The Annals of Probability
2003, Vol. 31, No. 3, 1679–1712

COMPARISONS FOR MEASURE VALUED PROCESSES WITH INTERACTIONS

BY SAUL JACKA AND ROGER TRIBE

University of Warwick

This paper considers some measure-valued processes $\{X_t : t \in [0,T]\}$ based on an underlying critical branching particle structure with random branching rates. In the case of constant branching these processes are Dawson–Watanabe processes. Sufficient conditions on functionals Φ of the process are given that imply that the expectations $E(\Phi(X_T))$ are comparable to the constant branching case. Applications to hitting estimates and regularity of solutions are discussed. The result is established via the martingale optimality principle of stochastic control theory. Key steps, which are of independent interest, are the proof of a version of Itô's lemma for $\Phi(X_t)$, suitable for a large class of functions of measures (Theorem 3) and the proof of various smoothing properties of the Dawson–Watanabe transition semigroup (Section 3).

1. Introduction and statement of results.

1.1. *Introduction.* We start by describing the stochastic processes that we study, which will be solutions to a certain martingale problem. Let E be a compact metric space, with Borel sigma field $\mathcal{E}$, and let $\mathcal{M}$ be the space of finite Borel measures on $(E, \mathcal{E})$, on which we put the topology of weak convergence. We write either (μ, f) or $\mu(f)$ for the integral of a function $f : E \to \mathbb{R}$ with respect to a measure $\mu \in \mathcal{M}$, whenever this is well defined. Let $C(E)$ [resp. $B(E)$] be the space of continuous (resp. bounded measurable) functions on E, with the supremum norm $\|f\|_E$, and let A be the generator of a strongly continuous Markov semigroup $\{P_t : t \geq 0\}$ on $C(E)$.

Suppose $\{X_t : t \in [0,T]\}$ is an adapted process defined on a filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$ that has continuous paths with values in $\mathcal{M}$. The terminal time T will be fixed throughout the paper. Let $\mathcal{P}$ be the predictable sets for this probability space. Let $\sigma : [0,T] \times \Omega \times E \to [0,\infty)$ be $\mathcal{P} \otimes \mathcal{E}$ measurable. We call $\{X_t\}$ a solution to the martingale problem $M(A,\sigma)$ if for all $\phi \in D(A)$ the process

$$Z_t(\phi) = (X_t, \phi) - (X_0, \phi) - \int_0^t (X_s, A\phi)\, ds \tag{1}$$

Received March 2002; revised July 2002.
AMS 2000 subject classifications. Primary 60G57; secondary 49K27, 60H15, 60J55, 60J80, 60K35, 60H99, 93C25.
Key words and phrases. Interaction, stochastic control, Itô's lemma, random branching, clustering.

is an $\mathcal{F}_t$ local martingale for $t \in [0, T]$ with quadratic variation

$$[Z(\phi)]_t = \int_0^t (X_s, \sigma_s \phi^2)\, ds. \tag{2}$$

One may think of X_t as a measure describing the position of an infinite cloud of infinitesimal particles that are independently moving according to the process with generator A, and that are continuously dying and branching into two, each at rate $\sigma/2$. We emphasize that $\sigma_t(x)$, the particle branching rate at time t and position x, may be random, for instance it may depend on the position of the other particles. This has the potential for modeling many aspects of populations, for example competition, mutualism or clustering. The convergence of particle systems to a measure valued limit satisfying $M(A, \sigma)$ is well known for constant branching rates (see [4], Section 4.6) and has been shown for some interacting branching rates σ in [8]. Our arguments will need only the martingale problem and will not use this associated particle picture.

The solutions for a constant branching rate $\sigma_t^c \equiv c$ are called Dawson–Watanabe processes and have been extensively studied (see [4]). The martingale problem $M(A, \sigma^c)$ has solutions that are unique in law. The constant branching rate means that disjoint sets of particles evolve independently. This makes the process quite tractable and a large number of qualitative properties have been established. Less is known about the processes with random branching rates σ. Existence of solutions to the martingale problem, studied in [7, 9–11], holds for a large class of branching rates. Uniqueness is typically unknown. Uniqueness for an extended version of the martingale problem (called a historical martingale problem) has been established in [11] for a restricted class of branching rates σ.

The aim of this paper is to establish a comparison principle for expectations $E(\Phi(X_t))$ of certain functionals $\Phi : \mathcal{M} \to [0, \infty]$. When the branching rate satisfies $\sigma \geq c$ we find conditions on Φ that ensure the expectation $E(\Phi(X_t))$ is greater than the corresponding expectation for the Dawson–Watanabe process with constant branching rate c. There is a corresponding result when $\sigma \leq c$. The intuition is that more branching should lead to more clustering which should lead to certain functionals increasing in expectation. This leads to easy proofs that certain properties of Dawson–Watanabe processes carry over to interactive branching processes. This result should be compared to those of Cox, Fleischmann and Greven [3], who studied a similar problem for functionals of systems of stochastic differential equations (SDEs) on a lattice and applied it to establish ergodic properties. By taking our processes to have a motion process on the lattice our results apply to systems of SDEs. The results of Cox, Fleischmann and Greven are then more general in that they treat the case of two comparable branching rates $\sigma_1 \leq \sigma_2$, although our results allow a greater class of interactions.

The analogous problem of comparing functionals of processes with different drift terms, for example, with the term $+\int_0^t (b_s, \phi)\, ds$ added into the martingale problem (1), can be treated via pathwise comparison results. These allow one to couple two processes if one has a larger drift than the other, from which one can deduce comparisons between the expectations of increasing functionals. There are various pathwise comparison arguments in the literature for SPDEs (see [1] and its bibliography). For measure valued branching processes when the drift terms come from immigration, or from mass creation and annihilation terms [i.e., they are of the form $\int_0^t (X_s, b_s\phi)\,ds$], a coupling can be constructed via a "thinning" procedure (see [2], Theorem 5.1, for a related result).

1.2. *Statement of main result.* We now discuss the hypotheses for the theorem. We will need two mild assumptions on the branching rate and on the underlying spatial motion generated by A, for which we give the following two definitions:

1. The branching rate σ is called *locally bounded* if there exist stopping times $T_n \uparrow \infty$ so that $\sigma_t I(t < T_n)$ are bounded, as functions on $[0, T] \times \Omega \times E$, for each n.
2. The generator A is called a *good generator* if there is a dense linear subspace D_0 of $C(E)$ that is an algebra and is closed under the mappings P_t for all $t \geq 0$.

The assumption that σ is locally bounded ensures that the integral in (2) is well defined. Most commonly studied motion processes have good generators. Without loss of generality we may, and shall, assume that D_0 contains the constant functions. Using a lemma of Watanabe (see [6], Proposition 1.3.3) the conditions on D_0 imply that D_0 is a core for A. Recall that D_0 is a core for $D(A)$ if whenever $f \in D(A)$ there exist $f_n \in D_0$ so that $f_n \to f$ and $Af_n \to Af$.

The key hypothesis on Φ is the following convexity hypothesis:

$$E\big(\Phi(\mu + Z + \bar{Z}) - \Phi(\mu + Z) - \Phi(\mu + \bar{Z}) + \Phi(\mu)\big) \geq 0 \tag{3}$$

for all $\mu \in \mathcal{M}$ and for all i.i.d. $\mathcal{M}$ valued variables $Z, \bar{Z}$ with bounded total mass. This is a randomized version of the following parallelogram condition:

$$\Phi(\mu + \nu + \eta) - \Phi(\mu + \nu) - \Phi(\mu + \eta) + \Phi(\mu) \geq 0 \qquad \text{for all } \mu,\ \nu,\ \eta \in \mathcal{M}. \tag{4}$$

Clearly (4) implies (3). In the case that Φ has two continuous directional derivatives, as defined by (10) in Section 2, the condition (4) is equivalent to $D_{xy}\Phi(\mu) \geq 0$ for all $x, y \in E$, $\mu \in \mathcal{M}$. Example 5, in Section 4, does not satisfy (4), but (3) applies. Furthermore, Example 6 in Section 4 shows that Φ being convex is not a sufficiently strong hypothesis.

We require one more hypothesis on the smoothing properties of the underlying motion process. We suppose the motion semigroup $\{P_t\}$ satisfies $P_t f \in D(A)$ for $t > 0$, $f \in B(E)$ and that there exists $\alpha < \infty$ and $\beta \in [0, 2^{1/2})$ so that

$$\|AP_t f\|_E \leq \alpha t^{-\beta} \|f\|_E \qquad \text{for all } t \in [0, T],\ f \in B(E). \tag{5}$$

Here is the main result of the paper.

THEOREM 1. *Suppose $\{X_t\}$ is a solution to the martingale problem $M(A,\sigma)$ for a locally bounded branching rate σ and good generator A satisfying the smoothing hypothesis* (5). *Suppose that $\Phi : \mathcal{M} \to [0,\infty)$ is continuous, satisfies hypothesis* (3) *and the growth condition*:

$$\Phi(\mu) \le \exp(\lambda(\mu, 1)) \qquad \text{for some } \lambda < 1/cT \text{ and } C < \infty. \tag{6}$$

Let $\{Y_t\}$ be a solution to the problem $M(A,\sigma^c)$, that is a Dawson–Watanabe process with the constant branching rate c, and whose initial condition Y_0 has the same law as X_0. Then the following comparisons hold:

(a) *if $0 \le \sigma \le c$ then $E(\Phi(X_t)) \le E(\Phi(Y_t))$ for all $t \in [0,T]$*;
(b) *if $c \le \sigma$ and Φ is bounded then $E(\Phi(X_t)) \ge E(\Phi(Y_t))$ for all $t \in [0,T]$*;
(c) *if $c \le \sigma \le \bar{c}$ and $\lambda < 1/2\bar{c}T$ then $E(\Phi(X_t)) \ge E(\Phi(Y_t))$ for all $t \in [0,T]$.*

REMARKS. The continuity and growth conditions on Φ are certainly not necessary, and can often be weakened. For example, if the conclusions of the theorem hold for a convergent sequence of functions Φ_n then it is often possible to deduce that they hold for the limit.

The smoothing hypothesis (5) on the underlying motion process should be totally unnecessary, and hence we have not sought a best possible bound on β. However, the hypothesis is satisfied by the Laplacian with the value $\beta = 1$ and this is sufficient for all our examples in Section 4. We make some more remarks on this at the end of Section 3.

A similar result, under the same hypotheses, holds for path functionals $\int_0^T f(t)\Phi(X_t)\,dt$ where $f \ge 0$. Since the comparison holds for each $E(\Phi(X_t))$ it must hold for the integral.

The sketch proof below makes it clear that it is enough for the convexity hypothesis (3) to hold at all μ in the range of X_t, that is on any set $\mathcal{M}_0 \subseteq \mathcal{M}$ for which $X_t \in \mathcal{M}_0$ for all $t \le T$ almost surely.

We now give a sketch of the method used for the proof of this result, restricting for simplicity to the case $0 \le \sigma \le c$. We write $\{U_t^c : t \ge 0\}$ for the transition semigroup of the Dawson–Watanabe process with constant branching rate c. The conclusions of the theorem, for example, part (a), can then be rewritten as

$$E(\Phi(X_t)) \le E\big(U_t \Phi(X_0)\big) \qquad \text{for all } t \in [0,T].$$

We shall use the ideas of control theory. We consider σ as a control and try to maximize the value of $E(\Phi(X_t))$ over all controls bounded above by a constant c. Under our hypotheses on Φ the constant control σ^c is optimal and so we define the value function, the reward under the optimal control, by

$$F(t,\mu) = E(\Phi(Y_t)) = U_t^c \Phi(\mu).$$

The martingale optimality argument is the heuristic that, if X_s is a solution to $M(A,\sigma)$, then the process $s \to F(s, X_{t-s})$ is a supermartingale if $\sigma \le c$ and a martingale for the constant branching case $\sigma = c$ implying

$$E(\Phi(X_t)) = E(F(0, X_t)) \ge E(F(t, X_0)) = E\big(U_t^c \Phi(X_0)\big)$$

which is the desired conclusion. To implement this idea we need the drift in the semimartingale decomposition for a process $F(s, X_s)$. In Section 2 we show, for a general class of functions $F(s,\mu)$, that this is given by $\int_0^t L^\sigma F(s, X_s)\, ds$ where

$$L^\sigma F(s,\mu) = D_s F(s,\mu) + \int_E \big(A^{(x)} D_x F(s,\mu) + \sigma_s(x) D_{xx} F(s,\mu)\big)\, \mu(dx). \tag{7}$$

The derivatives D_x and D_{xx} are first and second derivatives in the direction of the point mass δ_x, as defined in Section 2, and we write $A^{(x)}$ to indicate the variable on which the operator A is acting. This formula is well known and easy to establish for certain simple explicit functionals F. We establish it for F which only need to satisfy certain smoothness conditions. When σ takes the constant value c this gives a formula for the generator L^c of the Dawson–Watanabe process acting on smooth functions. Comparing L^σ with L^c we see they differ only in the term involving the second directional derivative.

Using the semigroup property of U_t^c one expects that

$$L^c F(T-s,\mu) = 0 \qquad \text{for all } \mu \in \mathcal{M} \text{ and } 0 < s \le T. \tag{8}$$

Then suppose X is a solution to $M(\sigma, A)$. Formally, we expect

$$\begin{aligned}
&E(\Phi(X_t)) - E\big(U_t^c \Phi(X_0)\big)\\
&\qquad = E(F(0, X_t)) - E(F(t, X_0))\\
&\qquad = E\bigg(\int_0^t L^\sigma F(t-s, X_s)\, ds\bigg)\\
&\qquad = E\bigg(\int_0^t L^c F(t-s, X_s)\, ds\bigg)\\
&\qquad\qquad + E\bigg(\int_0^t \int \big(\sigma_s(x) - c\big) D_{xx} F(t-s, X_s) X_s(dx)\, ds\bigg)\\
&\qquad = E\bigg(\int_0^t \int \big(\sigma_s(x) - c\big) D_{xx} F(t-s, X_s) X_s\,(dx)\, ds\bigg) \qquad \text{[using (8)]}.
\end{aligned}$$

So what is needed to complete the proof is that

$$D_{xx} F(s,\mu) = D_{xx} U_s^c \Phi(\mu) \ge 0. \tag{9}$$

We will show that the convexity hypothesis (3) implies this by finding a representation, in Section 3, for the derivative $D_{xx} U_s^c \Phi(\mu)$.

The main technical difficulty in implementing this heuristic proof is that we do not know whether the value function $U_t^c \Phi(\mu)$ satisfies the smoothness assumptions required to apply the formula (7). In Section 3 we investigate smoothing properties of the Dawson–Watanabe transition semigroup $\{U_t^c\}$ and show that directional derivatives always exist. In Section 4 we complete the proof of Theorem 1 and we give a number of examples of functions Φ satisfying all the required assumptions. We use the one point compactification of a locally compact space to show how our results apply to processes on $\mathbb{R}^d$ and $\mathbb{Z}^d$. We then choose suitable functionals Φ to establish several properties of interacting measure-valued processes that are already known for the constant branching case, such as local extinction, hitting estimates and absolute continuity or singularity of the measures. The heuristic that "more branching leads to more clustering" always holds true. In many cases the proofs of these properties for the Dawson–Watanabe process would carry over to the interacting processes. However application of a comparison argument, when applicable, is very simple. It would be good to find a comparison result for two random branching rates that are comparable, $\sigma_1 \le \sigma_2$, as may occur when there is a scalar parameter in front of an interacting branching mechanism. However our control theory argument, which we felt was a natural approach to the problem, fell foul of the problem of smoothing an infinite dimensional value function. The present proof uses the fact that we are comparing with a constant branching rate to show the smoothness of the value function, which, given integrability, follows from explicit formulae for the required derivatives, established by exploiting the branching property for Dawson–Watanabe processes.

We end this section with a moment estimate, useful throughout the paper, for the total mass $(X_t, 1)$ of solutions to $M(A, \sigma)$.

LEMMA 2. *Suppose $\{X_t\}$ is a solution to the martingale problem $M(A,\sigma)$ satisfying $(X_0, 1) \le K$ and $\sigma \le L$, almost surely. Then*

$$E\Big(\sup_{t\le T} \exp\big((X_t, 1)/LT\big)\Big) \le 4\exp(2K/LT).$$

PROOF. The stopping times $T_n = \inf\{t : (X_t, 1) \ge n\}$ reduce the local martingales $Z_t(1)$, as follows from (2) and the bound on σ. Set $E_t = \exp(2(X_t, 1)/L(T+t))$. Using Itô's formula and the martingale problem $M(A,\sigma)$, we have

$$dE_t \le \big(2E_t/L(T+t)\big)\, dZ_t(1).$$

The right-hand side is also reduced by T_n so by optional stopping, for $t \le T$,

$$\begin{aligned} E\big(\exp\big((X_{t\wedge T_n}, 1)/LT\big)\big) &\le E(E_{t\wedge T_n}) \le E(E_0) \\ &= E\big(\exp\big(2(X_0, 1)/LT\big)\big) \le \exp(2K/LT). \end{aligned}$$

Itô's formula also implies that $\exp(\lambda(X_{t\wedge T_n}, 1))$ is a submartingale for any λ. So by Doob's L^2 inequality we have

$$E\Big(\sup_{t\le T} \exp\big((X_{t\wedge T_n}, 1)/LT\big)\Big) \le 4\exp(2K/LT).$$

Letting $n \to \infty$ and applying monotone convergence completes the proof. □

2. Semimartingale decompositions. In this section we suppose $\{X_t\}$ is a solution to the martingale problem $M(A,\sigma)$ for a locally bounded branching rate σ and good generator A. We shall show that for sufficiently smooth functions $F(t,\mu)$ the process $F(t,X_t)$ is a semimartingale and give an expression for the finite variation part.

We define the first directional derivatives $D_xF : E \times \mathcal{M} \to \mathbb{R}$ by

$$D_xF(\mu) = \lim_{\varepsilon\to 0} \frac{1}{\varepsilon}\big(F(\mu+\varepsilon\delta_x) - F(\mu)\big). \tag{10}$$

We will need only these derivatives in the directions of point masses. When $D_xF(\mu)$ is continuous, directional derivatives in the direction of a general element $\nu \in \mathcal{M}$ can be expressed in terms of the function $D_xF(\mu)$ [see Lemma 4(a)]. We define second directional derivatives $D_{xy}F : E^2 \times \mathcal{M} \to \mathbb{R}$ by taking a further derivative so that $D_{xy}F = D_xD_yF$. For derivatives in time and mixed derivatives we write D_sF, $D_{sx}F$, $D_{sxy}F$. If the mixed derivatives are continuous then they may be taken in any order. We write $A^{(x)}$ or $P_t^{(x)}$ for the generator or semigroup applied in the variable x, whenever the action is unclear.

THEOREM 3. *Suppose* $F : [0,T] \times \mathcal{M} \to \mathbb{R}$ *satisfies*:

(i) *the functions* $F, D_xF, D_{xy}F, D_{xyz}F, D_sF, D_{sx}F, D_{sxy}F, D_{sxyz}F$ *exist and are continuous*,

(ii) *for fixed* s, y, z, μ *the maps* $x \to D_xF(s,\mu)$, $x \to D_{xy}F(s,\mu)$, $x \to D_{xyz}F(s,\mu)$ *are in the domain of the generator* A,

(iii) *the functions* $A^{(x)}D_xF(s,\mu)$, $A^{(x)}D_{xy}F(s,\mu)$, $A^{(x)}D_{xyz}F(s,\mu)$ *are continuous in* s, x, y, z, μ.

Then $F(t,X_t) - \int_0^t L^\sigma F(s,X_s)\,ds$ *is a local* $(\mathcal{F}_t)$ *martingale for* $t \in [0,T]$ *where* $L^\sigma F$ *is given by* (7), *which can be written in short as* $L^\sigma F = D_sF + (A^{(x)}D_xF + \sigma D_{xx}F, \mu(dx))$.

REMARKS. In the case of constant branching $\sigma = \sigma^c$, the theorem shows that (7) is a deterministic formula for the Markov generator of the Dawson–Watanabe process acting on suitably smooth F. The Dawson–Watanabe semigroup can be generated (see [14]) via a Trotter product formula that mixes the semigroup due to pure branching and the dual semigroup P_t^* describing the heat flow of measures. Therefore we expect the generator to be the sum of the two corresponding generators and the last two terms of (7) can be identified as such (for the heat flow see Lemma 7).

It would be natural to require hypotheses only on those derivatives that are involved in the expression $L^\sigma F$. The reason for requiring $D_{xyz}F$ in the domain $D(A)$ and not just D_xF is that we shall approximate the worst derivatives $D_{xyz}F$ and $A^{(x)}D_{xyz}F$ first and then integrate up to get all the lesser derivatives.

The rest of this section contains the proof of this result. Formula (7), at least as a formal expression, is well known. This is presumably based on the fact that it is easy to verify for a class of simple functions F, as we now show. Let $C^1([0,T])$ be the space of bounded functions $\psi : [0,T] \to \mathbb{R}$ with one bounded continuous derivative. Suppose that $\psi \in C^1([0,T])$, $\phi^i \in D(A)$ for $i = 1, \dots, n$ and define

$$F(\mu, t) = \psi(t) \prod_{i=1}^{n} (\mu, \phi_i).$$

Notice that we may obtain the formula for the covariation of $Z(\phi_1)$ and $Z(\phi_2)$:

$$[Z(\phi_1), Z(\phi_2)]_t = \int_0^t (X_s, \sigma_s \phi_1 \phi_2)\, ds, \tag{11}$$

from equation (2) by polarisation. Then applying Itô's formula, using (11) and the decompositions (1) and (2), we have that

$$\begin{aligned} F(X_t, t) - F(X_0, 0) &= \int_0^t D_s \psi(s) \prod_{i=1}^{n} (X_s, \phi_i)\, ds \\ &\quad + \int_0^t \psi(s) \sum_{i=1}^{n} \left(\prod_{j \neq i} (X_s, \phi_j) \right) (X_s, A\phi_i)\, ds \\ &\quad + \int_0^t \psi(s) \sum_{i,j=1, j \neq i}^{n} \left(\prod_{k \neq i,j} (X_s, \phi_k) \right) (X_s, \sigma_s \phi_i \phi_j)\, ds \\ &\quad + \int_0^t \psi(s) \sum_{i=1}^{n} \left(\prod_{j \neq i} (X_s, \phi_j) \right) dZ_s(\phi_i). \end{aligned} \tag{12}$$

The last term is a local martingale and the first three terms on the right-hand side can easily be identified with the three terms of the expression for the weak generator (7) applied to the simple product function F. The proof for general F now consists of an approximation argument using the simple functions above. We shall simultaneously approximate F and all the derivatives of F that occur in the formula for $L^\sigma F$. For functions on $\mathbb{R}^n$ approximating derivatives can be done elegantly using Fourier transforms. In this infinite-dimensional setting we shall do it the hard way, approximating the second derivative $D_{xy} F$ first and integrating up to get approximations to lesser derivatives. We need to take care to ensure that, after integrating up, we remain in the class of simple product functions. Readers who believe this can be done will wish to skip to the next section.

In what follows we shall repeatedly need a type of fundamental theorem of calculus for functions $F : \mathcal{M} \to \mathbb{R}$ to allow us to reconstruct F from its derivatives.

LEMMA 4. (a) *Suppose $F : \mathcal{M} \to \mathbb{R}$ is continuous and has a continuous derivative $D_x F : E \times \mathcal{M} \to \mathbb{R}$. Then, writing* 0 *for the zero measure*,

$$F(\mu) = F(0) + \int_0^1 \int D_x F(\theta \mu) \mu(dx)\, d\theta. \tag{13}$$

(b) *Suppose $G(x,\mu)$ is continuous and has one spatial derivative $D_yG(x,\mu)$ that is continuous in $x, y \in E$, $\mu \in \mathcal{M}$. Suppose also that $D_yG(x,\mu) = D_xG(y,\mu)$. Define*

$$F(\mu) = \int_0^1 \int G(y, \theta\mu)\mu(dy)\, d\theta. \tag{14}$$

Then $D_xF(\mu) = G(x,\mu)$ for $x \in E$, $\mu \in \mathcal{M}$.

In particular, this holds if $G(x,\mu) = \int \phi(x,z)\mu^k(dz)$ for some $\phi \in C(E^{k+1})$ which is symmetric under permutations of its variables.

PROOF. The continuity of $D_xF(\mu)$ implies that the function

$$H(\theta_1,\dots,\theta_n) := F(\theta_1\delta_{x_1} + \cdots + \theta_n\delta_{x_n})$$

is continuously differentiable on $[0,\infty)^n$. For a weighted sum of point masses $\mu = \sum_{i=1}^n c_i\delta_{x_i}$, part (a) holds by applying the fundamental theorem of calculus on [0, 1] to the function $\theta \to H(\theta c_1,\dots,\theta c_n) = F(\theta\mu)$. Weighted sums of point masses are dense in $\mathcal{M}$ and, for fixed $\mu \in \mathcal{M}$, we may take a sequence of such sums μ_ε so that $\mu_\varepsilon \to \mu$ in the weak topology. Then we can pass to the limit in equation (13) for μ_ε, to obtain the same equation for μ, by using the fact that $D_xF(\theta\mu_\varepsilon) \to D_xF(\theta\mu)$ uniformly over $x \in E$.

To prove part (b) of the lemma we differentiate (14) from the definition to obtain

$$\begin{aligned} D_xF(\mu) &= \lim_{\varepsilon\to 0}\int_0^1\int \frac{G(y,\theta(\mu+\varepsilon\delta_x)) - G(y,\theta\mu)}{\varepsilon}\mu(dy)\, d\theta \\ &\quad + \lim_{\varepsilon\to 0}\int_0^1 G\big(x,\theta(\mu+\varepsilon\delta_x)\big)\, d\theta \\ &= \lim_{\varepsilon\to 0}\int_0^1\int\int_0^1 \theta D_xG\big(y,\theta(\mu+\theta'\varepsilon\delta_x)\big)\, d\theta'\, \mu(dy)\, d\theta \\ &\quad + \int_0^1 G(x,\theta\mu)\, d\theta \\ &= \int_0^1\int \theta D_xG(y,\theta\mu)\mu(dy)\, d\theta + \int_0^1 G(x,\theta\mu)\, d\theta \\ &= \int_0^1\int \theta D_yG(x,\theta\mu)\mu(dy)\, d\theta + \int_0^1 G(x,\theta\mu)\, d\theta, \end{aligned} \tag{15}$$

where in the second equality we applied part (a) of this lemma. Using part (a) again we have that

$$\begin{aligned} \int_0^1 G(x,\theta\mu)\, d\theta &= \int_0^1 \left(G(x,0) + \int_0^1\int D_yG(x,\theta'\theta\mu)\theta\, \mu(dy)\, d\theta'\right) d\theta \\ &= G(x,0) + \int_0^1\int (1-\theta'')D_yG(x,\theta''\mu)\mu(dy)\, d\theta''. \end{aligned} \tag{16}$$

Combining (15) with (16) gives

$$D_x F(\mu) = G(x,0) + \int_0^1 \int D_y G(x,\theta\mu)\,\mu(dy)\,d\theta = G(x,\mu)$$

using part (a) of this lemma again. If $G(x,\mu) = \int \phi(x,z)\mu^k(dz)$ then

$$\begin{aligned} D_x G(y,\mu) &= \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\biggl(\int \phi(y,z_1,\dots,z_k)\bigl(\mu(dz_1)+\varepsilon\delta_x\bigr)\cdots\bigl(\mu(dz_k)+\varepsilon\delta_x\bigr) \\ &\qquad\qquad - \int \phi(y,z)\mu^k(dz)\biggr) \\ &= \sum_{i=1}^k \int \phi(y,z_1,\dots,z_{i-1},x,z_{i+1},\dots,z_k)\,\mu^{k-1} \\ &\qquad \times (dz_1,\dots,dz_{i-1},dz_{i+1},\dots,dz_k), \end{aligned}$$

which is symmetric in x and y if ϕ is symmetric in its variables. □

Iterating the fundamental theorem gives a corollary showing one way to reconstruct F from its second partial derivatives.

COROLLARY 5. (a) *Suppose* $F:\mathcal{M}\to\mathbb{R}$ *has continuous derivatives* $D_x F$, $D_{xy}F$. *Then*

$$\begin{aligned} F(\mu) &= F(0) + \int D_x F(0)\mu(dx) \\ &\quad + \int_0^1\int_0^1 \iint D_{xy}F(\theta\theta'\mu)\theta\mu(dx)\mu(dy)\,d\theta\,d\theta'. \end{aligned} \tag{17}$$

(b) *Suppose* $F:[0,T]\times\mathcal{M}\to\mathbb{R}$ *has continuous derivatives* $D_x F$, $D_{xy}F$, $D_s F$, $D_{sxy}F$. *Then*

$$\begin{aligned} F(t,\mu) &= F(t,0) + \int\biggl(D_x F(0,0) + \int_0^t D_{sx}F(s,0)\,ds\biggr)\mu(dx) \\ &\quad + \int_0^1\int_0^1\iint\biggl(D_{xy}F(0,\theta'\theta\mu) + \int_0^t D_{sxy}F(s,\theta\theta'\mu)\,ds\biggr) \\ &\qquad \times \theta\,\mu(dx)\,\mu(dy)\,d\theta\,d\theta'. \end{aligned} \tag{18}$$

PROOF. Fixing x and applying (13) to $D_x F(\theta\mu)$ gives

$$D_x F(\theta\mu) = D_x F(0) + \int_0^1 \int D_{yx}F(\theta'\theta\mu)\theta\mu(dy)\,d\theta'. \tag{19}$$

Substituting this into (13) gives (17). For part (b) fix $t\in[0,T]$ and apply (17) to $F(t,\mu)$. Then expand the derivatives $D_{xy}F(t,\theta'\theta\mu)$ and $D_x F(t,0)$ using the usual fundamental theorem for real functions over $[0,t]$ to obtain (18). □

NOTATION. We write $\|f\|_X$ for the supremum norm on the space $C(X)$ of continuous functions on any compact metric space X. The space $\mathcal{M}$ is locally compact with compact subsets $\mathcal{M}(K) = \{\mu \in \mathcal{M} : (\mu, 1) \le K\}$. By using suitable stopping arguments we will be able to restrict to $\mathcal{M}(K)$ in our proof of Theorem 3 and hence we fix $K > 0$ for the remainder of this section. We now define some spaces of simple approximating functions. Let S_n denote the space of permutations on $\{1, \dots, n\}$. For any function $\phi : E^n \to \mathbb{R}$ we define its symmetrization ϕ^{sym} by

$$\phi^{\mathrm{sym}}(x_1, \dots, x_n) = \frac{1}{n!} \sum_{\pi \in S_n} \phi(x_{\pi_1}, \dots, x_{\pi_n}).$$

Recall D_0 is the particular dense linear subspace of $C(E)$ described in the definition of a good generator A. For $k \ge 1$ let $D_0^{\mathrm{prod}}(E^k)$ be the linear span generated by the functions $\prod_{i=1}^k \phi_i(x_i)$ where $\phi_i \in D_0$. If $k = 0$ we let this set of functions be just the constant functions. Then define

$$D_0^{\mathrm{sym}}(E^k) = \{\phi^{\mathrm{sym}} : \phi \in D_0^{\mathrm{prod}}(E^k)\}.$$

Note that $D_0^{\mathrm{sym}}(E^k)$ consists of exactly the symmetric functions in $D_0^{\mathrm{prod}}(E^k)$. Let μ^k be the k-fold product measure of μ. Let $C_0^1([0,T])$ be those $\psi : [0,T] \to \mathbb{R}$ that have one continuous derivative that vanishes at T. Define, for each $n \ge 0$

$$\mathcal{A}_n^{\mathrm{sym}} = \left\{ \sum_{i=1}^m \int_{E^{k_i}} \psi_i(t) \phi_{k_i}(x, z) \mu^{k_i}(dz) : \right.$$
$$\left. \psi_i \in C_0^1([0,T]), \phi_{k_i} \in D_0^{\mathrm{sym}}(E^{k_i+n}),\ k_i, m \ge 0 \right\}.$$

The functions in $\mathcal{A}_n^{\mathrm{sym}}$ act on the variables $t \in [0,T]$, $x \in E^n$ and $\mu \in \mathcal{M}(K)$, and thus $\mathcal{A}_n^{\mathrm{sym}} \subseteq C([0,T] \times E^n \times \mathcal{M}(K))$. Let $\mathcal{A}_n^{\mathrm{prod}}$ be the same set but with $D_0^{\mathrm{sym}}(E^{k_i+n})$ replaced by $D_0^{\mathrm{prod}}(E^{k_i+n})$. The functions in $\mathcal{A}_2^{\mathrm{sym}}$ will be used to approximate $D_{xy}F(s,\mu)$. The functions $F \in \mathcal{A}_0^{\mathrm{sym}}$ are sums of the simple products for which we used Itô's formula directly in (12) to find $L^\sigma F$. Finally we define $C^n([0,T] \times \mathcal{M})$ to be the collection of functions in $C([0,T] \times \mathcal{M}(K))$ possessing n continuous directional derivatives, that is, the derivatives exist for $(\mu, 1) < K$ and have a continuous extension to the closed ball $(\mu, 1) \le K$. Define

$$\mathcal{H}^{k,n} = \{D_{x_1 \cdots x_k} F(t, \mu) : F \in C^n([0,T] \times \mathcal{M}(K))\} \qquad \text{for } k = 0, 1, \dots, n.$$

LEMMA 6. *For each $n \ge 0$ and $K > 0$:*

(a) $\mathcal{A}_n^{\mathrm{prod}}$ *is dense in* $C([0,T] \times E^n \times \mathcal{M}(K))$.
(b) $\mathcal{A}_n^{\mathrm{sym}}$ *is a dense subset of* $\mathcal{H}^{n,n}$ *in* $C([0,T] \times E^n \times \mathcal{M}(K))$.

PROOF. $\mathcal{A}_n^{\text{prod}}$ is a linear subspace and it is easy to check that it is an algebra, since $C^1([0,T])$ is an algebra, D_0 is an algebra and if $\phi_1 \in C(E^{n+l})$ and $\phi_2 \in C(E^{n+m})$ then

$$\int_{E^l} \phi_1(x,z)\mu^l(dz) \int_{E^m} \phi_2(x,w)\mu^m(dw) = \int_{E^{l+m}} \phi(x,z,w)\mu^{l+m}(dz\,dw),$$

where $\phi(x,z,w) = \phi_1(x,z)\phi_2(x,w)$. Moreover, it is not hard to show that $\mathcal{A}_n^{\text{prod}}$ separates points. Part (a) follows by the Stone–Weierstrass theorem.

Since $\mathcal{A}_0^{\text{prod}} = \mathcal{A}_0^{\text{sym}}$, we may now consider $n \geq 1$ in part (b). If $G \in \mathcal{A}_n^{\text{sym}}$ and we define

$$F(t, x_1, \dots, x_{n-1}, \mu) = \int_0^1 \int G(t, x_1, \dots, x_n, \theta\mu)\mu(dx_n)\,d\theta,$$

it is easy to verify that $F \in \mathcal{A}_{n-1}^{\text{sym}}$. Moreover, $D_{x_n} F(t, x_1, \dots, x_{n-1}, \mu) = G(t, x_1, \dots, x_n, \mu)$, which follows either by direct calculation or from Lemma 4(b). Using this and induction we see that $\mathcal{A}_n^{\text{sym}}$ is a subset of $\mathcal{H}^{n,n}$. In the rest of this proof we shall show that $\mathcal{H}^{n,n+1} \subseteq \overline{\mathcal{A}_n^{\text{sym}}}$. In Corollary 12 we show that $\mathcal{H}^{n,n} \subseteq \overline{\mathcal{H}^{n,n+1}}$ which will therefore complete the proof.

Fix $D_{x_1 \cdots x_n} F(t,\mu) \in \mathcal{H}^{n,n+1}$. Then $D_{x_1 \cdots x_n x_{n+1}} F(t,\mu) \in \mathcal{H}^{n+1,n+1}$ and so, by part (a) of this lemma, for any $\varepsilon > 0$ and $K \geq 0$, we can find $G_1 \in \mathcal{A}_{n+1}^{\text{prod}}$ with

$$\|G_1(t, x_1, \dots, x_{n+1}, \mu) - D_{x_1 \cdots x_{n+1}} F(t,\mu)\|_{[0,T]\times E^{n+1} \times \mathcal{M}(K)} \leq \varepsilon. \tag{20}$$

Since $D_{x_1 \cdots x_{n+1}} F(t,\mu)$ is symmetric in the variables $x_1, \dots, x_{n+1}$, we may symmetrize G_1 in these variables without changing the bound (20) and still have $G_1 \in \mathcal{A}_{n+1}^{\text{prod}}$. In the same way we may find $G_0 \in \mathcal{A}_n^{\text{prod}}$ that lies within ε of $D_{x_1 \cdots x_n} F(t,0)$ and is symmetric in $x_1, \dots, x_n$. Now define

$$\begin{aligned} F^{\varepsilon}(t, x_1, \dots, x_n, \mu) &= G_0(t, x_1, \dots, x_n, \mu) \\ &\quad + \int_0^1 \int_E G_1(t, x_1, \dots, x_n, x_{n+1}, \theta\mu)\mu(dx_{n+1})\,d\theta. \end{aligned} \tag{21}$$

If we compare this with the reconstruction formula (13) for $D_{x_1 \cdots x_n} F(t,\mu)$ in terms of $D_{x_1 \cdots x_{n+1}} F$, we see that

$$\|F^{\varepsilon}(t, x_1, \dots, x_n, \mu) - D_{x_1 \cdots x_n} F(t,\mu)\|_{[0,T]\times E^n \times \mathcal{M}(K)} \leq (1+K)\varepsilon.$$

It remains only to show that $F^{\varepsilon} \in \mathcal{A}_n^{\text{sym}}$. Since $D_{x_1 \cdots x_n} F(t,0)$ does not depend on μ, we may choose G_0 independent of μ (indeed we may replace $G_0(t, x_1, \dots, x_n, \mu)$ by $G_0(t, x_1, \dots, x_n, 0)$). This, and the symmetry in $x_1, \dots, x_n$ imply that G_0 is actually a member of $\mathcal{A}_n^{\text{sym}}$. The function G_1 may, since it was chosen from $\mathcal{A}_{n+1}^{\text{prod}}$, be written as a linear combination of terms of the form

$$\int_{E^k} \psi(t)\phi_k(x_1, \dots, x_n, x_{n+1}, z_1, \dots, z_k)\mu^k(dz)$$
$$\text{with } \psi \in C_0^1([0,T]) \text{ and } \phi_k \in C(E^{n+k+1}).$$

By our earlier remark, we may assume that ϕ_k is symmetric in the variables $x_1, \dots x_{n+1}$. This term enters into the formula (21) for F^ε as

$$\frac{\theta^{k+1}}{k+1} \int_E \int_{E^k} \phi_k(x_1, \dots, x_n, x_{n+1}, z_1, \dots, z_k) \mu^k(dz) \mu(dx_{n+1}). \tag{22}$$

The integral in (22) is with respect to the product measure μ^{k+1} so we may symmetrize ϕ_k in its last $k+1$ arguments. Since ϕ_k is also symmetric in its first $n+1$ arguments, and S_{n+k+1} is generated by the collection consisting of permutations of $\{1, \dots, n+1\}$ and of $\{n+1, \dots, n+k+1\}$, ϕ_k may be replaced by ϕ_k^{sym} and hence $F^\varepsilon \in \mathcal{A}_n^{\mathrm{sym}}$. $\square$

The above lemma shows we can approximate $D_{xy}F$ by elements of $\mathcal{A}_2^{\mathrm{sym}}$. We now turn to the approximation of $D_s F$ and $\int_E A^{(x)} D_x F \mu(dx)$. To do this we introduce some more notation.

NOTATION. For each $n \geq 0$, define the (stopped) semigroup (V_t^n) on $C([0, T] \times E^n \times \mathcal{M}(K))$ by

$$V_s^n F(t, x_1, \dots, x_n, \mu) = \begin{cases} P_s^{(x_1)} \cdots P_s^{(x_n)} F(t+s, x_1, \dots, x_n, P_s^* \mu), & \text{if } s+t \leq T, \\ P_{T-t}^{(x_1)} \cdots P_{T-t}^{(x_n)} F(T, x_1, \dots, x_n, P_{T-t}^* \mu), & \text{if } s+t \geq T. \end{cases}$$

Here $\{P_t^*\}$ is the dual semigroup to $\{P_t\}$, acting on $\mathcal{M}$. Let $D_s + Q^n$ be the generator of (V_s^n) acting on $C([0, T] \times E^n \times \mathcal{M}(K))$.

The operator V_s^n acts independently on the variables $t, x_1, \dots, x_n, \mu$ and this makes the semigroup property clear. The domain of its generator is described in the following lemma.

LEMMA 7. *Fix $K > 0$.*

(a) *Suppose for some $F \in C([0, T] \times E^n \times \mathcal{M}(K))$ that $D_s F$, $A^{(x_i)} F$, $D_z F$ and $A^{(z)} D_z F$ exist and are continuous in all variables. Then F is in the domain of $D_s + Q^n$ and*

$$\begin{aligned} (D_s + Q^n) F(t, x_1, \dots, x_n, \mu) &= D_s F(t, x_1, \dots, x_n, \mu) + \sum_{i=1}^n A^{(x_i)} F(t, x_1, \dots, x_n, \mu) \\ &\quad + \int_E A^{(z)} D_z F(t, x_1, \dots, x_n, \mu) \, \mu(dz). \end{aligned} \tag{23}$$

(b) *Suppose* $F \in C^n([0,T] \times \mathcal{M}(K))$ *and that* $D_{x_1\cdots x_n}F$ *satisfies the hypotheses of part* (a). *Then*

$$D_{x_1\cdots x_n}\big((D_s + Q^0)F(t,\mu)\big) = (D_s + Q^n)D_{x_1\cdots x_n}F(t,\mu).$$

PROOF. The proofs of both of parts of this lemma are fairly routine. We only sketch some steps and leave the details to the reader. For part (a) one can use the fact that the semigroup (V_t^n) is made up of separate semigroups acting in each variable. The expression (23) is simply the sum of the generators for the individual semigroups. In particular, the derivative of the heat flow $\mu \to P_t^*\mu$ is given, for suitable G, by

$$\frac{d}{dt}G(P_t^*\mu) = \int_E A^{(z)} D_z G(P_t^*\mu)\,\mu(dz).$$

Part (b) follows once one has shown that

$$D_{x_k}\big((D_s + Q^{k-1})D_{x_1\cdots x_{k-1}}F(t,\mu)\big) = (D_s + Q^k)D_{x_1\cdots x_k}F(t,\mu)$$

for $k = 1,\dots,n$. To show this, one applies directly the definition of the directional derivative D_{x_k}. The key point is the fact that, for $i \in \{1,\dots,k-1\}$,

$$\begin{aligned}
&D_{x_k}\big(A^{(x_i)} D_{x_1\cdots x_{k-1}}F(t,\mu)\big)\\
&\quad = \lim_{\varepsilon\downarrow 0} \frac{A^{(x_i)} D_{x_1\cdots x_{k-1}}F(t,\mu+\varepsilon\delta_{x_k}) - A^{(x_i)} D_{x_1\cdots x_{k-1}}F(t,\mu)}{\varepsilon}\\
&\quad = \lim_{\varepsilon\downarrow 0}\lim_{u\downarrow 0} \frac{P_u^{(x_i)} - I}{u}\,\frac{D_{x_1\cdots x_{k-1}}F(t,\mu+\varepsilon\delta_{x_k}) - D_{x_1\cdots x_{k-1}}F(t,\mu)}{\varepsilon}\\
&\quad = \lim_{\varepsilon\downarrow 0}\lim_{u\downarrow 0} \frac{1}{u}\int_0^1\int_0^u P_s^{(x_i)} A^{(x_i)} D_{x_1\cdots x_k}F(t,\mu+\varepsilon\theta\delta_{x_k})\,ds\,d\theta \qquad \text{[using (13)]}\\
&\quad = A^{(x_i)} D_{x_1\cdots x_k}F(t,\mu).
\end{aligned}$$

The fact that $(x_1,\dots,x_k,\theta,s) \to P_s^{(x_i)} A^{(x_i)} D_{x_1\cdots x_k}F(t,\mu+\varepsilon\theta\delta_{x_k})$ is uniformly continuous in all variables means that the limits $t \downarrow 0$ and $\varepsilon \downarrow 0$ can be taken in either order and that the convergence is uniform in the other variables $(x_1,\dots,x_{k-1})$. □

We now briefly explain our strategy in the rest of the proof. Note that $(D_s + Q^0)F$ gives precisely the terms D_sF and $\int_E A^{(x)} D_xF\mu(dx)$ which we need to approximate in the drift (7) of $F(X_t)$ in Theorem 3. To approximate F, $(D_s + Q^0)F$ and $D_{xy}F$ simultaneously however we shall approximate $D_{xy}F$ and $(D_s + Q^2)D_{xy}F$ simultaneously and integrate up twice to show we have approximated F and $(D_s + Q^0)F$ as well. The following lemma is the key to implementing this idea.

LEMMA 8. *For each $n \geq 0$:*

(a) *$\mathcal{H}^{n,n}$ is a closed linear subspace of $C([0,T] \times E^n \times \mathcal{M}(K))$;*
(b) *$V_t^n : \mathcal{H}^{n,n} \to \mathcal{H}^{n,n}$ for all $t \geq 0$;*
(c) *$\mathcal{A}_n^{\mathrm{sym}}$ is a core for the generator $D_s + Q^n$ of the semigroup (V_t^n) acting on $\mathcal{H}^{n,n}$.*

PROOF. The standard proof of part (a) uses induction on n. The result is trivial for $n = 0$ since $H^{0,0} = C([0,T] \times \mathcal{M}(K))$. Take a sequence $F_m \in \mathcal{H}^{n,n}$ converging to $F \in C([0,T] \times E^n \times \mathcal{M}(K))$. We may suppose that $F_m(t, x_1, \dots, x_n, \mu) = D_{x_1 \cdots x_n} G_m(t, \mu)$. Define

$$H_m(t, x_1, \dots, x_{n-1}, \mu) = \int_0^1 \int F_m(t, x_1, \dots, x_{n-1}, x_n, \theta\mu)\mu(dx_n)\, d\theta. \tag{24}$$

By Lemma 4(a) we know that H_m differs from $D_{x_1 \cdots x_{n-1}} G_m$ by a function independent of μ, namely $D_{x_1 \cdots x_{n-1}} G_m(t, 0)$. For any symmetric continuous $F : E^n \to \mathbb{R}$, by differentiating from the definition, we have

$$F(x_1, \dots, x_n) = \frac{1}{n!} D_{x_1 \cdots x_n}\left(\int_{E^n} F(y_1, \dots, y_n)\mu^n(dy)\right).$$

Applying this to $D_{x_1 \cdots x_{n-1}} G_m(t, 0)$, we see that $D_{x_1 \cdots x_{n-1}} G_m(t, 0) \in \mathcal{H}^{n-1,n-1}$. Thus $H_m \in \mathcal{H}^{n-1,n-1}$ also. Using the uniform convergence of F_m to F, (24) implies that H_m converge uniformly to continuous H. Also by Lemma 4(a), we have

$$\begin{aligned} &H_m(t, x_1, \dots, x_{n-1}, \mu + \varepsilon\delta_x) - H_m(t, x_1, \dots, x_{n-1}, \mu) \\ &\quad = \int_0^\varepsilon F_m(t, x_1, \dots, x_n, \mu + \eta\delta_x)\, d\eta. \end{aligned}$$

Taking the limit $m \to \infty$ we obtain the same identity with H_m, F_m replaced by H, F. But this identity implies that $D_z H = F$. By the inductive hypothesis $H \in \mathcal{H}^{n-1,n-1}$ and this implies the conclusion of part (a).

For part (b) we suppose $G(t, x_1, \dots, x_n, \mu) = D_{x_1 \cdots x_n} G(t, \mu)$ for some $G \in C^n([0,T] \times \mathcal{M}(K))$. Then, when $t + s \leq T$, by direct differentiation, one can check that $(t, \mu) \to G(t+s, P_s^*\mu) \in C^n([0,T] \times \mathcal{M}(K))$ and that $D_{x_1 \cdots x_n} G(t+s, P_s^*\mu) = V_s^n F(t, x_1, \dots, x_n, \mu)$. This proves part (b) when $t + s \leq T$. The proof when $t + s \geq T$ is entirely similar.

Since $\mathcal{A}_n^{\mathrm{sym}}$ is a dense subspace of $\mathcal{H}^{n,n}$, we can apply again Watanabe's lemma (given in [6], Proposition 1.3.3). This shows it is sufficient, to prove part (c), to establish, for each s, that $V_s^n : \mathcal{A}_n^{\mathrm{sym}} \to \mathcal{A}_n^{\mathrm{sym}}$. Applying V_s^n to a typical term in the sum which constitutes an element of $\mathcal{A}_n^{\mathrm{sym}}$, we have, for some $\psi \in C_0^1([0,T])$

and $\phi \in D_0^{\mathrm{sym}}(E^{k+n})$,

$$V_s^n\left(\psi(t)\int_{E^k}\phi(x_1,\dots,x_n,z_1,\dots,z_k)\mu^k(dz)\right)$$

$$= \begin{cases} \psi(t+s)\displaystyle\int_{E^k} P_s^{(x_1)}\cdots P_s^{(x_n)}\phi(x_1,\dots,x_n,z_1,\dots z_k)(P_s^*\mu)^k(dz), \\ \hfill \text{if } t+s\le T, \\ \psi(T)\displaystyle\int_{E^k} P_{T-t}^{(x_1)}\cdots P_{T-t}^{(x_n)}\phi(x_1,\dots,x_n,z_1,\dots z_k)(P_{T-t}^*\mu)^k(dz), \\ \hfill \text{if } t+s\ge T, \end{cases}$$

$$= \begin{cases} \psi(t+s)\displaystyle\int_{E^k} P_s^{(x_1)}\cdots P_s^{(x_n)}P_s^{(z_1)}\cdots P_s^{(z_k)}\phi(x_1,\dots,x_n,z_1,\dots z_k)\mu^k(dz), \\ \hfill \text{if } t+s\le T, \\ \psi(T)\displaystyle\int_{E^k} P_{T-t}^{(x_1)}\cdots P_{T-t}^{(x_n)}P_{T-t}^{(z_1)}\cdots P_{T-t}^{(z_k)}\phi(x_1,\dots,x_n,z_1,\dots z_k)\mu^k(dz) \\ \hfill \text{if } t+s\ge T. \end{cases}$$

The definition of $C_0^1([0,T])$ implies that for $\psi \in C_0^1([0,T])$ the function $t \to \psi(t+s\wedge T)$ is still an element of $C_0^1([0,T])$. Recalling the form of $\phi \in D_0^{\mathrm{sym}}(E^{k+n})$, and that $P_t : D_0 \to D_0$, the result follows. $\square$

LEMMA 9. *For any F satisfying the conditions of Theorem* 3, *and any $\varepsilon > 0$ and $K > 0$, there exists $F^\varepsilon \in \mathcal{A}_0^{\mathrm{sym}}$ such that*

$$\|F - F^\varepsilon\|_{[0,T]\times\mathcal{M}(K)} \le \varepsilon,$$
$$\|D_{xy}F - D_{xy}F^\varepsilon\|_{[0,T]\times E^2\times\mathcal{M}(K)} \le \varepsilon,$$
$$\|(D_s+Q^0)F - (D_s+Q^0)F^\varepsilon\|_{[0,T]\times\mathcal{M}(K)} \le \varepsilon.$$

PROOF. Throughout this proof K is fixed and the norm $\|\cdot\|$, without a subscript, is the supremum norm of $C([0,T]\times E^m\times\mathcal{M}(K))$ for a relevant value of m.

Fix $F:[0,T]\times\mathcal{M}\to\mathbb{R}$ as in the statement of Theorem 3. By Lemma 7 the hypotheses on F imply that $D_{xy}F(t,\mu)$ is in the domain of the generator D_s+Q^2, that $D_xF(t,0)$ is in the domain of D_s+Q^1 and that $F(t,0)$ is in the domain of D_s+Q^0. Hence, given $\varepsilon>0$, we may, by Lemma 8, pick $G_2^\varepsilon\in\mathcal{A}_2^{\mathrm{sym}}$, $G_1^\varepsilon\in\mathcal{A}_1^{\mathrm{sym}}$, and $G_0^\varepsilon\in\mathcal{A}_0^{\mathrm{sym}}$ so that

$$\begin{aligned} &\|G_2^\varepsilon - D_{xy}F\| + \|(D_s+Q^2)(G_2^\varepsilon - D_{xy}F)\| \le \varepsilon, \\ &\|G_1^\varepsilon - D_xF| + \|(D_s+Q^1)(G_1^\varepsilon - D_xF)\| \le \varepsilon, \\ &\|G_0^\varepsilon - F\| + \|(D_s+Q^0)(G_0^\varepsilon - F)\| \le \varepsilon. \end{aligned} \tag{25}$$

Since $F(t,0)$ and $D_x F(t,0)$ are independent of μ, we may also pick G_1^ε and G_0^ε independent of μ. Now we define

$$
\begin{aligned}
F^\varepsilon(t,\mu) &= G_0^\varepsilon(t) + \int_E G_1^\varepsilon(t,x)\mu(dx) \\
&\quad + \int_0^1\int_0^1\int_E\int_E G_2^\varepsilon(t,x,y,\theta\theta'\mu)\,d\theta\,d\theta'\,\mu(dx)\,\mu(dy)\,d\theta\,d\theta'.
\end{aligned} \tag{26}
$$

Note that F^ε is an element of $\mathcal{A}_0^{\mathrm{sym}}$ and, using Lemma 4(b), that $D_{xy}F^\varepsilon(t,\mu) = G_2^\varepsilon(t,x,y,\mu)$ and $D_x F^\varepsilon(t,0) = G_1^\varepsilon(t,x)$. The bound $\|D_{xy}F^\varepsilon - D_{xy}F\| \le \varepsilon$ follows immediately from (25). Comparing (26) and the reconstruction formula (17) for F, and using the estimates from (25), we see that $\|F^\varepsilon - F\| \le \varepsilon(1+K+K^2)$.

Lemma 7 also shows that $(D_s + Q^0)F$ is twice differentiable and identifies the derivatives. Applying the reconstruction (18), we get

$$
\begin{aligned}
&(D_s + Q^0)F(t,\mu) \\
&\quad = D_s F(t,0) + \int_E (D_s + Q^1) D_x F(t,0)\mu(dx) \\
&\qquad + \int_0^1\int_0^1\int_E\int_E (D_s + Q^2) D_{xy} F(t,\theta\theta'\mu)\,d\theta\,d\theta'\,\mu(dx)\,\mu(dy).
\end{aligned} \tag{27}
$$

Applying this formula with the choice $F = F^\varepsilon$ gives

$$
\begin{aligned}
&(D_s + Q^0)F^\varepsilon(t,\mu) \\
&\quad = D_s G_0^\varepsilon(t) + \int_E (D_s + Q^1) G_1^\varepsilon(t,x)\mu(dx) \\
&\qquad + \int_0^1\int_0^1\int_E\int_E (D_s + Q^2) G_2^\varepsilon(t,x,y,\theta\theta'\mu)\,d\theta\,d\theta'\mu(dx)\mu(dy).
\end{aligned} \tag{28}
$$

Comparing (27) and (28) and using the estimates in (25) shows that

$$
\|(D_s + Q^0)F^\varepsilon - (D_s + Q^0)F\| \le \varepsilon(1+K+K^2)
$$

which completes the proof. □

PROOF OF THEOREM 3. We make the following reductive assumption: there exists $K > 0$ so that, with probability 1,

$$
(X_t, 1) \le K \quad \text{and} \quad |\sigma_t| \le K \qquad \text{for all } t \in [0,T]. \tag{29}
$$

We claim that if we can prove Theorem 3 when this assumption holds then we can prove the general case. To see this suppose that $\{X_t\}$ is as in the statement of Theorem 3. Using the local boundedness of σ, choose stopping times T_K^1

so that $\sigma_t^K := \sigma_t I(t < T_K)$ is bounded by K and $T_K^1 \uparrow \infty$ as $K \to \infty$. Set $T_K^2 = \inf\{t : (X_t, 1) \geq K\}$ and $T_K = T_K^1 \wedge T_K^2$. Let $\Omega_K = \{(X_0, 1) \leq K\}$ and define

$$X_t^K = \begin{cases} 0, & \text{on } \Omega_K^c, \\ X_t, & \text{on } \Omega_K \cap \{t < T_K\}, \\ P^*_{t-T_K} X_{T_K}, & \text{on } \Omega_K \cap \{t \geq T_K\}. \end{cases} \tag{30}$$

It is straightforward to show that (X_t^K) is a solution on $(\Omega, \mathcal{F}_t, P)$ to the martingale problem $M(\sigma^K, A)$. Moreover, $d(X_t^K, 1) = 0$ for $t > T_K$ so the total mass process never exceeds K. So (X_t^K) and (σ_t^K) satisfy the assumption (29) and so $M_t^K := F(t, X_t^K) - \int_0^t L^{\sigma^K} F(s, X_s^K)\, ds$ is a local martingale on $[0, T]$. But on the set Ω_K

$$F(t \wedge T_K, X_{t\wedge T_K}) - \int_0^{t\wedge T_K} L^\sigma F(s, X_s)\, ds = M^K_{t\wedge T_K}.$$

Since Ω_K is $\mathcal{F}_0$ measurable the process $F(t \wedge T_K, X_{t\wedge T_k}) - \int_0^{t\wedge T_K} L^\sigma F(s, X_s)\, ds$ is also a local martingale on $[0, T]$. Since $T_K \uparrow \infty$ this completes the reduction of Theorem 3 to the case where assumption (29) holds.

We now fix a constant K where (29) holds. For F as in the hypotheses of Theorem 3, we pick F^ε as in Lemma 9. The function F_ε is an element of $\mathcal{A}_0^{\mathrm{sym}}$ so we may apply Itô's formula to $F_\varepsilon(t, X_t)$ as in (12) to see that $F_\varepsilon(t, X_t) - \int_0^t L^\sigma F_\varepsilon(s, X_s)\, ds$ is a local martingale. Since, by assumption (29), both σ and the total mass $(X_t, 1)$ are bounded by K, the functions F and $L^\sigma F$ are evaluated only on the compact set and hence are bounded. So the process is a true martingale and for any bounded $\mathcal{F}_s$ measurable variable Z_s we have

$$E\biggl(Z_s\biggl(F_\varepsilon(t, X_t) - F_\varepsilon(s, X_s) - \int_s^t L^\sigma F_\varepsilon(r, X_r)\, dr\biggr)\biggr) = 0 \quad \text{for } s \leq t \leq T. \tag{31}$$

Now let $\varepsilon \to 0$ in this expectation. Using the various uniform convergence estimates in Lemma 9 and the fact that $(X_t, 1) \leq K$, we obtain the same (31) with F_ε replaced by F. Hence $F(t, X_t) - \int_0^t L^\sigma F(s, X_s)$ is an $(\mathcal{F}_t)$ martingale on $[0, T]$, completing the proof of Theorem 3. □

3. Smoothing properties of the Dawson–Watanabe semigroup. It is not always obvious how to smooth functions on infinite-dimensional spaces. The properties we develop in this section suggest that smoothing using the Dawson–Watanabe transition semigroup (U_t^c) is a useful method. The key to our proof of the smoothness of $U_t^c \Phi(\mu)$ is the branching structure underlying Dawson–Watanabe processes. The branching rate $c > 0$ will be fixed throughout this section. We start by collecting the three facts we shall use.

FACT 1. Define Q^t_μ to be the law of a Dawson–Watanabe process at time t, with initial condition μ and constant branching c, so that $U^c_t \Phi(\mu) = Q^t_\mu(\Phi)$. The branching property of the Dawson–Watanabe process can be expressed as

$$Q^t_{\mu+\lambda}(\Phi) = Q^t_\mu * Q^t_\lambda(\Phi) := \iint \Phi(\nu_1 + \nu_2) Q^t_\mu(d\nu_1) Q^t_\lambda(d\nu_2). \tag{32}$$

This is thought of intuitively as the fact that disjoint sets of particles evolve independently.

FACT 2. Q^t_μ is the law of a Cox cluster random measure (see [4], Sections 3 and 4). Intuitively the measure is thought of as a Poisson number of clusters, rooted at points chosen according to μ, where each cluster represents the surviving ancestors of one individual at time zero. This can be expressed as follows: For each $t > 0$ and $x \in E$, there is a probability kernel $(R^t_x(A) : A \subseteq \mathcal{M})$, satisfying $R^t_x(\{0\}) = 0$, so that Q^t_μ is the law of $\int_{\mathcal{M}} \nu\, \eta^\mu_t(d\nu)$, where η^μ_t is a Poisson random measure on $\mathcal{M}$ with finite intensity $(2/ct)\int_E \mu(dx) R^t_x(d\nu)$. The kernel $R^t_x(A)$ is characterized by its Laplace functional given, for continuous $\phi : E \to [0, \infty)$, by

$$\int_{\mathcal{M}} \exp(-(\nu, \phi)) R^t_x(d\nu) = 1 - \frac{ct}{2} u_t(x) \tag{33}$$

where $(u_s(x) : 0 \le s \le t, x \in E)$ is the unique nonnegative solution to the differential equation

$$\partial_t u = Au - \frac{cu^2}{2}, \qquad u_0(x) = \phi(x). \tag{34}$$

In the case that $\mu = \varepsilon\delta_x$ we can write the measure $\int \nu\, \eta^\mu_t(d\nu)$ as a finite sum of a Poisson number N, mean $2\varepsilon/ct$, of i.i.d. random measures $\{Z^i_t\}$, independent of N, as follows:

$$\int_{\mathcal{M}} \nu\, \eta^{\varepsilon\delta_x}_t(d\nu) = \sum_{i=1}^N Z^i_t \tag{35}$$

where each Z^i_t has the law $P(Z^i_t \in A) = R^t_x(A)$. The sum is zero if $N = 0$. The Laplace functional of R^t_x can be used to show that $(2/ct)Z^i_t$ converge in law, as $t \downarrow 0$, to a point mass at x, with a weight given by an exponential variable with mean 1.

FACT 3. The total mass $(\nu, 1)$ under the law $R^t_x(d\nu)$ has an exponential distribution with mean $ct/2$, as can be checked from (33). We can form a further disintegration by conditioning on this total mass to obtain, for each $m > 0$, $t > s > 0$, $x \in E$, a probability kernel $(R^{s,t}_{x,m}(A) : A \subseteq \mathcal{M})$ so that

$$R^t_x(\Phi) = \int_0^\infty dm\, e_t(m) R^{0,t}_{x,m}(\Phi) \qquad \text{where } e_t(m) = \frac{2}{ct} e^{-2m/ct}. \tag{36}$$

The kernels $R^{s,t}_{x,m}$ can be defined via the following probabilistic description, due to Dawson and Perkins (see [4], Theorem 12.4.6), which they call the "splitting atom process." A particle of mass m starts at x at time s and moves according to the underlying motion process. At the inhomogeneous rate $2mtc^{-1}(t-r)^{-2}\,dr$, for $r \in [s,t)$, the particle splits. At the splitting time two particles are formed with masses um and $(1-u)m$ where u is chosen independently and uniformly over $[0,1]$. After the splitting the two particles continue independently using the same rules as the parent particle. This measure valued process converges to a limit at time t and the law of this limiting random measure is $R^{s,t}_{x,m}(d\nu)$. Note that the inhomogeneous rate has infinite intensity on $[s,t)$ ensuring that a split (and then infinitely many splits) occur. Write $p_t(x,dy)$ for a measurable probability kernel that generates the operators $\{P_t\}$. By conditioning on the time of the first split, we have that

$$R^{s,t}_{x,m}(\Phi) = \int_s^t dr\,\pi_{m,s,t}(r)\int_E p_{r-s}(x,dy)\int_0^1 du\,R^{r,t}_{y,um} * R^{r,t}_{y,(1-u)m}(\Phi), \tag{37}$$

where

$$\begin{aligned}\pi_{m,s,t}(r) &= \frac{2mt}{c(t-r)^2}\exp\left(-\int_s^r \frac{2mt}{c(t-q)^2}\,dq\right)\\ &= \frac{2mt}{c(t-r)^2}\exp\left(-\frac{2mt(r-s)}{c(t-r)(t-s)}\right).\end{aligned}$$

This representation implies a certain smoothness of the law $R^t_x(d\nu)$ in x and will lead, for suitable underlying motion, to the regularity of the derivatives $D_x U^c_t\Phi$ in x.

One immediate consequence of the last fact is the following lemma:

LEMMA 10. *If $\{P_t\}$ is a strong Feller semigroup then, for bounded measurable $\Phi : \mathcal{M} \to \mathbb{R}$, the map $(t,\mu) \to U^c_t\Phi(\mu)$ is continuous on $(0,\infty)\times\mathcal{M}$. In particular, $\{U^c_t\}$ is a strong Feller semigroup.*

PROOF. Fix a bounded measurable Φ. By subtracting a constant we may assume that that $\Phi(0) = 0$. The representation of Q^t_μ as the law of a Poisson random measure allows us to calculate $U^c_t\Phi(\mu) = Q^t_\mu(\Phi)$ in terms of the intensity of the Poisson random measure as

$$\begin{aligned}U^c_t\Phi(\mu) &= \frac{2}{ct}\int_E \mu(dx)R^t_x(\Phi)\\ &= \int_E \mu(dx)\int_0^t dr\int_E p_r(x,dy)H^\Phi(r,t,y)\end{aligned} \tag{38}$$

where, using the decompositions (36) and (37), $H^\Phi(r,t,y)$ is defined by

$$H^\Phi(r,t,y) = \frac{2}{ct}\int_0^\infty dm\, e_t(m)\pi_{m,0,t}(r)\int_0^1 du\, R^{r,t}_{y,um} * R^{r,t}_{y,(1-u)m}(\Phi).$$

We can bound $|H^\Phi(r,t,y)|$ by

$$\begin{aligned}\frac{2}{ct}\int_0^\infty dm\, e_t(m)\pi_{m,0,t}(r)\|\Phi\|_{\mathcal{M}} \\ = \frac{2t}{c(rt+t-r)^2}\|\Phi\|_{\mathcal{M}} \\ \le \frac{2}{c}\max\{t^{-1},t^{-3}\}\|\Phi\|_{\mathcal{M}}.\end{aligned}$$

The strong Feller property of $\{P_t\}$ imply that the map $x \to \int_0^t dr \int_E p_r(x,dy) \times H^\Phi(r,t,y)$ is bounded and continuous. This and (38) show that $U_t^c\Phi(\mu)$ is continuous in μ. In particular, the semigroup $\{U_t^c\}$ is strong Feller. For the joint continuity in (t,μ), we fix $s<t$ and write $U_t^c\Phi(\mu) = U_{t-s}^c\Psi(\mu)$ where $\Psi = U_s^c\Phi$ is bounded and continuous. The joint continuity is now a consequence of the Feller property of $\{U_t^c\}$. □

The next result shows that smoothing with the Dawson–Watanabe semigroup yields derivatives of all orders.

LEMMA 11. *Fix $\Phi : \mathcal{M} \to [0,\infty)$ and $x_1,\dots,x_n \in E$.*

(a) *Suppose, for some $t>0$, that $U_t^c\Phi(\mu) < \infty$ for all $\mu \in \mathcal{M}$. Then the derivatives $D_{x_1\cdots x_n}U_t^c\Phi(\mu)$ exist and are given by the following expression*:

$$D_{x_1\cdots x_n}U_t^c\Phi(\mu) = \left(\frac{2}{ct}\right)^n \sum_{A\subseteq\{1,\dots,n\}} (-1)^{n-|A|} E\left(\Phi\left(Y_{t,\mu} + \sum_{i\in A} Z_t^i\right)\right),$$

where $Y_{t,\mu}$ has the law Q_μ^t of a Dawson–Watanabe process at time t started at μ, $(Z_t^i : i = 1,2,\dots)$ is an independent sequence of independent random measures and Z_t^i has the cluster law R_{t,x_i}. The sum above is over all subsets A of $\{1,\dots,n\}$ and $|A|$ denotes the cardinality of A.

(b) *If $U_t^c\Phi(\mu)$ is continuous for $t \in (0,T]$, $\mu \in \mathcal{M}$ then the derivatives $D_{x_1\cdots x_n}U_t^c\Phi(\mu)$ are continuous for $x_i \in E$, $\mu \in \mathcal{M}$, $t \in (0,T]$.*

(c) *If Φ is continuous and satisfies the growth condition* (6) *then $U_t^c\Phi(\mu)$ is continuous over $t \in [0,T]$, $\mu \in \mathcal{M}$.*

PROOF. Consider the case $n=1$. Fix $x \in E$ and let $(Z_{t,x}^i : i = 1,2,\dots)$ be an i.i.d. sequence of random measures with the cluster law $R_{t,x}$. Let N be a Poisson variable, with mean $2\varepsilon/ct$, independent of $Y_{t,\mu}$ and of $(Z_{t,x}^i : i = 1,2,\dots)$. Using the branching property (32) and the representation (35) we have

$$
\begin{aligned}
D_x U_t^c \Phi(\mu) &= \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} E\left(\Phi\left(Y_{t,\mu} + \sum_{i=1}^{N} Z_{t,x}^i\right) - \Phi(Y_{t,\mu})\right) \\
&= \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \sum_{k=1}^{\infty} \frac{e^{-2\varepsilon/ct}(2\varepsilon/ct)^k}{k!} E\left(\Phi\left(Y_{t,\mu} + \sum_{i=1}^{k} Z_{t,x}^i\right) - \Phi(Y_{t,\mu})\right) \\
&= \left(\frac{2}{ct}\right) E\big(\Phi(Y_{t,\mu} + Z_{t,x}^1) - \Phi(Y_{t,\mu})\big).
\end{aligned}
$$

To justify the interchange of the limit and the sum over k, we use the dominated convergence theorem with the domination, over $\varepsilon \in (0, 1]$,

$$
\begin{aligned}
&\sum_{k=0}^{\infty} \frac{(2/ct)^k}{k!} E\left(\Phi\left(Y_{t,\mu} + \sum_{i=1}^{k} Z_{t,x}^i\right) + \Phi(Y_{t,\mu})\right) \\
&\qquad = e^{2c/t}\big(U_t^c \Phi(\mu + \delta_x) + U_t^c \Phi(\mu)\big)
\end{aligned}
\tag{39}
$$

which is finite by assumption. The existence of the nth order derivative follows by a very similar argument, using induction on n, and using the finiteness of $U_t^c \Phi(\mu + n\delta_x)$ for the dominated convergence step.

The map $x \to \delta_x \in \mathcal{M}$ is continuous. So under the continuity assumption of part (b) the map $U_t^c \Phi(\mu + \varepsilon\delta_x)$ is continuous in $x \in E, \mu \in \mathcal{M}, t \in (0, T]$. The definition of the derivative shows that $D_x U_t^c \Phi(\mu)$ is the limit of functions that are continuous in x, μ, t. The domination above can be used to show the limit is uniform over $t \in (t_0, T]$ for any $t_0 > 0$. Again the argument for the higher derivatives is similar.

If Φ is continuous then the Feller property of (U_t^c) implies that $U_t^c(\Phi \wedge n)(\mu)$ is continuous for $t \in [0, T]$, $\mu \in \mathcal{M}$. Under the growth condition, we have for some $p > 1$, using Lemma 2, that

$$
\sup_{\mu \in \mathcal{M}(K)} \sup_{t \in [0,T]} U_t^c(\Phi^p)(\mu) \le 4 \exp(2K/cT) < \infty.
$$

Using this one can show $U_t^c \Phi(\mu)$ is the limit of $U_t^c(\Phi \wedge n)(\mu)$ as $n \to \infty$, uniformly over $t \in [0, T]$, $\mu \in \mathcal{M}(K)$. $\square$

We now finish one unproved step from Section 2.

COROLLARY 12. *Using the notation introduced before Lemma* 6 *in Section* 2, $\mathcal{H}^{n,n+1}$ *is dense in* $\mathcal{H}^{n,n}$.

PROOF. Fix $K > 0$, $F \in \mathcal{H}^{n,n}$ and $G \in C^n([0, T] \times \mathcal{M}(K))$ with $F = D_{x_1 \cdots x_n} G$. Extend G to a function $\tilde{G}$ defined on $[0, T] \times \mathcal{M}$, that is continuous and bounded. Now set, for $\varepsilon, \delta > 0$,

$$
G_{\varepsilon,\delta}(t, \mu) = U_\varepsilon^c \tilde{G}(t, (1 + \delta)^{-1}\mu) \quad \text{and} \quad F_{\varepsilon,\delta} = D_{x_1 \cdots x_n} G_{\varepsilon,\delta}.
$$

By Lemma 11, we have $G_{\varepsilon,\delta} \in C^{n+1}([0,T] \times \mathcal{M})$ and hence $F_{\varepsilon,\delta} \in \mathcal{H}^{n,n+1}$. We shall show that

$$\lim_{\delta\downarrow 0}\lim_{\varepsilon\downarrow 0} \|F_{\varepsilon,\delta} - F\|_{[0,T]\times E^n \times \mathcal{M}(K)} = 0 \tag{40}$$

which will complete the proof. Using the representation for the derivatives of $F_{\varepsilon,\delta}$ from Lemma 11, we obtain

$$\begin{aligned} &F_{\varepsilon,\delta}(t, x_1, \dots, x_n, \mu) \\ &\quad = \left(\frac{2}{c\varepsilon}\right)^n \sum_{A\subseteq\{1,\dots,n\}} (-1)^{n-|A|} \\ &\qquad\qquad \times E\left(\tilde{G}\left(t, (1+\delta)^{-1}\left(Y_{\varepsilon,\mu} + \sum_{i\in A} Z_\varepsilon^i\right)\right)\right) \\ &\quad = \left(\frac{2}{c\varepsilon}\right)^n \sum_{A\subseteq\{1,\dots,n\}} (-1)^{n-|A|} \\ &\qquad\qquad \times E\left(G\left(t, (1+\delta)^{-1}\left(Y_{\varepsilon,\mu} + \sum_{i\in A} Z_\varepsilon^i\right)\right) \right. \\ &\qquad\qquad\qquad \left. \times I\left((Y_{\varepsilon,\mu}, 1) + \sum_{i=1}^n (Z_\varepsilon^i, 1) \le K(1+\delta)\right)\right) \\ &\qquad + \mathrm{Error}(t, x_1, \dots, x_n, \mu), \end{aligned} \tag{41}$$

where

$$\begin{aligned} &|\mathrm{Error}(t, x_1, \dots, x_n, \mu)| \\ &\quad \le \left(\frac{2}{c\varepsilon}\right)^n 2^n \|\tilde{G}\|_{[0,T]\times\mathcal{M}} P\left((Y_{\varepsilon,\mu}, 1) + \sum_{i=1}^n (Z_\varepsilon^i, 1) > K(1+\delta)\right). \end{aligned}$$

The process $(Y_{t,\mu}, 1)$ follows a Feller diffusion. There are positive exponential moments given by

$$E\big(\exp(\lambda Y_{t,\mu}(1))\big) = \exp\big(2\lambda\mu(1)/(2 - c\lambda t)\big) \qquad \text{when } \lambda < 2/ct.$$

A Chebyshev argument using these moments shows that $P((Y_{\varepsilon,\mu}, 1) > (\mu, 1) + \eta)$ is exponentially small in ε^{-1}. Using the the Laplace transform of the independent masses $(Z_\varepsilon^i, 1)$ given by (33) and corresponding exponential moments, a Chebyshev argument shows a similar bound for $P((Z_\varepsilon^i, 1) > \eta)$. Together these show that the probability in the error term, for fixed $\delta > 0$, is exponentially small in ε^{-1}, uniformly over $(\mu, 1) \le K$. We omit the details.

So we may concentrate on the first term on the right-hand side of (41). Note that the terms in the summation evaluate G at the vertices of a n-dimensional parallelopiped. The alternating sign allows us to combine the terms in terms of the derivatives of G using Lemma 4(a) n times, yielding

$$\left(\frac{2}{c\varepsilon}\right)^n \int_0^1 d\theta_1 \cdots \int_0^1 d\theta_n \int_{E^n} E\Bigg(D_{z_1\cdots z_n} G\Bigg(t, (1+\delta)^{-1}\Bigg(Y_{\varepsilon,\mu} + \sum_{i=1}^n \theta_i Z_\varepsilon^i\Bigg)\Bigg)$$

$$\times Z_\varepsilon^1(dz_1)\cdots Z_\varepsilon^n(dz_n) I\Bigg((Y_{\varepsilon,\mu}, 1) + \sum_{i=1}^n (Z_\varepsilon^i, 1) \le K(1+\delta)\Bigg)\Bigg).$$

Now we use convergence of the measures $Y_{\varepsilon,\mu} \to \mu$ and $(2/c\varepsilon)Z_\varepsilon^i \to \mathcal{E}_i \delta_{x_i}$ in law, where $\mathcal{E}_i$ are i.i.d. exponential variables with mean 1. We may, changing the probability space if necessary via Skorokhod's lemma, assume the convergence is almost sure. Then applying the dominated convergence theorem to (42) we may pass to the limit, as $\varepsilon \downarrow 0$, to obtain $D_{x_1\cdots x_n} G(t, (1+\delta)^{-1}\mu)$. Moreover, it is not too hard to show that this convergence is uniform on $(\mu, 1) \le K$. Finally, using the continuity of $D_{x_1\cdots x_n} G$, we let $\delta \downarrow 0$ to complete the proof of (40). □

The example where the underlying motion is uniform motion on a torus shows that the map $x \to D_x U_t^c \Phi$ need not be in the domain $D(A)$ and further smoothing is required to apply Theorem 3. We failed to find a smoothing of Φ that would work and thus we are led to making the smoothing assumption (5) on the underlying motion process.

LEMMA 13. *Suppose the motion semigroup* $\{P_t\}$ *satisfies the smoothing property* (5). *Suppose also that* Φ *is continuous and satisfies the growth condition* (6). *Then for some* $T' > T$, *the maps* $A^{(x_1)} D_{x_1\cdots x_n} U_t^c \Phi(\mu)$ *exist and are continuous in* $x_i \in E$, $\mu \in \mathcal{M}$, *and* $A^{(x_1)} D_{x_1\cdots x_n} U_t^c \Phi(\mu)$ *is locally bounded in* $t \in (0, T']$, *uniformly over* $x_i \in E$, $\mu \in \mathcal{M}(m)$ *for any* m.

PROOF. In this proof $C(t, \alpha, \ldots)$ will denote a quantity whose value may change from line to line, but which is locally bounded as a function of t. We first claim, for continuous Φ and $t > 0$, that

$$\big|R_{y,m}^{s,t}(\Phi) - R_{y,m}^{0,t}(\Phi)\big| \le C(t, c, \alpha)(1+m)s^{1/(1+\beta)} \|\Phi\|_{\mathcal{M}(m)} \quad \text{for } s \le t/2. \tag{42}$$

Indeed, we may use the decomposition (37) to write the difference $R_{y,m}^{0,t}(\Phi) - R_{y,m}^{s,t}(\Phi)$ as

$$\int_0^s dr\, \pi_{m,0,t}(r) \int_E p_r(x, dy) \int_0^1 du\, R_{y,um}^{r,t} * R_{y,(1-u)m}^{r,t}(\Phi)$$

$$+ \int_s^t dr\big(\pi_{m,0,t}(r) - \pi_{m,s,t}(r)\big)$$

$$\times \int_E p_r(x, dy) \int_0^1 du\, R_{y,um}^{r,t} * R_{y,(1-u)m}^{r,t}(\Phi) \tag{43}$$

$$+\int_s^t dr\,\pi_{m,s,t}(r)\int_E \big(p_r(x,dy)-p_{r-s}(x,dy)\big)$$
$$\times\int_0^1 du\, R^{r,t}_{y,um} * R^{r,t}_{y,(1-u)m}(\Phi).$$

Apply the inequality $\pi_{m,0,t}(r)\le 2mtc^{-1}(t-r)^{-2}\le 32mt^{-1}c^{-1}$ when $r\le 3t/4$ to bound the first term of (43) by $C(c)mst^{-1}\|\Phi\|_{\mathcal{M}(m)}$. Using the estimate

$$|\pi_{m,0,t}(r)-\pi_{m,s,t}(r)|\le \pi_{m,s,t}(r)2msc^{-1}(t-s)^{-1},$$

we can bound the second term of (43) by the same quantity. From the smoothing hypothesis (5) on $\{P_t\}$, we have

$$\|P_r\phi-P_{r-s}\phi\|_E=\left\|\int_{r-s}^r AP_q\phi\,dq\right\|_E\le \alpha s(r-s)^{-\beta}\|\phi\|_E.$$

Using this we can bound the third term of (43), when $s^{1/(1+\beta)}\le t/4$, by

$$\begin{aligned}
&\|\Phi\|_{\mathcal{M}(m)}\int_s^t \pi_{m,s,t}(r)\min\{\alpha s(r-s)^{-\beta},2\}\,dr\\
&\quad\le \|\Phi\|_{\mathcal{M}(m)}\left(2\int_s^{s+s^{1/(1+\beta)}}\pi_{m,s,t}(r)\,dr+\alpha s^{1/(1+\beta)}\int_{s+s^{1/(1+\beta)}}^t \pi_{m,s,t}(r)\,dr\right)\\
&\quad\le \|\Phi\|_{\mathcal{M}(m)}\left(2\int_s^{(s+s^{1/(1+\beta)})}\frac{2mt}{c(t-r)^2}\,dr+\alpha s^{1/(1+\beta)}\int_s^t \pi_{m,s,t}(r)\,dr\right)\\
&\quad\le C(t,c,\alpha)(1+m)s^{1/(1+\beta)}\|\Phi\|_{\mathcal{M}(m)}.
\end{aligned}$$

When $s^{1/(1+\beta)}\ge t/4$ we use the simple bound $2\|\Phi\|_{\mathcal{M}(m)}\le C(t)s\|\Phi\|_{\mathcal{M}(m)}$. Combining the bounds on the terms in (43) establishes the claim (42).

Our second claim is that the map $x\to R^{0,t}_{x,m}(\Phi)$ is in $D(A)$ and

$$\big|AR^{0,t}_{x,m}(\Phi)\big|\le C(t,c,\alpha)(1+m)\|\Phi\|_{\mathcal{M}(m)}\qquad\text{for all } x\in E. \tag{44}$$

To show this, we use (37) to write

$$\begin{aligned}
R^{0,t}_{x,m}(\Phi)&=\int_0^t \pi_{m,0,t}(r)P_rH_r(x)\,dr\\
&=\int_0^t \pi_{m,0,t}(r)P_rH_0(x)\,dr+\int_0^t \pi_{m,0,t}(r)P_r(H_r-H_0)(x)\,dr
\end{aligned}\tag{45}$$

where $H_r(y)=\int_0^1 du\,R^{r,t}_{y,um} * R^{r,t}_{y,(1-u)m}(\Phi)$. A standard argument shows that the first term on the right-hand side of (45) is in the domain $D(A)$ and

$$\begin{aligned}
&A\left(\int_0^t \pi_{m,0,t}(s)P_sH_0\,ds\right)\\
&\qquad=-\pi_{m,0,t}(0)H_0-\int_0^t \partial_r\pi_{m,0,t}(r)P_rH_0\,dr\\
&\qquad=\pi_{m,0,t}(0)(P_tH_0-H_0)+\int_0^t \partial_r\pi_{m,0,t}(r)(P_t-P_r)H_0\,dr.
\end{aligned}$$

[The unusual rearrangement here is to avoid a term of the form $\int_0^t |\partial_r \pi_{m,0,t}(r)|\,dr = O(m^{-1})$.] We now estimate the size of these two terms. Using $\|H_s\| \le \|\Phi\|_{\mathcal{M}(m)}$, we can bound the first term by $4mt^{-1}c^{-1}\|\Phi\|_{\mathcal{M}(m)}$. Using (5), we have $\|(P_t - P_r)H_0\|_E \le C(t-r)\alpha r^{-\beta}\|\Phi\|_{\mathcal{M}(m)}$. Combining this with the trivial bound $\|(P_t - P_r)H_0\|_E \le 2\|\Phi\|_{\mathcal{M}(m)}$ leads to

$$\|(P_t - P_r)H_0\|_E \le C(t,\alpha)(t-r)\|\Phi\|_{\mathcal{M}(m)} \qquad \text{for all } 0 < r \le t.$$

So

$$\begin{aligned}&\left\| \int_0^t \partial_r \pi_{m,0,t}(r)(P_t - P_r)H_0\,dr \right\|_E \\ &\qquad \le C(t,c,\alpha)\|\Phi\|_{\mathcal{M}(m)} \int_0^t |\partial_r \pi_{m,0,t}(r)|(t-r)\,dr,\end{aligned}$$

and an explicit computation shows this is bounded by $C(t,c,\alpha)(1+m)\|\Phi\|_{\mathcal{M}(m)}$. For the second term on the right-hand side of (45) we claim that

$$\begin{aligned} A\left(\int_0^t \pi_{m,0,t}(r)P_r(H_r - H_0)\,dr\right) &= \lim_{\delta \to 0} \int_0^t \pi_{m,0,t}(r)\frac{P_\delta - I}{\delta}P_r(H_r - H_0)\,dr \\ &= \int_0^t \pi_{m,0,t}(r)AP_r(H_r - H_0)\,dr. \end{aligned}$$

To justify this we use (42) twice to bound $\|H_r - H_0\|_E \le C(t,c,\alpha)(1 + m)r^{1/(1+\beta)}\|\Phi\|_{\mathcal{M}(m)}$ for $r \le t/2$. Then, using (5), we bound

$$\begin{aligned} \left\| \frac{P_\delta - I}{\delta} P_r(H_r - H_0) \right\|_E &\le \|AP_r(H_r - H_0)\|_E \\ &\le C(t,c,\alpha)(1+m)r^{-\beta}r^{1/(1+\beta)}\|\Phi\|_{\mathcal{M}(m)} \end{aligned}$$

whenever $r \le t/2$. The power $r^{-\beta}r^{1/(1+\beta)}$ is integrable near zero when $\beta \in (0, 2^{1/2})$. This leads to the domination required to pass to the limit $\delta \to 0$ in the above and also can be used to show that the result is bounded by $C(t,c,\alpha)(1 + m)\|\Phi\|_{\mathcal{M}(m)}$. Combining the various estimates, we have completed the proof of the claim (44).

Now we prove the lemma for a first derivative $D_x U_t^c \Phi$. Take Φ satisfying the growth conditions and let $\Phi^\nu(\mu) = \Phi(\nu + \mu)$. Lemma 11(a) shows that

$$\begin{aligned} \frac{P_\delta - I}{\delta} D_x U_t^c \Phi(\mu) &= \frac{2}{ct}\frac{P_\delta - I}{\delta} \iint \Phi(\nu_1 + \nu_2)R_x^t(d\nu_1)Q_\mu^t(d\nu_2) \\ &= \frac{2}{ct} \iint_0^\infty \left(\frac{P_\delta - I}{\delta} R_{x,m}^{0,t}(\Phi^\nu)\right) e_t(m)\,dm\,Q_\mu^t(d\nu) \\ &\to \frac{2}{ct} \iint_0^\infty AR_{x,m}^{0,t}(\Phi^\nu)e_t(m)\,dm\,Q_\mu^t(d\nu). \end{aligned}$$

To justify taking the limit under the integrals here we use the bound from (44) and the growth bound on Φ to show the domination

$$\begin{aligned}
&\frac{2}{ct}\int\int_0^\infty \|AR^{0,t}_{x,m}(\Phi^\nu)\|e_t(m)\,dm\,Q^t_\mu(d\nu)\\
&\quad\le C(t,c,\alpha)\int\int_0^\infty \|\Phi^\nu\|_{\mathcal{M}(m)}(1+m)e_t(m)\,dm\,Q^t_\mu(d\nu)\\
&\quad\le C(t,c,\alpha)\int\int_0^\infty \exp\bigl(\lambda\bigl(m+(\nu,1)\bigr)\bigr)(1+m)e_t(m)\,dm\,Q^t_\mu(d\nu)\\
&\quad\le C(t,c,\alpha)\int \exp(\lambda(\nu,1))\,Q^t_\mu(d\nu)\\
&\quad\le C(t,c,\alpha)\exp\bigl(2(\mu,1)/cT\bigr).
\end{aligned}$$

The final inequality is valid for $t\le T'$, for suitably chosen $T'>T$, by Lemma 2. A similar argument shows that the higher derivatives $x_1\to D_{x_1\cdots x_n}U^c_t\Phi$ are also in the domain $D(A)$.

Finally we come to the regularity of the first derivatives. The argument for the higher derivatives is very similar. Presumably the derivative $AD_xU^c_t\Phi(\mu)$ is continuous in $t>0, x, \mu$. We prove the slightly weaker conclusion of the theorem as this is all we need in the next section.

Examining all the terms in the expression for $AR^{0,t}_{x,m}(\Phi)$ given in this proof one sees, using the assumed continuity of P_rH_q and AP_rH_q when $r>0$, that for fixed $t>0$ and ν the map $x\to AR^{0,t}_{x,m}(\Phi^\nu)$ is continuous. Using this in (46) one finds that, for fixed $t>0$ and μ, the map $x\to AD_xU^c_t\Phi(\mu)$ is continuous. Using the growth condition on Φ and the bound (44), one can show that the maps

$$\left(\nu\to\int_0^\infty AR^{0,t}_{x,m'}(\Phi^\nu)e_t(m')\,dm'\right)_{x\in E}$$

are equicontinuous on $\mathcal{M}(m)$. Using this in (46), the growth condition and the Feller property of U^c_t, one finds that, for fixed $t>0$, the map $\mu\to AD_xU^c_t\Phi(\mu)$ is continuous, uniformly in x. Finally, the domination that guaranteed (46) also shows that $AD_xU^c_t\Phi(\mu)$ is locally bounded in t, uniformly over $x\in E$, $\mu\in\mathcal{M}(m)$ for any m. □

4. The martingale optimality argument and examples.

PROOF OF THEOREM 1. We fix Φ as in the statement of the theorem. We need to smooth the value function in time. Choose a mollifier as follows: Let $h:(0,\infty)\to[0,\infty)$ be smooth, supported in $(1,2)$ and satisfy $\int h(r)\,dr=1$. For $\delta>0$, set $h_\delta(r)=\delta^{-1}h(r\delta^{-1})$. Define

$$F_\delta(t,\mu)=\int_0^\infty h_\delta(s)U^c_{t+s}\Phi(\mu)\,ds. \tag{46}$$

The growth condition (6) on Φ involves the strict inequality $\lambda < 1/cT$. It therefore holds also for some $\lambda' < 1/cT'$ where $T' > T$. So, for $\delta > 0$ small enough, Lemma 11(c) implies that

$$\text{the functions } D_{x_1 \cdots x_n} U_t \Phi(\mu) \quad \text{are continuous for } t \le T + 2\delta,\ x_i \in E,\ \mu \in \mathcal{M}. \tag{47}$$

After the smoothing in time we find, using Lemma 13, that $F_\delta(t, \mu)$ satisfies all the hypotheses of Theorem 3. □

To follow the argument sketched in the Introduction, we need the next two lemmas, which are rigorous versions of (8) and (9).

LEMMA 14. *For $\delta > 0$ sufficiently small, we have*

$$L^c F_\delta(t, \mu) = 0 \qquad \text{for all } \mu \in \mathcal{M} \text{ and } t \le T.$$

PROOF. We work with δ small enough that (47) applies. Let $\{Y_t\}$ be a Dawson–Watanabe process with motion semigroup $\{P_t\}$, constant branching rate c and initial condition $Y_0 = \mu$. Fix t so that $t \le T + 2\delta$. For $s \le t$, let $\tau_K^s = \inf\{r : (Y_r, 1) \ge K\} \wedge s$ for values of $K > (\mu, 1)$. Applying the strong Markov property for this stopping time, we have

$$U_t^c \Phi = E(\Phi(Y_t)) = E\big(U^c_{t-\tau_K^s} \Phi(Y_{\tau_K^s})\big).$$

By integrating over the t variable, we obtain

$$F_\delta(t, \mu) = E\big(F_\delta(t - \tau_K^s, Y_{\tau_K^s})\big) \qquad \text{when } t \le T.$$

Now apply Theorem 3 to the function $F_\delta(t - q, Y_q)$ for $q \in [0, \tau_K^s]$. The stopping time and the continuity of F_δ and $L^c F_\delta$ ensure that the local martingale in this theorem is a true martingale. Hence we obtain

$$\begin{aligned} 0 &= \frac{1}{s}\big(E\big(F_\delta(t - \tau_K^s, Y_{\tau_K^s})\big) - F_\delta(t, \mu)\big) \\ &= E\bigg(\frac{1}{s}\int_0^{\tau_K^s} L^c F_\delta(t - q, Y_q)\, dq\bigg) \\ &= E\bigg(\frac{1}{s}\int_0^{s} I_{\{(Y_q, 1) < K\}} L^c F_\delta(t - q, Y_q)\, dq\bigg). \end{aligned}$$

Using the continuity of $L^c F_\delta$ and the continuous paths, we may let $s \downarrow 0$ in this equation and by dominated convergence obtain $L^c F_\delta(t, \mu) = 0$. □

LEMMA 15. *For $\delta > 0$ sufficiently small, we have*

$$D_{xx} F_\delta(t, \mu) \ge 0 \qquad \text{for all } \mu \in \mathcal{M} \text{ and } t \le T.$$

PROOF. We use Lemma 11 to represent the derivative $D_{xx}U_t^c\Phi(\mu)$. Let (Y_t^1, Y_t^2) be independent identically distributed random measures with law R_x^t. Let $Y_t^i(n)$ be finite approximations given by $Y_t^i(n) = nY_t^i(n \vee (Y_t^i, 1))^{-1}$. Then

$$\begin{aligned} D_{xx}U_t^c\Phi(\mu) & \\ &= E\big(\Phi(X_{t,\mu} + Y_t^1 + Y_t^2) - \Phi(X_{t,\mu} + Y_t^1) - \Phi(X_{t,\mu} + Y_t^2) + \Phi(X_{t,\mu})\big) \\ &= \lim_{n\to\infty} E\Big(\Phi\big(X_{t,\mu} + Y_t^1(n) + Y_t^2(n)\big) \\ &\qquad - \Phi\big(X_{t,\mu} + Y_t^1(n)\big) - \Phi\big(X_{t,\mu} + Y_t^2(n)\big) + \Phi(X_{t,\mu})\Big). \end{aligned}$$

The second equality follows from the growth bound on Φ. Since (Y_t^1, Y_t^2) are independent of $X_{t,\mu}$, the convexity hypothesis (3) implies that $D_{xx}U_t^c\Phi(\mu) \geq 0$ and integrating over the t variable completes the proof. □

Before starting the proof of Theorem 1, we make a reduction. Suppose that $\{X_t\}$ is a solution to $M(A,\sigma)$. We claim we may assume that $(X_0, 1) \leq L$ for some L. Suppose the theorem is proved under such a restriction. Define $\Omega_L = \{(X_0, 1) \leq L\}$ and $X_t^L = X_t \mathrm{I}_{\Omega_L}$. Then X^L is a still a solution to $M(\sigma, A)$ and has initial mass bounded by L. The conclusion of the theorem then compares $E(\Phi(X_t^L))$ and $E(U_t^c\Phi(X_0^L))$. Splitting both expectations into two parts, one over Ω_L and one over Ω_L^c, we can apply monotone convergence as $L \to \infty$ to obtain the conclusion for X_t. Thus we now assume $(X_0, 1) \leq L$.

Using the local boundedness of σ, choose stopping times T_K^1 so that $T_K^1 \uparrow \infty$ as $K \to \infty$ and $|\sigma_t| I(t < T_K^1) \leq K$. Set $T_K^2 = \inf\{t : (X_t, 1) \geq K\}$, $T_K = T_K^1 \wedge T_K^2$. Fix $t \in (0, T]$. We apply Theorem 3 to the function $F_\delta(t - s, X_s)$ for $s \in [0, t \wedge T_K]$. The definition of the stopped processes and the continuity of F_δ and its derivatives imply that the local martingale in this theorem is a true martingale. Hence

$$\begin{aligned} & E\big(F_\delta\big(t - (t \wedge T_K), X_{t\wedge T_K}\big)\big) - E(F_\delta(t, X_0)) \\ &\quad = E\bigg(\int_0^{t\wedge T_K} L^\sigma F_\delta\big(t - s, X_s\big)\, ds\bigg) \\ &\quad = E\bigg(\int_0^{t\wedge T_K} L^c F_\delta(t - s, X_s)\, ds\bigg) \\ &\qquad + E\bigg(\int_0^{t\wedge T_K} \int \big(\sigma_s(x) - c\big) D_{xx} F_\delta(t - s, X_s) X_s(dx)\, ds\bigg). \end{aligned} \tag{48}$$

The second equality follows by comparing the expressions for L^σ and L^c. The first term on the right-hand side of (48) is zero by Lemma 14. Lemma 15 shows that the last term on the right-hand side of (48) is nonnegative if $\sigma \geq c$ and nonpositive if $\sigma \leq c$.

Turning to the left-hand side of (48), we know that $F_\delta(t,\mu) \to U_t^c\Phi(\mu)$ as $\delta \to 0$, uniformly over compacts in t, μ. Passing to this limit in (48), we obtain

$$E\big(U^c_{t-(t\wedge T_K)}\Phi(X_{t\wedge T_K})\big) \le E\big(U_t^c\Phi(X_0)\big) \tag{49}$$

when $\sigma \le c$ and the reverse inequality when $\sigma \ge c$. It remains only to let $K \to \infty$ and we consider each of the three cases stated in the theorem. When $\sigma \le c$ we can use Fatou's lemma to obtain the desired comparison. When Φ is bounded we can apply the dominated convergence theorem. This leaves only case (c) where $c \le \sigma \le \bar{c}$ and $\Phi(\mu) \le C\exp(\lambda(\mu,1))$ for some $\lambda < 1/2\bar{c}T$. We split the left-hand side of (49) into two parts. On the set $\{T_K > t\}$ we have

$$E\big(\Phi(X_t)I_{\{T_K>t\}}\big) \to E(\Phi(X_t)) \qquad \text{as } K\to\infty.$$

Lemma 2 implies that $U_t^c\Phi(\mu) \le C\exp(2\lambda(\mu,1))$ for all $t \le T$. So on the set $\{T_K \le t\}$ we have

$$\begin{aligned}
&E\big(U^c_{t-T_K}\Phi(X_{T_K})I_{\{T_K\le t\}}\big)\\
&\quad \le C\exp(2\lambda K)P(T_K \le t)\\
&\quad \le C\exp(2\lambda K)\exp(-K/\bar{c}T)E\Big(\sup_{t\le T}\exp\big((1/\bar{c}T)(X_s,1)\big)\Big)\\
&\quad \le C\exp(2\lambda K)\exp(-K/\bar{c}T)\exp(2L/\bar{c}T),
\end{aligned}$$

using Markov's inequality and Lemma 2 for the last two inequalities. Letting $K \to \infty$ and combining the two parts gives the comparison in the third and final case. □

EXAMPLES.

Extension to locally compact E. Many examples of measure valued processes are studied when $E = \mathbb{R}^d$ or $E = \mathbb{Z}^d$. To apply the comparison in these cases one can consider them as living on the compactification of E, as follows. Let E be a locally compact metric space and let $C_0(E)$ be the space of continuous functions on E that converge to 0 at infinity. Let $\{P_t\}$ be a strongly continuous Markov $C_0(E)$ semigroup with generator A. Suppose $\{X_t\}$ is a process taking values in the space of finite measures on E and that:

(i) $t \to X_t(\phi)$ is continuous for all $\phi \in C_0(E)$ and for the constant function $\phi = 1$;

(ii) $\{X_t\}$ solves equations (1) and (2), for some locally bounded σ, for all $\phi \in D(A)$ and for the constant function $\phi = 1$ (where we set $A1 = 0$).

Let $\bar{E}$ be the one point compactification of E. We can identify the space $C(\bar{E})$ with the functions on E that have a limit at infinity, each of which can be written as a constant plus a function in $C_0(E)$. The semigroup $\{P_t\}$ extends to a semigroup $\{\bar{P}_t\}$ on $C(\bar{E})$ by setting, for $\bar{\phi} \in C(\bar{E})$,

$$\bar{P}_t\bar{\phi}(x) = \begin{cases} P_t\phi(x), & \text{if } x \neq \infty, \text{ where } \phi := \bar{\phi}|_E, \\ \bar{\phi}(\infty), & \text{if } x = \infty, \text{ where } \bar{\phi}(\infty) = \lim_{x\to\infty} \bar{\phi}(x). \end{cases}$$

It can then be checked that $\{\bar{P}_t\}$ is still strongly continuous and Markov and has generator $\bar{A}$ where $D(\bar{A}) = \{\bar{\phi} : \bar{\phi}|_E - \bar{\phi}(\infty)$ is an element of $D(A)\}$ and

$$\bar{A}\bar{\phi}(x) = \begin{cases} A\phi(x), & \text{if } x \neq \infty, \\ 0, & \text{if } x = \infty. \end{cases}$$

Finally, we extend σ to $\bar{\sigma}$ by setting $\bar{\sigma}(t, \infty) = 0$. If we now consider $\{X_t\}$ as a process taking values in the space of finite measures on $\bar{E}$, giving no mass to the point at infinity, then $t \to X_t(\bar{\phi})$ is continuous for all $\bar{\phi} \in C(\bar{E})$ and $\{X_t\}$ is a solution to $M(\bar{A}, \bar{\sigma})$ and we may apply the results of the theorem.

To apply our results to two important cases discussed in the literature, we need to check the hypothesis on the generator A. If A is the Laplacian on $\mathbb{R}^d$, we may take the good core to consist of the algebra generated by the Schwarz space of rapidly decaying test functions and the function 1. If A is a bounded generator on the lattice $\mathbb{Z}^d$, for example the generator of a continuous time Markov chain with bounded jump rates, we may take the algebra generated by $C_0(\mathbb{Z}^d)$ and the function 1. In both cases the smoothing hypothesis (5) holds.

Ergodicity. The application studied in Cox, Fleischmann and Greven [3] was to studying ergodicity problems for systems of interacting SDEs indexed by the lattice $\mathbb{Z}^d$. Here the interest is in translation invariant initial conditions. Therefore, to obtain analogous results, one needs to extend our results to processes with infinite mass. Typically one expects, although uniqueness in law would be a usual ingredient of the proof, that solutions with initial conditions having infinite mass can be approximated by solutions with finite initial mass, and then the comparison results will extend to the more general setting. As an example consider the case of Dawson–Watanabe process with a Brownian motion process, which is known as super-Brownian motion. In dimensions $d = 1, 2$ solutions with translation invariant initial conditions become locally extinct, in that $X_t(\phi) \to 0$ in probability, for compactly supported $\phi \geq 0$. Using the functional $\Phi(\mu) = \exp(-\lambda\mu(\phi))$, one can use the comparison argument to show the same holds for interacting processes with branching rates that are bounded below. In dimensions $d \geq 3$ there are nonzero stationary measures. Using second moments $\Phi(\mu) = (\mu(\phi))^2$, one can then use the comparison principle to show that local extinction does not occur for interacting models with branching rates bounded above. Together with compactness arguments this leads to the proof of existence of nonzero invariant measures. See [3] and cited references for these techniques.

Hitting sets. For super-Brownian motion on $\mathbb{R}^d$ there are useful bounds on the probability of charging small balls, see Dawson, Iscoe and Perkins [5]. For example, in $d \geq 3$, if $Y_0 = \mu$,

$$P\big(Y_t(B(x_0, \varepsilon) > 0\big) \leq C\varepsilon^{d-2} \int (2\pi t)^{-d/2} \exp(-|x_0 - x|^2/2t)\mu(dx). \tag{50}$$

These imply that if C has zero $d-2$ Hausdorff measure then $P(Y_t(C) > 0) = 0$. Let d be the metric on the space E and fix a closed subset C. Define, for $\alpha, \varepsilon > 0$,

$$\phi_\varepsilon(x) = 1 - \frac{d(x, C) \wedge \varepsilon}{\varepsilon}, \qquad \Phi(\mu) = \exp\big(-\alpha(\mu, \phi_\varepsilon)\big).$$

Function Φ is continuous and satisfies the parallelogram rule (4). Letting $\varepsilon \downarrow 0$ the function ϕ_ε converges to the indicator of C and then as $\alpha \to \infty$ the function Φ converges to the indicator of the set $\{\mu(C) = 0\}$. Applying the comparison result to Φ and taking the above limits, we obtain the following comparison: If $\{X_t\}$ solves $M(A, \sigma)$ and $\{Y_t\}$ is a Dawson–Watanabe process with branching rate c and with the same initial condition then

$$P\big(X_t(C) > 0\big) \geq P\big(Y_t(C) > 0\big) \qquad \text{for closed } C,$$

when $\sigma \leq c$ and the reverse inequality when $\sigma \geq c$. This confirms the intuition that the more branching there is the greater the clustering and the lower the chance of hitting sets.

Regularity of solutions. It is well known (see Dawson [4]) that, for super-Brownian motion in dimensions $d \geq 2$, the closed support of Y_t at time $t > 0$ has Hausdorff dimension 2. The hitting estimates (50) provide a simple proof of this fact. Indeed covering $\mathbb{R}^d$ by a lattice of boxes of length r the hitting estimates lead immediately to the bound $E(N(r)) \leq Cr^{-2}$ on the first moment of the number $N(r)$ of boxes that are charged by Y_t. Using this, a Borel–Cantelli argument gives a sequence of covers for the support of Y_t showing that the $2 + \varepsilon$ Hausdorff measure of the support is zero for any $\varepsilon > 0$. Since the hitting estimates carry over by the comparison argument one obtains the same singularity for interacting models with branching bounded below by a constant.

To obtain lower bounds on the dimension of the support note that the usual Frostman energy approach uses the functional

$$E(\Phi(Y_t)) = E\bigg(\iint |x - y|^{-\alpha} Y_t(dx) Y_t(dy)\bigg).$$

This energy is finite for super-Brownian motion started from deterministic finite initial measures when $\alpha < 2$ and $t > 0$. For $\phi \geq 0$ the functional $\Phi(\mu) = \iint \phi(x, y)\mu(dx)\mu(dy)$ satisfies the convexity hypothesis (4). By approximating the energy by continuous second moments of this sort the comparison argument applies. This implies the support has dimension at least 2 for a class of interacting models with branching bounded above by a constant.

Existence of densities. Super-Brownian motion process in dimension $d = 1$ has a continuous density. Roelly–Coppoletta [12] used spectral methods as a simple way to investigate densities at a fixed time. Set $e_\theta = \exp(i\theta x)$, acting on $E = \mathbb{R}$, and

$$\Phi_N(\mu) = \int_{-N}^{N} |\mu(e_\theta)|^2 \, d\theta.$$

The randomized parallelogram rule (3) becomes, after some simplification,

$$\begin{aligned} &E\big(\Phi_N(\mu + Z + \bar{Z}) - \Phi_N(\mu + Z) - \Phi_N(\mu + \bar{Z}) + \Phi_N(\mu)\big) \\ &\quad = \int_{-N}^{N} E\big(Z(e_\theta)\bar{Z}(e_{-\theta}) + Z(e_{-\theta})\bar{Z}(e_\theta)\big)\, d\theta \\ &\quad = 2\int_{-N}^{N} \big(E(Z(\cos(\theta\cdot)))\big)^2 + \big(E(Z(\sin(\theta\cdot)))\big)^2 \, d\theta \\ &\quad \ge 0. \end{aligned}$$

[Note that in this example the parallelogram rule (4) fails.] So the comparison theorem is applicable and, by letting $N \to \infty$, we obtain the comparison for the function

$$E\bigg(\int_{\mathbb{R}} |X_t(e_\theta)|^2 \, d\theta\bigg).$$

For initial conditions $\mu(dx) = f(x)\,dx$ with $f \in L^1 \cap L^2$, this expectation is finite for one-dimensional super-Brownian motion. By comparison it is finite for solutions to $M(\Delta/2, \sigma)$ when $\sigma \le c$, and Plancherel's theorem then implies that these processes have an L^2 density at any fixed time $t > 0$. In [8], a class of interacting branching processes on $\mathbb{R}$ is shown to have continuous densities.

A counterexample. We searched for some time for simpler sufficient conditions on Φ ensuring the comparison result holds. The following example, which we found surprising, stopped us wasting time on certain false conjectures. Fix nonnegative $f, g \in C(E)$ and consider the functional $\Phi(\mu) = \max\{\mu(f), \mu(g)\}$. Note that Φ is nice: it has quadratic growth and, being the maximum of two linear functions, Φ is a convex function on $\mathcal{M}$. The representation for the second derivatives in Lemma 11 and the simple fact that

$$\max\{a + b, c + d\} - \max\{a, c\} - \max\{b, d\} \le 0 \qquad \text{for all real } a, b, c, d,$$

show that $D_{xx} U_t^c \Phi(0) \le 0$ for all $t > 0$. Moreover, it is clear that except in very special circumstances this will be a strict inequality, so that $U_t^c \Phi$ is not convex in the direction of point masses. Moreover, using the continuity of $D_{xx} U_t^c \Phi(\mu)$, one expects Theorem 1 to fail, in that for solutions X_t to the martingale problem $M(A, \sigma)$ with $\sigma \le c$ and $X_0 = \mu$, one expects the comparison

$$E(\Phi(X_t)) > U_t^c \Phi(\mu) \qquad \text{for } \mu \text{ with } (\mu, 1) \text{ sufficiently small.}$$

REFERENCES

[1] ASSING, S. (1999). Comparison of systems of stochastic partial differential equations. *Stochastic Process. Appl.* **82** 259–282.

[2] BARLOW, M. T., EVANS, S. N. and PERKINS, E. A. (1991). Collision local times and measure-valued processes. *Canad. J. Math.* **43** 897–938.

[3] COX, T., FLEISCHMANN, K. and GREVEN, A. (1996). Comparison of interacting diffusions and an application to their ergodic theory. *Probab. Theory Related Fields* **105** 513–528.

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

[5] DAWSON, D. A., ISCOE, I. and PERKINS, E. A. (1989). Super-Brownian motion: Path properties and hitting properties. *Probab. Theory Related Fields* **83** 135–205.

[6] ETHIER, S. N. and KURTZ, T. (1986). *Markov Processes*: *Characterization and Convergence*. Wiley, New York.

[7] KURTZ, T. (2001). Particle representations for measure-valued population processes with spatially varying birth rates. Preprint.

[8] LOPEZ, M. M. (1996). Path properties and convergence of interacting superprocesses. Ph.D. thesis, Univ. British Columbia.

[9] MELEARD, S. and ROELLY, S. (1992). Interacting branching measure processes. In *Stochastic Partial Differential Equations and Applications* (G. Da Prato and L. Tubaro, eds.) 246–256. Longman, New York.

[10] METIVIER, M. (1985). Weak convergence of measure-valued processes using Sobolev imbedding techniques. *Stochastic Partial Differential Equations and Applications. Lecture Notes in Math.* **1236** 172–183. Springer, Berlin.

[11] PERKINS, E. A. (1995). On the martingale problem for interactive measure-valued branching diffusions. *Mem. Amer. Math. Soc.* **115** 1–89.

[12] ROELLY-COPPOLETTA, S. (1986). A criterion of convergence of measure-valued processes: Application to measure branching processes. *Stochastics* **17** 43–65.

[13] ROGERS, L. C. G. and WILLIAMS, D. (1994). *Diffusions*, *Markov Processes and Martingales* **1**: *Foundations*, 2nd ed. Wiley.

[14] WATANABE, S. (1968). A limit theorem of branching processes and continuous state branching. *J. Math. Kyoto Univ.* **8** 141–167.

DEPARTMENT OF STATISTICS
UNIVERSITY OF WARWICK
COVENTRY CV4 7AL
UNITED KINGDOM
E-MAIL: s.d.jacka@warwick.ac.uk
r.tribe@warwick.ac.uk