$, and we endow this space with the usual compact open topology. Then this is the canonical path space for super-Brownian motion X, which is a diffusion with values in $M_p(\mathbb{R}^n)$ and generator

$$Lu = \langle D^2 u(\mu), \mu \rangle + \tfrac{1}{2} \langle \Delta Du(\mu), \mu \rangle, \; u \in FC_b^\infty(\mathbf{D}), \tag{3.1}$$

where $\mathbf{D}$ equals here the set of all smooth functions on $\mathbb{R}^n$ with compact support. Let $\mathbb{P}_\mu$ denote its law when starting from $\mu \in M_p(\mathbb{R}^n)$. Assume now that $\mu \in M_p(\mathbb{R}^n)$ has full support. It has been shown in Corollary 3 of Schied (1996) that, as $\varepsilon \downarrow 0$, the rescaled processes $X_t^\varepsilon := X_{\varepsilon t}$ $(0 \le t \le 1)$ satisfy a large deviation principle with good rate function

$$I_\mu^0(\omega) = \begin{cases} 2E(\omega), & \text{if } \omega(0) = \mu \text{ and } \omega \in D, \\ \infty, & \text{otherwise,} \end{cases} \tag{3.2}$$

where D is the set of all paths $\omega \in C([0,1]: M_p(\mathbb{R}^n))$ having decreasing support. The appearance of the intrinsic metric in the rate function shows that the Kukutani–Hellinger distance has some probablistic significance for Dawson–Watanabe superprocesses.

We now combine the above large deviation principle with the knowledge of the minimizing geodesics provided by Theorem 2.1 to prove a weak convergence result for super-Brownian motion conditioned on local extinction.

THEOREM 3.1. *Fix $\mu \in M_p(\mathbb{R}^n)$ with full support, and choose a closed set $N \subset \mathbb{R}^n$ such that $\mu(N) < \infty$. For $\varepsilon > 0$, let X^ε denote the process $X^\varepsilon_t = X_{\varepsilon t}$ $(0 \le t \le 1)$. Then the laws of X^ε under $\mathbb{P}_\mu[\cdot \mid X^\varepsilon_1(N) = 0]$ converge weakly as $\varepsilon \downarrow 0$ to the Dirac mass in the curve $\gamma \in C([0,1]: M_p(\mathbb{R}^n))$ given by*

$$\gamma(t) = \mu - t(2-t)\mu_N, \qquad 0 \le t \le 1,$$

where μ_N is the restriction of μ to N.

4. Proofs. We will prove our results in the following order: Theorem 2.1, Theorem 2.2, Theorem 1.2, Theorem 1.1, Proposition 1.3, and finally Theorem 3.1.

All vector-valued integrals that appear below are to be understood in the sense of Bochner. See Hille and Phillips (1957).

PROOF OF THEOREM 2.1. First we will prove (2.3). To this end fix $\omega \in \mathbf{H}$ and define $\eta \in M$ by $\eta = \int_0^1 \omega(t)\, dt$. If $A \in B$ and $\eta(A) = 0$, then $\omega(t)(A) = 0$ for every $t \in [0,1]$ since $t \mapsto \omega(t)(A)$ is continuous by (2.1) and (2.2). Hence, for each $t \in [0,1]$, there is a function $\varphi(t) \in L^1(\eta)$ such that $d\omega(t) = \varphi(t)\, d\eta$. Since

$$\int_0^1 \left\| \frac{d\dot\omega(t)}{d\eta} \right\|_{L^1(\eta)} dt = \int_0^1 \left\| \frac{d\dot\omega(t)}{d\omega(t)} \right\|_{L^1(\omega(t))} dt \le c(\omega) \int_0^1 \left\| \frac{d\dot\omega(t)}{d\omega(t)} \right\|_{L^2(\omega(t))} dt < \infty,$$

by (2.2), $t \mapsto d\dot\omega(t)/d\eta$ is Bochner integrable in $L^1(\eta)$ [cf. Theorem 3.7.4 in Hille and Phillips (1957)]. By testing with bounded and measurable functions it follows easily that, for all $t \in [0,1]$,

$$\varphi(t) = \varphi(0) + \int_0^t \frac{d\dot\omega(s)}{d\eta}\, ds,$$

holds in $L^1(\eta)$. Therefore the mapping $t \mapsto \varphi(t) \in L^1(\eta)$ is strongly differentiable almost everywhere and possesses the derivative $\dot\varphi(t) = d\dot\omega(t)/d\eta$ [cf. Theorem 3.8.5 in Hille and Phillips (1957)].

LEMMA 4.1. *For φ as above, the identity*

$$\sqrt{\varphi(t)} - \sqrt{\varphi(0)} = \tfrac{1}{2}\int_0^t \dot\varphi(s)\varphi(s)^{-1/2}\, ds, \qquad 0 \le t \le 1 \tag{4.1}$$

holds in $L^2(\eta)$.

PROOF. Define functions f_n' and f_n on $[0,\infty)$ by $f_n'(x) = n \wedge (4x)^{-1/2}$ and $f_n(x) = \int_0^x f_n'(y)\, dy$ $(n = 1, 2, \dots)$. Equations (53) and (54) in Schied (1996) assert that

$$f_n(\varphi(t)) - f_n(\varphi(0)) = \int_0^t f_n'(\varphi(s))\dot\varphi(s)\, ds$$

in $L^2(\eta)$. For almost every s,

$$f_n'(\varphi(s))\dot\varphi(s) \to \tfrac{1}{2}\dot\varphi(s)\varphi(s)^{-1/2} \qquad \text{as } n\uparrow\infty, \tag{4.2}$$

pointwise on E. But, for all n and almost every s,

$$\left(f_n'(\varphi(s))\dot\varphi(s)\right)^2 \le \tfrac{1}{4}\dot\varphi(s)^2\varphi(s)^{-1} \in L^1(\eta).$$

Hence (4.2) takes place even in $L^2(\eta)$. Moreover,

$$\int_0^1 \|\dot\varphi(s)\varphi(s)^{-1/2}\|_{L^2(\eta)}\,ds = \int_0^1 \left\|\frac{d\dot\omega(s)}{d\omega(s)}\right\|_{L^2(\omega(s))} ds < \infty$$

by (2.2), and hence, for each t,

$$\int_0^t f_n'(\varphi(s))\dot\varphi(s)\,ds \to \tfrac{1}{2}\int_0^t \dot\varphi(s)\varphi(s)^{-1/2}\,ds \qquad \text{as } n\uparrow\infty,$$

in $L^2(\eta)$ by dominated convergence for Bochner integrals [see Hille and Phillips (1957), Theorem 3.7.9]. On the other hand, by monotone convergence,

$$f_n(\varphi(t)) \to \sqrt{\varphi(t)} \qquad \text{as } n\uparrow\infty$$

in $L^2(\eta)$. Therefore the lemma is proved. □

It follows from the lemma that

$$d(\omega(s),\omega(t)) = \left\|\tfrac{1}{2}\int_s^t \dot\varphi(u)\varphi(u)^{-1/2}\,du\right\|_{L^2(\eta)} \le \tfrac{1}{2}\int_s^t \|\dot\varphi(u)\varphi(u)^{-1/2}\|_{L^2(\eta)}\,du.$$

Hence

$$L(\omega) \le \frac{1}{2}\int_0^1 \|\dot\varphi(u)\varphi(u)^{-1/2}\|_{L^2(\eta)}\,du = \frac{1}{2}\int_0^1 \left\|\frac{d\dot\omega(u)}{d\omega(u)}\right\|_{L^2(\omega(u))} dt.$$

To prove the reverse inequality, choose $\varepsilon > 0$ and a step function $\psi: [0,1] \to L^2(\eta)$ such that $\int_0^1 \|\dot\varphi(u)\varphi(u)^{-1/2}/2 - \psi(u)\|_{L^2(\eta)}\,du < \varepsilon$. See Hille and Phillips (1957), page 86, for existence. Let $\Delta = \{s_0,\dots,s_k\}$ denote an ordered partition of $[0,1]$ such that ψ is constant a.e. on $[s_{i-1}, s_i]$ $(i = 1,\dots,n)$. We can find a refinement $\Delta' = \{t_0,\dots,t_l\}$ of Δ such that

$$L(\omega) < \sum_{i=1}^{l} d(\omega(t_{i-1}),\omega(t_i)) + \varepsilon = \sum_{i=1}^{l} \left\|\int_{t_{i-1}}^{t_i} \dot\varphi(s)\varphi(s)^{-1/2}\,ds\right\|_{L^2(\eta)} + \varepsilon,$$

by Lemma 4.1. Then we have that

$$\sum_{i=1}^{l} \left|\left\|\int_{t_{i-1}}^{t_i} \tfrac{1}{2}\dot\varphi(s)\varphi(s)^{-1/2}\,ds\right\|_{L_2(\eta)} - \left\|\int_{t_{i-1}}^{t_i} \psi(s)\,ds\right\|_{L^2(\eta}\right|$$
$$\le \int_0^1 \left\|\tfrac{1}{2}\dot\varphi(s)\varphi(s)^{-1/2} - \psi(s)\right\|_{L^2(\eta)} ds < \varepsilon,$$

and

$$\left|\tfrac{1}{2}\int_0^1 \|\dot{\varphi}(s)\varphi(s)^{-1/2}\|_{L^2(\eta)}\,ds - \sum_{i=1}^{l}\left\|\int_{t_{i-1}}^{t_i}\psi(s)\,ds\right\|_{L^2(\eta)}\right|$$

$$=\left|\tfrac{1}{2}\int_0^1 \|\dot{\varphi}(s)\varphi(s)^{-1/2}\|_{L^2(\eta)}\,ds - \int_0^1 \|\psi(s)\|_{L^2(\eta)}\,ds\right|$$

$$\le \int_0^1 \left\|\tfrac{1}{2}\dot{\varphi}(s)\varphi(s)^{-1/2} - \psi(s)\right\|_{L^2(\eta)}ds < \varepsilon.$$

Putting these estimates together, we arrive at

$$\left| L(\omega) - \tfrac{1}{2}\int_0^1 \|\dot{\varphi}(s)\varphi(s)^{-1/2}\|_{L^2(\eta)}\,ds\right| < 3\varepsilon.$$

Since ε was arbitrary, (2.3) is proved.

Now consider the curve γ given by (2.4). Then, since $t \mapsto \sqrt{d\gamma(t)/d\eta}$ is a straight line in $L^2(\eta)$, we have that $d(\gamma(s),\gamma(t)) = |t-s|d(\nu,\mu)$. Hence $L(\gamma) = d(\nu,\mu)$, $E(\gamma) = d(\nu,\mu)^2/2$ and $\gamma \in \mathbf{H}$. Since $d(\nu,\mu) \le L(\omega)$ and $d(\nu,\mu)^2 \le 2E(\omega)$ are trivial for any curve ω connecting ν and μ, the above implies (2.5) and (2.6).

Now suppose $L(\omega) = d(\nu,\mu)$. We will show first that this implies that the graphs of ω and γ coincide:

$$\{\omega(t) \mid 0 \le t \le 1\} = \{\gamma(t) \mid 0 \le t \le 1\}. \tag{4.3}$$

Observe that it suffices to prove only the inclusion $\subset$, because ω is continuous. Clearly, for any $t \in [0,1]$,

$$d(\nu,\mu) = L(\omega) \ge d(\nu,\omega(t)) + d(\omega(t),\mu) \ge d(\nu,\mu), \tag{4.4}$$

and the inequalities are here in fact equalities. If we define η_0 by $\eta_0 = \nu + \mu + \omega(t)$, (4.4) implies that

$$\left\|\sqrt{\frac{d\nu}{d\eta_0}} - \sqrt{\frac{d\omega(t)}{d\eta_0}}\right\|_{L^2(\eta_0)} + \left\|\sqrt{\frac{d\omega(t)}{d\eta_0}} - \sqrt{\frac{d\mu}{d\eta_0}}\right\|_{L^2(\eta_0)}$$

$$= \left\|\sqrt{\frac{d\nu}{d\eta_0}} - \sqrt{\frac{d\mu}{d\eta_0}}\right\|_{L^2(\eta_0)}.$$

Hence $(d\omega(t)/d\eta_0)^{1/2}$ must lie on the straight line from $(d\nu/d\eta_0)^{1/2}$ to $(d\mu/d\eta_0)^{1/2}$ in the space $L^2(\eta_0)$. Since (2.4) is independent of η, the above shows that $\omega(t)$ is an element of the graph of γ, and (4.3) is proved. To construct the desired reparametrization it suffices to observe that $t \mapsto d(\nu,\omega(t))$ is increasing. Indeed, $d(\nu,\omega(t+h)) = d(\nu,\omega(t)) + d(\omega(t),\omega(t+h))$ as can be shown by arguing as in (4.4).

Next suppose that $E(\omega) = d(\nu,\mu)^2/2$. Equation (2.7) now yields $2E(\omega) = L(\omega)^2 = d(\nu,\mu)^2$. Hence ω coincides with γ at least modulo reparametrization. In addition $E(\omega)$ is finite and thus $\omega \in \mathbf{H}$. Therefore, by (2.3),

$$\frac{1}{4}\left(\int_0^1 \left\| \frac{d\dot{\omega}(t)}{d\omega(t)} \right\|_{L^2(\omega(t))} dt\right)^2 = L(\omega)^2 = 2E(\omega) = \frac{1}{4}\int_0^1 \left\| \frac{d\dot{\omega}(t)}{d\omega(t)} \right\|^2_{L^2(\omega(t))} dt.$$

But this can only happen if $\| d\dot{\omega}(t)/d\omega(t)\|_{L^2(\omega(t))}$ is constant in t. This easily implies $\gamma \equiv \omega$ and Theorem 2.1 is proved. □

PROOF OF THEOREM 2.2. First let us show that δ is indeed a metric on M_1. Only the triangle inequality is not obvious. To prove it, suppose $\nu, \mu, \lambda \in M_1$, fix $\eta \gg \nu, \mu, \lambda$, and denote by φ_ν, φ_μ and φ_λ the square-roots of the corresponding Radon–Nikodym derivatives. Then let $\theta_1 := \delta(\nu, \mu)$, $\theta_2 := \delta(\mu, \lambda)$ and

$$\hat{\varphi}_\nu := \varphi_\nu \frac{1}{\sin\theta_1} - \varphi_\mu \cot\theta_1 \quad \text{and} \quad \hat{\varphi}_\lambda := \varphi_\lambda \frac{1}{\sin\theta_2} - \varphi_\mu \cot\theta_2.$$

Then $\hat{\varphi}_\lambda, \hat{\varphi}_\nu \perp \varphi_\mu$ in $L^2(\eta)$, $\varphi_\nu = \varphi_\mu \cos\theta_1 + \hat{\varphi}_\nu \sin\theta_1$ and $\varphi_\lambda = \varphi_\mu \cos\theta_2 + \hat{\varphi}_\lambda \sin\theta_2$. Thus

$$\begin{aligned}\cos\delta(\nu, \lambda) &= \cos\theta_1 \cos\theta_2 - \langle \hat{\varphi}_\nu, \hat{\varphi}_\lambda \rangle_{L^2(\eta)} \sin\theta_1 \sin\theta_2 \\ &\geq \cos\theta_1\cos\theta_2 - \sin\theta_1 \sin\theta_2 \\ &= \cos(\delta(\nu,\mu) + \delta(\mu,\lambda)).\end{aligned}$$

As δ only takes values in $[0, \pi/2]$, this shows that

$$\delta(\nu, \lambda) \leq \delta(\nu, \mu) + \delta(\mu, \lambda) \quad \text{with equality iff } \hat{\varphi}_\nu = \hat{\varphi}_\mu. \tag{4.5}$$

To prove (2.9) observe that

$$\delta(\nu, \mu) = \arccos\left(1 - \tfrac{1}{2} d(\nu, \mu)^2\right). \tag{4.6}$$

But $0 < d(\nu, \mu) \leq \sqrt{2}$ and $1 < \arccos(1 - x^2/2)/x \leq \pi/\sqrt{8}$, for $0 < x \leq \sqrt{2}$.

Now fix a curve ω into M_1. We have $\overline{L}(\omega) \geq L(\omega)$ and $\overline{E}(\omega) \geq E(\omega)$ from (2.9). To prove the converse assertion for the arc length, define $\overline{L}_\Delta(\omega) := \sum_{i=1}^n \delta(\omega(t_i), \omega(t_{i-1}))$ if $\Delta = \{t_0, t_1, \dots, t_n\}$ is any ordered partition of the unit interval. Next we remark that $\arccos(1 - x^2/2)/x \downarrow 1$ as $x \downarrow 0$. Hence we infer from (4.6) that for any $\varepsilon > 0$ there is some $r > 0$ such that $\delta(\alpha, \lambda) \leq (1 + \varepsilon) d(\alpha, \lambda)$ whenever $\alpha, \lambda \in M_1$ are such that $d(\alpha, \lambda) \leq r$. This shows that $\overline{L}_\Delta(\omega) \leq (1 + \varepsilon) L(\omega)$, if we suppose that $d(\omega(t_i), \omega(t_{i-1})) \leq r$, for all $t_i \in \Delta$. But δ is a metric, and hence

$$\overline{L}_{\Delta'}(\omega) \geq \overline{L}_\Delta(\omega) \quad \text{if } \Delta' \text{ is finer than } \Delta. \tag{4.7}$$

Since ω was supposed to be strongly continuous, we arrive at $\overline{L}(\omega) = \sup_\Delta \overline{L}_\Delta(\omega) \leq (1 + \varepsilon) L(\omega)$, where the supremem is taken over all partitions Δ that are fine enough in the above sense.

In principle the same reasoning applies to $\overline{E}(\omega)$. However, some additional care is needed, because the property analogous to (4.7) is uncertain in general. But it remains true if Δ and Δ' are dyadic partitions, and this is sufficient to prove the assertion. See Schied (1996), Lemma 24 for the details.

Now we proceed to show that $\bar{\gamma}$ is indeed a minimizing geodesic. Choose η dominating μ and ν, and let $\psi(t)$ denote the Radon–Nikodym derivative $d\bar{\gamma}(t)/d\eta$. Clearly, $t \mapsto \psi(t) \in L^1(\eta)$ is strongly differentiable, and its derivative satisfies

$$\dot{\psi}(t) = \frac{d}{dt}\psi(t) = 2\theta\Big(\hat{\varphi}\cos\theta t - \sqrt{\psi(0)}\,\sin\theta t\Big)\sqrt{\psi(t)}\,, \tag{4.8}$$

where $\hat{\varphi} = (\sin\theta)^{-1}\sqrt{\psi(1)} - \cot\theta\sqrt{\psi(0)}$. Therefore $t \mapsto \psi(t) \in L^1(\eta)$ is of strong bounded variation. Moreover ψ is weakly absolutely continuous, and so $\psi(t) = \psi(0) + \int_0^t \dot{\psi}(s)\,ds$ holds in $L^1(\eta)$ by Theorem 3.6.8 in Hille and Phillips (1957). Thus (2.1) follows for $\bar{\gamma}$, with $\dot{\bar{\gamma}}$ given by $d\dot{\bar{\gamma}}(t) = \dot{\psi}(t)\,d\eta$.

From (4.8) we conclude that $\dot{\bar{\gamma}}(t) \ll \bar{\gamma}(t)$, for all t. To calculate the corresponding density observe that, for any measurable set A,

$$\dot{\bar{\gamma}}(t)(A) = \int_A \frac{d\dot{\bar{\gamma}}(t)}{d\bar{\gamma}(t)}\cdot\frac{d\bar{\gamma}(t)}{d\eta}\,d\eta = \int_A \dot{\psi}(t)\,d\eta.$$

Thus it follows that

$$\frac{d\dot{\bar{\gamma}}(t)}{d\bar{\gamma}(t)}\psi(t) = \dot{\psi}(t) = 2\theta\Big(\hat{\varphi}\cos\theta t - \sqrt{\psi(0)}\,\sin\theta t\Big)\sqrt{\psi(t)}\,, \quad \eta\text{-a.e.,}$$

which in turn implies that

$$\frac{d\dot{\bar{\gamma}}(t)}{d\bar{\gamma}(t)} = 2\theta\Big(\hat{\varphi}\cos\theta t - \sqrt{\psi(0)}\,\sin\theta t\Big)\psi(t)^{-1/2}, \quad \bar{\gamma}(t)\text{-a.e.}$$

Using $\|\hat{\varphi}\|_{L^2(\eta)} = 1$ and $\hat{\varphi} \perp \sqrt{\psi(0)}$ in $L^2(\eta)$, this yields

$$\left\|\frac{d\dot{\bar{\gamma}}(t)}{d\bar{\gamma}(t)}\right\|_{L^2(\bar{\gamma}(t))} = 2\theta\Big\|\hat{\varphi}\cos\theta t - \sqrt{\psi(0)}\,\sin\theta t\Big\|_{L^2(\eta)} = 2\delta(\mu,\nu).$$

Therefore $\bar{\gamma} \in \mathbf{H}$,

$$\delta(\mu,\nu) = \frac{1}{2}\int_0^1 \left\|\frac{d\dot{\bar{\gamma}}(t)}{d\bar{\gamma}(t)}\right\|_{L^2(\bar{\gamma}(t))} dt = L(\bar{\gamma}),$$

and

$$\frac{1}{2}\delta(\nu,\mu)^2 = \frac{1}{8}\int_0^1 \left\|\frac{d\dot{\bar{\gamma}}(t)}{d\bar{\gamma}(t)}\right\|^2_{L^2(\bar{\gamma}(t))} dt = E(\bar{\gamma}).$$

Hence Theorem 2.2 is proved. □

PROOF OF THEOREM 1.2. Suppose $u \in C^1(M)$ satisfies $\Gamma(u,u) \le 1$ and $\omega \in \mathbf{H}$ is a curve such that $\omega(0) = \lambda$ and $\omega(1) = \mu$. Then Hölder's inequality and (2.3) imply that

$$\begin{aligned} u(\lambda) - u(\mu) &= \int_0^1 \frac{d}{dt}u(\omega(t))\,dt = \int_0^1 \langle Du(\omega(t)), \dot{\omega}(t)\rangle\,dt \\ &\le \int_0^1 \sqrt{\Gamma(u,u)(\omega(t))}\left\|\frac{d\dot{\omega}(t)}{d\omega(t)}\right\|_{L^2(\omega(t))} dt \le 2L(\omega). \end{aligned}$$

Equation (2.5) now shows that

$$\rho(\lambda, \mu) := \sup\{u(\lambda) - u(\mu) \mid \Gamma(u, u) \le 1, u \in C^1(M)\} \le 2\, d(\lambda, \mu).$$

To prove the reverse inequality we will distinguish between the cases where E is countable or uncountable. Let us start with the latter case. Since the assertion of our theorem only depends on the measurable structure of (E, B), we can assume that $E = \mathbb{R}$ and that B is the usual Borel field [see, e.g., Theorem 2.12 of Parthasarathy (1967)]. Denote by ν_σ the normal distribution with mean 0 and variance $\sigma > 0$ and choose $\eta = \lambda + \mu + \nu_1$. Then, for any $\alpha \in M$ and each $\sigma > 0$, the convolution $\nu_\sigma * \alpha$ of ν_σ and α is absolutely continuous with respect to η. Now fix $\varepsilon, \sigma > 0$ and a bounded measurable function f, and define

$$u(\alpha) = \int f\left(\varepsilon + \frac{d\nu_\sigma * \alpha}{d\eta}\right)^{1/2} d\eta, \qquad \alpha \in M. \tag{4.9}$$

Clearly, for $\beta \in M$, $t \mapsto u(\alpha + t\beta)$ is differentiable and

$$\begin{aligned}\frac{d}{dt}u(\alpha + t\beta) &= \frac{1}{2}\int f\left(\varepsilon + \frac{d\nu_\sigma *(\alpha + t\beta)}{d\eta}\right)^{-1/2} \frac{d\nu_\sigma * \beta}{d\eta}\, d\eta \\ &= \frac{1}{2}\int \nu_\sigma * \left[f\left(\varepsilon + \frac{d\nu_\sigma *(\alpha + t\beta)}{d\eta}\right)^{-1/2}\right] d\beta.\end{aligned}$$

Therefore $u \in C^1(M)$.

By Jensen's inequality $(\nu_\sigma * g)^2 \le \nu_\sigma * g^2$, for any bounded and measurable function g on $\mathbb{R}$. Hence

$$\begin{aligned}\Gamma(u, u)(\alpha) &= \frac{1}{4}\int \left(\nu_\sigma * \left[f\left(\varepsilon + \frac{d\nu_\sigma * \alpha}{d\eta}\right)^{-1/2}\right]\right)^2 d\alpha \\ &\le \frac{1}{4}\int f^2 \left(\frac{d\nu_\sigma * \alpha}{d\eta}\right)^{-1} d\nu_\sigma * \alpha \\ &\le \frac{1}{4}\int f^2\, d\eta.\end{aligned}$$

Hence $\Gamma(u, u) \le 1$ if $\|f\|_{L^2(\eta)} \le 2$. Therefore we can choose f depending on $\lambda, \mu, \varepsilon$ and σ such that

$$\begin{aligned}u(\lambda) - u(\mu) &= \int f\left(\left(\varepsilon + \frac{d\nu_\sigma * \lambda}{d\eta}\right)^{1/2} - \left(\varepsilon + \frac{d\nu_\sigma * \mu}{d\eta}\right)^{1/2}\right) d\eta \\ &\ge 2(1-\varepsilon)\left\|\left(\varepsilon + \frac{d\nu_\sigma * \lambda}{d\eta}\right)^{1/2} - \left(\varepsilon + \frac{d\nu_\sigma * \mu}{d\eta}\right)^{1/2}\right\|_{L^2(\eta)}.\end{aligned}$$

Letting ε tend to 0 we conclude that $\rho(\lambda, \mu) \ge 2\, d(\nu_\sigma * \lambda, \nu_\sigma * \mu)$. The assertion will now follow from

$$\liminf_{\sigma \downarrow 0} d(\nu_\sigma * \lambda, \nu_\sigma * \mu) \ge d(\lambda, \mu). \tag{4.10}$$

To prove (4.10) suppose first that the support of λ is contained in the support of μ. Of course, the support of a measure should here be regarded with respect to the Euclidean topology on $E = \mathbb{R}$. In this case, Lemma 23 in Schied (1996) asserts that

$$d(\lambda, \mu) = \sup_{f \in C_c(\mathbb{R})} \big(\langle f, \lambda \rangle - \langle V_1 f, \mu \rangle \big), \tag{4.11}$$

where $C_c(\mathbb{R})$ denotes the space of continuous functions with compact support on $\mathbb{R}$ and V_t is, for $t \geq 0$, given by

$$V_t x = \begin{cases} \dfrac{x}{1 - tx}, & \text{if } x < t, \\ \infty, & \text{otherwise.} \end{cases} \tag{4.12}$$

Compare also with Lynch and Sethuraman (1987). Applying this variational formula to $d(\nu_\sigma * \lambda, \nu_\sigma * \mu)$ as well, we get with Jensen's inequality that

$$\begin{aligned} &\liminf_{\sigma \downarrow 0} d(\nu_\sigma * \lambda, \nu_\sigma * \mu) \\ &\quad \geq \sup_{n \in \mathbb{N}} \liminf_{\sigma \downarrow 0} \sup_{f \in C_c(\mathbb{R}),\, f \leq 1 - 1/n} \big(\langle f, \nu_\sigma * \lambda \rangle - \langle V_1 f, \nu_\sigma * \mu \rangle \big) \\ &\quad \geq \sup_{n \in \mathbb{N}} \liminf_{\sigma \downarrow 0} \sup_{f \in C_c(\mathbb{R}),\, f \leq 1 - 1/n} \big(\langle \nu_\sigma * f, \lambda \rangle - \langle V_1(\nu_\sigma * f), \mu \rangle \big). \end{aligned}$$

Now choose $\varepsilon > 0$, $n \in \mathbb{N}$ large enough and $g \in C_c(\mathbb{R})$ such that $g \leq 1 - 1/n$ and

$$\langle g, \lambda \rangle - \langle V_1 g, \mu \rangle \geq \sup_{f \in C_c(\mathbb{R}),\, f \leq 1 - 1/n} \big(\langle f, \lambda \rangle - \langle V_1 f, \mu \rangle \big) - \varepsilon.$$

Since $\nu_\sigma * g \to g$ pointwise as $\sigma \downarrow 0$ and $\nu_\sigma * g \leq 1 - 1/n$, for all $\sigma > 0$, dominated convergence yields that $\langle \nu_\sigma * g, \lambda \rangle - \langle V_1(\nu_\sigma * g), \mu \rangle \to \langle g, \lambda \rangle - \langle V_1 g, \mu \rangle$. Thus

$$\begin{aligned} &\liminf_{\sigma \downarrow 0} \sup_{f \in C_c(\mathbb{R}),\, f \leq 1 - 1/n} \big(\langle \nu_\sigma * f, \lambda \rangle - \langle V_1(\nu_\sigma * f), \mu \rangle \big) \\ &\qquad \geq \sup_{f \in C_c(\mathbb{R}),\, f \leq 1 - 1/n} \big(\langle f, \lambda \rangle - \langle V_1 f, \mu \rangle \big) \end{aligned}$$

and (4.10) is proved in the case where the support of λ is contained in the support of μ. To handle the general case, observe first that

$$d(\alpha, \beta) \leq \|\alpha - \beta\|_{\mathrm{var}}^{1/2}, \qquad \alpha, \beta \in M.$$

This can be seen as in part (i) of the proof of Proposition V.4.4 in Jacod and Shiryaev (1987). Hence, for σ, $\varepsilon > 0$,

$$\begin{aligned} d\big(\nu_\sigma * \mu, \nu_\sigma * (\mu + \varepsilon\lambda)\big) &\leq \|\nu_\sigma * \mu - \nu_\sigma * (\mu + \varepsilon\lambda)\|_{\mathrm{var}}^{1/2} \\ &\leq \|\nu_\sigma\|_{\mathrm{var}}^{1/2} \|\varepsilon\lambda\|_{\mathrm{var}}^{1/2} = \sqrt{\varepsilon \|\lambda\|_{\mathrm{var}}}\,. \end{aligned}$$

Since the support of λ trivially is contained in the support of $\mu + \varepsilon\lambda$, we get that

$$
\begin{aligned}
&\liminf_{\sigma\downarrow 0} d(\nu_\sigma * \lambda, \nu_\sigma * \mu)\\
&\qquad \geq \liminf_{\sigma\downarrow 0} d\big(\nu_\sigma * \lambda, \nu_\sigma *(\mu+\varepsilon\lambda)\big) - \limsup_{\sigma\downarrow 0} d\big(\nu_\sigma * \mu, \nu_\sigma *(\mu+\varepsilon\lambda)\big)\\
&\qquad \geq d(\lambda,\mu) - d(\mu, \mu+\varepsilon\lambda) - \sqrt{\varepsilon\|\lambda\|_{\mathrm{var}}} \geq d(\lambda,\mu) - 2\sqrt{\varepsilon\|\lambda\|_{\mathrm{var}}}.
\end{aligned}
$$

Since ε was arbitrary, (4.10) is proved.

This proves $\rho = \sqrt{2}\, d$ in the case where E is uncountable. If E is countable, E can be regarded as discrete subset of $\mathbb{R}$. Then $M(E) \hookrightarrow M(\mathbb{R})$, and hence every function $u \in C^1(M(\mathbb{R}))$ satisfying $\Gamma(u,u) \leq 1$ yields a function u' on $M(E)$ with the same property. Therefore

$$
\begin{aligned}
&\sup\big\{u'(\lambda) - u'(\mu) \mid u' \in C^1(M(E)), \Gamma(u', u') \leq 1\big\}\\
&\qquad \geq \sup\big\{u(\lambda) - u(\mu) \mid u \in C^1(M(\mathbb{R})), \Gamma(u,u) \leq 1\big\} = 2\, d(\lambda,\mu)
\end{aligned}
$$

and Theorem 1.2 is proved. □

PROOF OF THEOREM 1.1. Suppose $u \in C^1(M_1)$ satisfies $\overline{\Gamma}(u,u) \leq 1$ and $\omega \in \mathbf{H}$ is a curve such that $\omega(0) = \lambda$, $\omega(1) = \mu$ and $\omega(t) \in M_1$, for all $t \in [0,1]$. Then, as above,

$$
\begin{aligned}
u(\lambda) - u(\mu) &= \int_0^1 \langle \overline{D}u(\omega(t)), \dot{\omega}(t)\rangle\, dt\\
&\leq \int_0^1 \sqrt{\overline{\Gamma}(u,u)(\omega(t))}\left\| \frac{d\dot{\omega}(t)}{d\omega(t)}\right\|_{L^2(\omega(t))} dt \leq 2\, L(\omega).
\end{aligned}
$$

Choosing $\omega = \overline{\gamma}$ as defined in (2.10), we get

$$
\overline{\rho}(\lambda,\mu) := \sup\big\{u(\lambda) - u(\mu) \mid \overline{\Gamma}(u,u) \leq 1, u \in C^1(M_1)\big\} \leq 2\delta(\lambda,\mu).
$$

For the prove of $\overline{\rho} \geq 2\delta$ we may restrict ourselves as above to the case where $E = \mathbb{R}$. Choose, $\sigma, \varepsilon > 0$, let ν_σ denote the normal distribution with mean 0 and variance σ, let η denote the measure $\eta = (\lambda + \mu + \nu_1)/3$, and define f as (an η-version of)

$$
f = (1+\varepsilon)^{-2}\left(\varepsilon + \frac{d\eta_\sigma * \mu}{d\eta}\right)^{1/2}.
$$

Note that f is bounded, since $d\nu_\sigma * \mu/d\eta$ is bounded by some constant. With these choices define a function u on the whole of M as in (4.9), and let w denote the function

$$
w(\alpha) = \arccos\big(u(\alpha) - \sqrt{\varepsilon}\langle f, \eta\rangle\big), \qquad \alpha \in M_1.
$$

By Hölder's inequality,

$$
u(\alpha) \leq \|f\|_{L^2(\eta)}(\|\alpha\|_{\mathrm{var}} + \varepsilon) \leq 1 \qquad \forall\, \alpha \in M_1
$$

and so ω is well defined and in $C^1(M_1)$. For $\alpha \in M_1$ fixed, let φ denote the density $d\nu_\sigma * \alpha / d\eta$. Then we have that, for $\beta \in M_1$,

$$\langle \overline{D}w(\alpha), \beta \rangle = \frac{-1}{\sqrt{1 - (\cos w(\alpha))^2}} \langle Du(\alpha), \beta - \alpha \rangle$$

$$= \frac{-1}{2 \sin w(\alpha)} \Big\langle \nu_\sigma * \big[f(\varepsilon + \varphi)^{-1/2} \big], \beta - \alpha \Big\rangle.$$

Therefore, with C denoting $(2 \sin w(\alpha))^{-2}$,

$$\langle (\overline{D}w(\alpha))^2, \alpha \rangle = C\Big(\Big\langle \Big(\nu_\sigma * \big[f(\varepsilon + \varphi)^{-1/2} \big] \Big)^2, \alpha \Big\rangle - \Big\langle f(\varepsilon + \varphi)^{-1/2}, \nu_\sigma * \alpha \Big\rangle^2 \Big)$$

$$\leq C\Big(\Big\langle f^2(\varepsilon + \varphi)^{-1}, \nu_\sigma * \alpha \Big\rangle - \Big\langle f(\varepsilon + \varphi)^{-1/2} \varphi, \eta \Big\rangle^2 \Big)$$

$$\leq C\Big(\langle f^2, \eta \rangle - \Big(\Big\langle f(\varepsilon + \varphi)^{1/2}, \eta \Big\rangle - \varepsilon \Big\langle f(\varepsilon + \varphi)^{-1/2}, \eta \Big\rangle \Big)^2 \Big).$$

Now

$$\varepsilon \Big\langle f(\varepsilon + \varphi)^{-1/2}, \eta \Big\rangle \leq \sqrt{\varepsilon} \langle f, \eta \rangle \leq \Big\langle f(\varepsilon + \varphi)^{1/2}, \eta \Big\rangle = u(\alpha),$$

and hence

$$\Big\langle (\overline{D}w(\alpha))^2, \alpha \Big\rangle \leq \frac{1 - \big(u(\alpha) - \sqrt{\varepsilon} \langle f, \eta \rangle \big)^2}{4 \sin^2 w(\alpha)} = \frac{1}{4}$$

Hence $\overline{\Gamma}(2w, 2w) \leq 1$ and $\overline{\rho}(\lambda, \mu) \geq 2w(\lambda) - 2w(\mu)$. Letting ε tend to 0 we arrive at

$$\overline{\rho}(\lambda, \mu) \geq 2\delta(\nu_\sigma * \lambda, \nu_\sigma * \mu).$$

But from (4.6) and (4.10),

$$\lim_{\sigma \downarrow 0} \delta(\nu_\sigma * \lambda, \nu_\sigma * \mu) = \liminf_{\sigma \downarrow 0} \arccos\Big(1 - d(\nu_\sigma * \lambda, \nu_\sigma * \mu)^2 / 2\Big) \geq \delta(\lambda, \mu).$$

Hence the theorem is proved. □

REMARK. The choice of the particular function w in the above proof can be guessed by expressing the *carré du champs* operator Γ in terms of "polar coordinates." That is, we write $\mu \in M$ as $\mu = r \cdot \hat{\mu}$, where $r \geq 0$ and $\hat{\mu} \in M_1$, and we consider a function u on $[0, \infty) \times M_1$ as function on M by writing $u(\mu) = u(r, \hat{\mu})$. Then, at least formally,

$$\Gamma(u, u)(\mu) = r\left(\frac{\partial}{\partial r} u(r, \hat{\mu}) \right)^2 + \frac{1}{r} \overline{\Gamma}(u(r, \cdot), u(r, \cdot))(\hat{\mu}).$$

The r appearing on the right-hand side can be eliminated by choosing $u(r) = r^{1/2} u_0(\hat{\mu})$, and we then get that

$$\Gamma(u, u)(\mu) = u_0(\hat{\mu})^2 / 4 + \overline{\Gamma}(u_0, u_0)(\hat{\mu}).$$

So if $w \in C^1(M_1)$ is some function such that $\overline{\Gamma}(w, w) \leq 1$, we can choose $u_0 = 2\cos(w/2)$ to get $\Gamma(u, u) \leq 1$. This argument is close to those that can be used to derive the results of Etheridge and March (1991) and Perkins (1991).

PROOF OF PROPOSITION 1.3. We will only prove the part of the proposition concerned with ρ. The assertion for $\bar{\rho}$ can be obtained analogously.

Again we can put ourselves in the situation where $E = \mathbb{R}$ and B is its Borel field. It suffices to show that, for ν and $\mu \in M$ fixed, we can find a sequence $(u_n) \subset FC_b^\infty(\mathbf{D})$ such that $\Gamma(u_n, u_n) \le 1$ and $u_n(\nu) \to u(\nu)$ and $u_n(\mu) \to u(\mu)$ if u is given by (4.9). To this end let f, ε, and η be the same as in (4.9), and choose pairwise disjoint sets $A_1, \dots, A_n \in B$ and $g_1, \dots, g_n \in \mathbf{D}$ such that $0 \le g_i \le 1$ $(i = 1, \dots, n)$. Now define $v_i \in FC_b^\infty(\mathbf{D})$ $(i = 1 \dots n)$ by

$$v_i(\alpha) = \int_{A_i} f\left(\varepsilon + \frac{\langle g_i, \alpha \rangle}{\eta(A_i)}\right)^{1/2} d\eta, \qquad \alpha \in M.$$

Then

$$Dv_i(\alpha, x) = \tfrac{1}{2}\langle fI_{A_i}, \eta\rangle\big(\varepsilon\eta(A_i)^2 + \eta(A_i)\langle g_i, \alpha\rangle\big)^{-1/2} g_i(x).$$

Now let w be given by $w = \sum_{i=1}^n v_i$. We get

$$\Gamma(w, w)(\alpha) \le \frac{1}{4} \sum_{i,j=1}^{n} \frac{\langle fI_{A_i}, \eta\rangle \langle fI_{A_j}, \eta\rangle}{\sqrt{\eta(A_i)\eta(A_j)}} \frac{\langle g_i g_j, \alpha\rangle}{\sqrt{\langle g_i, \alpha\rangle \langle g_j, \alpha\rangle}},$$

but $0 \le g_i \le 1$ and hence $\langle g_i g_j, \alpha\rangle \le \sqrt{\langle g_i, \alpha\rangle\langle g_j, \alpha\rangle}$ $(i, j = 1, \dots, n)$. Thus

$$\Gamma(w, w)(\alpha) \le \frac{1}{4}\left(\sum_{i=1}^{n} \frac{\langle fI_{A_i}, \eta\rangle}{\sqrt{\eta(A_i)}}\right)^2 \le \frac{1}{4}\int f^2 \, d\eta.$$

By our assumptions on $\mathbf{D}$ we can find functions $\tilde{g}_i$ that coincide η-a.e. (and hence ν and μ-a.e.) with $\nu_\sigma * I_{A_i}$, and that are bounded pointwise limits of sequences $(g_i^k) \subset \mathbf{D}$ with $0 \le g_i^k \le 1$ $(i = 1, \dots, n,\ k \in \mathbb{N})$. The function w^k corresponding to $(g_1^k, \dots, g_n^k)$ then satisfies $\Gamma(w^k, w^k) \le \|f\|_{L^2(\eta)}^2/4$, and (w^k) converges pointwise to some function $\tilde{w}$. For $\alpha \ll \eta$ and in particular for $\alpha \in \{\nu, \mu\}$, $\tilde{w}(\alpha)$ takes the form

$$\tilde{w}(\alpha) = \int f\left(\varepsilon + \frac{d\nu_\sigma * \alpha}{d\eta}\bigg|_{\sigma(A_1, \dots, A_n)}\right)^{1/2} d\eta.$$

But now martingale convergence implies that $u(\alpha)$ given in (4.9) can be approximated by functions $\tilde{w}(\alpha)$ as above when $\sigma(A_1, \dots, A_n) \to B$. This proves the part of the proposition concerned with ρ. □

PROOF OF THEOREM 3.1. We will show first that under the assumptions of Theorem 3.1,

$$\mu(N) = d(\mu, \mu_{N^c})^2 = \inf\big\{I_\mu^0(\omega) \mid \omega(1)(N) = 0\big\} = I_\mu^0(\gamma). \tag{4.13}$$

Indeed, the first equality is trivial. Obviously, by (2.6),

$$\begin{aligned} d(\mu, \mu_{N^c})^2 &= \inf\big\{d(\mu, \nu)^2 \mid \nu(N) = 0\big\} \\ &\le \inf\big\{I_\mu^0(\omega) \mid \omega(1)(N) = 0\big\} \le I_\mu^0(\gamma). \end{aligned}$$

But γ is the unique minimizing geodesic from μ to μ_{N^c}, and therefore $I_\mu^0(\gamma) = 2E(\gamma) = d(\mu, \mu_{N^c})^2$. Hence (4.13) is proved.

LEMMA 4.2. *Under the assumptions of Theorem* 3.1,

$$\lim_{\varepsilon \downarrow 0} \varepsilon \log \mathbb{P}_\mu[X_\varepsilon(N) = 0] = -\mu(N).$$

PROOF. Let $P_t(x, A) = (2\pi t)^{-d/2} \int_A \exp(-|x-y|^2/2t)\, dy$ denote the Brownian transition semigroup. Theorem 9 of Schied (1996) implies that, for $t \geq 0$ and f bounded and measurable,

$$\langle V_t P_t f, \mu \rangle \leq \log \mathbb{E}_\mu[\exp(\langle f, X_t \rangle)] \leq \langle P_t V_t f, \mu \rangle,$$

where V_t is defined in (4.12). Using the lower bound we get that

$$\begin{aligned} \log \mathbb{P}_\mu[X_\varepsilon(N) = 0] &= \lim_{\lambda \uparrow \infty} \log \mathbb{E}_\mu[\exp(-\lambda X_\varepsilon(N))] \\ &\geq -\lim_{\lambda \uparrow \infty} \int \frac{\lambda P_\varepsilon(x, N)}{1 + \varepsilon \lambda P_\varepsilon(x, N)}\, \mu(dx) \\ &= -\frac{1}{\varepsilon} \int P_\varepsilon(x, N)\, \mu(dx). \end{aligned}$$

Hence

$$\liminf_{\varepsilon \downarrow 0} \varepsilon \log \mathbb{P}_\mu[X_\varepsilon(N) = 0] \geq \limsup_{\varepsilon \downarrow 0} \int P_\varepsilon(x, N)\, \mu(dx) \geq -\mu(N),$$

because N is closed. On the other hand, $\limsup_{\varepsilon \downarrow 0} \varepsilon \log \mathbb{P}_\mu[X_\varepsilon(N) = 0] \leq \mu(N)$ follows from (4.13) and the upper bound of the large deviation principle stated in connection with (3.2). □

Now suppose $A \subset C([0,1] : M_p(\mathbb{R}^n))$ is closed. Let A_N denote the closed set $A_N = A \cap \{\omega \mid \omega(1)(N) = 0\}$. Then, by the above large deviation principle and Lemma 4.2,

$$\begin{aligned} \limsup_{\varepsilon \downarrow 0} \varepsilon \log \mathbb{P}_\mu[X^\varepsilon \in A \mid X_1^\varepsilon(N) = 0] &= \limsup_{\varepsilon \downarrow 0} \varepsilon \log \mathbb{P}_\mu[X^\varepsilon \in A_N] + \mu(N) \\ &\leq \mu(N) - \inf_{\omega \in A_N} I_\mu^0(\omega), \end{aligned}$$

but by Theorem 2.1 and (4.13) we have that

$$\inf_{\omega \in A_N} I_\mu^0(\omega) > \mu(N) = I_\mu^0(\gamma) \quad \Leftrightarrow \quad \gamma \notin A_N.$$

This shows that

$$\limsup_{\varepsilon \downarrow 0} \mathbb{P}_\mu[X^\varepsilon \in A \mid X_1^\varepsilon(N) = 0] \leq \delta_\gamma(A) \quad \forall \text{ closed } A \subset C([0,1] : M_p(\mathbb{R}^n))$$

and this is equivalent to the asserted weak convergence. □

REFERENCES

AMARI, S. (1985). Differential-geometrical methods in statistics. *Lecture Notes in Statist.* **28**. Springer, Berlin.

BALLMANN, W. (1990). Singular spaces of non-positive curvature. In *Sur les groupes Hyperboliques d'aprés Mikhael Gromov* 189–202. Birkhäuser, Boston.

CARLEN, E. A., KUSUOKA, S. and STROOK, D. W. (1987). Upper bounds for symmetric Markov transition functions. *Ann. Inst. H. Poincaré* **2** 245–287.

DAVIES, E. B. (1989). *Heat Kernels and Spectral Theory*. Cambridge Univ. Press.

DAWSON, D. A. (1993). Measure-valued Markov processes. *Ecole d'Eté de Probabilités de Saint-Flour XXI. Lecture Notes in Math*. **1541** 1–260. Springer, Berlin.

DUNFORD, N. and SCHWARTZ, J. (1967). *Linear Operators*. Interscience, New York.

ETHERIDGE, A. and MARCH, P. (1991). A note on superprocesses. *Probab. Theory Related Fields* **89** 141–147.

ETHIER, S. N and KURTZ, T. G. (1986). *Markov Processes—Characterization and Convergence*. Wiley, New York.

HILLE, E. and PHILLIPS, R. (1957). *Functional Analysis and Semi-groups*. Amer. Math. Soc., Providence, RI.

IKEDA, N. and WATANABE, S. (1989). *Stochastic Differential Equations and Diffusion Processes*, 2nd ed. North-Holland, Amsterdam.

JACOD, J. and SHIRYAEV, A. N. (1987). *Limit Theorems for Stochastic Processes*. Springer, Berlin.

KONNO, N. and SHIGA, T. (1988). Stochastic partial differential equations for some measure-valued diffusions. *Probab. Theory Related Fields* **79** 201–225.

KUWAE, K. and UEMURA, T. (1995). Weak convergence of symmetric diffusion processes. *Probab. Theory Related Fields*. To appear.

LYNCH, J. and SETHURAMAN, J. (1987). Large deviations for processes with independent increments. *Ann. Probab.* **15** 610–627.

OVERBECK, L. and RÖCKNER, M. (1996). Geometric aspects of finite and infinite dimensional Fleming–Viot processes. In *Random Operators and Stochastic Equations*. To appear.

PARTHASARATHY, K. R. (1967). *Probability Measures on Metric Spaces*. Academic Press, New York.

PERKINS, E. (1991). Conditional Dawson–Watanabe processes and Fleming–Viot processes. In *Seminar on Stochastic Processes Progr. Probab.* **29** 143–156. Birkhäuser, Basel.

SCHIED, A. (1996). Sample path large deviations for super-Brownian motion. *Probab. Theory Related Fields* **104** 319–348.

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

STURM, K.-TH. (1994). Analysis on local Dirichlet spaces I. Recurrence, conservativeness and L^p-Liouville properties. *J. Reine Angew. Math.* **456** 173–196.

STURM, K.-TH. (1995). On the geometry defined by Dirichlet forms. In *Seminar on Stochastic Analysis, Random Fields and Applications Prog. Probab.* **36** 231–242. Birkäuser, Basel.

INSTITUT FÜR MATHEMATIK
HUMBOLDT-UNIVERSITÄT
UNTER DEN LINDEN 6
D10099 BERLIN
GERMANY
E-MAIL: schied@mathematik.hu-berlin.de