RENORMALIZATION OF HIERARCHICALLY INTERACTING ISOTROPIC DIFFUSIONS

F. den Hollander
J. M. Swart

Abstract

We study a renormalization transformation arising in an infinite system of interacting diffusions. The components of the system are labeled by the N-dimensional hierarchical lattice ($N \geq 2$) and take values in the closure of a compact convex set $\overline{D} \subset \mathbb{R}^d$ ($d \geq 1$). Each component starts at some $\theta \in D$ and is subject to two motions: (1) an isotropic diffusion according to a local diffusion rate $g : \overline{D} \to [0, \infty)$ chosen from an appropriate class; (2) a linear drift towards an average of the surrounding components weighted according to their hierarchical distance. In the local mean-field limit $N \to \infty$, block averages of diffusions within a hierarchical distance k, on an appropriate time scale, are expected to perform a diffusion with local diffusion rate $F^{(k)}g$, where $F^{(k)}g = (F_{c_k} \circ \cdots \circ F_{c_1})g$ is the k-th iterate of *renormalization transformations* F_c ($c > 0$) applied to g. Here the c_k measure the strength of the interaction at hierarchical distance k. We identify F_c and study its orbit $(F^{(k)}g)_{k \geq 0}$. We show that there exists a 'fixed shape' g^* such that $\lim_{k\to\infty} \sigma_k F^{(k)} g = g^*$ for all g, where the σ_k are normalizing constants. In terms of the infinite system, this property means that there is *complete universal behavior* on large space-time scales.

Our results extend earlier work for $d = 1$ and $\overline{D} = [0, 1]$ resp. $[0, \infty)$. The renormalization transformation F_c is defined in terms of the ergodic measure of a d-dimensional diffusion. In $d = 1$ this diffusion allows a Yamada-Watanabe-type coupling, its ergodic measure is reversible and the renormalization transformation F_c is given by an explicit formula. All this breaks down in $d \geq 2$, which complicates the analysis considerably and forces us to new methods. Part of our results depend on a certain martingale problem being well-posed.

AMS 1991 subject classifications. 60K35 (primary), 47H99 (secondary).
Key words and phrases. Interacting diffusions, hierarchical lattice, renormalization, non-linear integral transform, universality.

Acknowledgment. Research supported by the Dutch Association for Mathematical Physics.

Contents

0 Introduction

In this paper we study a renormalization transformation that arises in the study of a system of hierarchically interacting diffusions. Our study is part of a larger area where the goal is to understand universal behavior on large space-time scales of stochastic systems with interacting components. In a recent series of papers ([1], [5], [6], [7]), it was shown how renormalization techniques can be used to give a rigorous analysis of a model, described below, consisting of interacting diffusions indexed by the hierarchical lattice and taking values in the state space $[0, 1]$. In the meantime the analysis has been generalized to the state space $[0, \infty)$ ([2], [8]).

So far, the model has only been treated completely in the case of a one-dimensional state space (although some limited results for the infinite-dimensional state space of probability measures on $[0, 1]$ can be found in [9], [10]). The present paper investigates a class of isotropic models with state space $\overline{D}$, where $D \subset \mathbb{R}^d$ $(d \geq 1)$ is open, bounded and convex. To help the reader, we use the remainder of this section to present an overview of the known results for the case $d = 1$, together with a heuristic view on what is behind these results. This overview provides the essential motivation for section 1, where we state our new results for the case $d \geq 2$ and formulate some open problems. Proofs appear in sections 2–4.

0.1 Genetic diffusions

Our model finds its origin in population dynamics. Consider a gene that comes in $d+1$ types ('alleles'). Consider a population consisting of n individuals, each carrying one copy of the gene ('haploid organisms'). At any time the population may be described by a point x in the discrete simplex

$$K_d^n := \{x = (x_1, \ldots, x_d) \in \tfrac{1}{n}\mathbb{Z}^d : x_i \geq 0,\ \sum_{i=1}^{d} x_i \leq 1\}. \tag{0.1}$$

We interpret $x_1, \ldots, x_d, 1 - \sum_{i=1}^{d} x_i$ as the proportions of alleles $1, \ldots, d+1$. Frequencies of alleles are supposed to change due to 'random sampling' and 'migration'.

Random sampling is a random process by which some alleles may occasionally produce more offspring than others. We can model it as a Markov evolution on K_d^n by replacing pairs of individuals after an exponential waiting time with mean 1. A pair is replaced in the following manner: we choose one individual of the pair at random, determine its allele and replace both individuals by individuals with this allele.

Migration is a random process that we can model by introducing a huge reservoir of individuals, with gene frequencies $\theta_1, \ldots, \theta_d, 1 - \sum_{i=1}^{d} \theta_i$, and letting each individual in the population be replaced with rate c by an individual of the reservoir.

The generator A of the resulting process (migration and random sampling) is given by

$$\begin{aligned}(A_n f)(x) &= cn \sum_{i,j=1}^{d+1} \theta_i x_j \left[f(x + \frac{e^i}{n} - \frac{e^j}{n}) - f(x) \right] \\ &\quad + n^2 \sum_{i,j=1}^{d+1} x_i x_j \left[f(x + \frac{e^i}{n} - \frac{e^j}{n}) - f(x) \right],\end{aligned} \tag{0.2}$$

where $x = (x_1, \ldots, x_d)$ and $e^i = (e^i_1, \ldots, e^i_d)$ with $e^i_j = \delta_{ij}$ for $i = 1, \ldots, d$. In (0.2) we have additionally defined $x_{d+1} = 1 - \sum_{i=1}^d x_i$ and $\theta_{d+1} = 1 - \sum_{i=1}^d \theta_i$, and put e^{d+1} to be the zero vector in $\mathbb{R}^d$.

In the limit $n \to \infty$ the gene frequencies take values in the d-dimensional simplex

$$K_d := \{x \in \mathbb{R}^d : x_i \geq 0, \ \sum_{i=1}^d x_i \leq 1\}. \tag{0.3}$$

On functions $f \in \mathcal{C}^2(K_d)$ the generator A_n can be seen to converge, in an appropriate sense, to

$$(Af)(x) = \Big(\sum_{i=1}^d c(\theta_i - x_i) \tfrac{\partial}{\partial x_i} + \sum_{i,j=1}^d x_i(\delta_{ij} - x_j) \tfrac{\partial^2}{\partial x_i \partial x_j} \Big) f(x). \tag{0.4}$$

The matrix $x_i(\delta_{ij} - x_j)$ is the Wright-Fisher diffusion matrix. Similar models, with slightly more complicated random sampling mechanisms, yield similar differential operators with different diffusion matrices. Provided the martingale problem for these operators is well-posed, it is often possible to show that the discrete process on K_d^n converges in law to the diffusion with generator A (see [11] for details).

Let us consider the case $d = 1$ and let us introduce the following objects.

1. ('state space') $K_1 = [0, 1]$.
2. ('diffusion function') $\mathcal{H}_{Lip}$ is the class of functions $g : [0, 1] \to [0, \infty)$ satisfying

$$\begin{array}{rl} \text{(i)} & g = 0 \text{ on } \{0, 1\} \\ \text{(ii)} & g > 0 \text{ on } (0, 1) \\ \text{(iii)} & g \text{ is Lipschitz continuous on } [0, 1]. \end{array} \tag{0.5}$$

3. ('attraction point') $\theta \in [0, 1]$.
4. ('attraction constant') $c \in (0, \infty)$.

With these ingredients we consider the following Stochastic Differential Equation (SDE) on $[0, 1]$:

$$dX_t = c(\theta - X_t)dt + \sqrt{2g(X_t)}dB_t \qquad (t \geq 0), \tag{0.6}$$

where $(B_t)_{t\geq 0}$ is standard Brownian motion. We call (0.6) 'the basic diffusion equation'. We define a linear operator A with domain $\mathcal{C}^2[0,1]$ (the two times continuously differentiable real functions on $[0,1]$) by putting

$$(Af)(x) := [c(\theta - x)\tfrac{\partial}{\partial x} + g(x)\tfrac{\partial^2}{\partial x^2}]f(x). \tag{0.7}$$

The following is known[1]:

Theorem 0.1 *For each $g \in \mathcal{H}_{Lip}$, $\theta \in [0,1]$ and $c \in [0,\infty)$, and for each initial distribution on $[0,1]$, the SDE (0.6) has a unique strong solution $(X_t)_{t\geq 0}$. The martingale problem for A in (0.7) is well-posed and the law of $(X_t)_{t\geq 0}$ solves the martingale problem for A. The operator A has a unique extension to a generator of a Feller semigroup and $(X_t)_{t\geq 0}$ is the associated Feller process.*

The choice $g(x) = x(1-x)$ corresponds to the Wright-Fisher case. The diffusion equation (0.6) and its generalizations to higher dimensions will play a key role in the present paper. In the following sections we show how it has been used as a starting point for the construction and analysis of an infinite system of interacting diffusions.

0.2 The hierarchical model

In the model described in the previous section, all individuals have equal chances of interaction with all other individuals. A more realistic model takes into account the effects of isolation by distance. To this aim, we introduce the following additional objects:

5. ('index space') For $N \geq 2$, let Ω_N be the N-dimensional hierarchical lattice

$$\Omega_N := \Big\{(\xi_i)_{i\geq 1} : \xi_i \in \{0,1,\ldots,N-1\}, \xi_i \neq 0 \text{ finitely often}\Big\}. \tag{0.8}$$

With componentwise addition (mod N), Ω_N is a countable group.

[1] To get these results, extend the diffusion function g by putting $g(x) = g(1)$ $(x \geq 1)$, $g(x) = g(0)$ $(x \leq 0)$, and extend the drift by the same recipe. It is easy to show that any solution $(X_t)_{t\geq 0}$ of the SDE on $\mathbb{R}$ satisfies $P[X_t \in [0,1]\ \forall t \geq 0] = 1$. Now, by Skorohod's Theorem ([14], Theorem 5.4.22), there exists a weak solution of (0.6). The Yamada-Watanabe argument ([14], Proposition 5.2.13) gives strong uniqueness, and therefore strong existence as well as weak uniqueness ([14], Propositions 5.3.23 and 5.3.20). It follows that the martingale-problem is well-posed ([14], Corollary 5.4.8 and 5.4.9). The process $(X_t)_{t\geq 0}$ has the Feller property ([20], Corollary 11.1.5) and its generator G clearly extends A. In fact, it is the only generator of a Feller semigroup to do so. For let $\tilde{G}$ be another generator extending A, then there exists an associated Feller process $(\tilde{X}_t)_{t\geq 0}$ ([11], Theorem 4.2.7) that solves the martingale problem for A ([11], Theorem 4.1.7). It follows that $(\tilde{X}_t)_{t\geq 0}$ and $(X_t)_{t\geq 0}$ have the same distribution for all initial conditions, and hence $G = \tilde{G}$.

In the special case that $g \in \mathcal{C}^2[0,1]$, it is known that G is the closure of A ([11], Theorem 8.2.1), but for general $g \in \mathcal{H}_{Lip}$ this seems to be an open problem. (In this respect, the loose remark in [1], page 7, that the closure of the operator G mentioned there generates a Feller semigroup seems unfounded.)

6. ('distance') Let $d : \Omega_N \times \Omega_N \to \mathbb{N}_0$ be the hierarchical distance

$$d(\xi, \eta) := \min\{j \geq 0 : \xi_i = \eta_i \text{ for all } i > j\}. \tag{0.9}$$

7. ('interaction constants') Let $(c_k)_{k\geq 1}$ be strictly positive constants, satisfying

$$\sum_{k=1}^{\infty} c_k^{-1} = \infty \tag{0.10}$$

$$\sum_{k=1}^{\infty} c_k N^{-k} < \infty. \tag{0.11}$$

8. ('noise') Let $(\{B_\xi(t)\}_{\xi\in\Omega_N})_{t\geq 0}$ be an i.i.d. collection of standard Brownian motions.

With the above ingredients, we consider the process

$$X^N = \left(X^N(t)\right)_{t\geq 0} = \left(\left\{X^N_\xi(t)\right\}_{\xi\in\Omega_N}\right)_{t\geq 0} \tag{0.12}$$

with state space $[0,1]^{\Omega_N}$ given by the following set of coupled SDE's:

$$\begin{aligned} dX^N_\xi(t) &= \textstyle\sum_{k=1}^{\infty} c_k N^{1-k}\left[X^{N,k}_\xi(t) - X^N_\xi(t)\right]dt + \sqrt{2g\left(X^N_\xi(t)\right)}\,dB_\xi(t) \\ X^N_\xi(0) &= \theta \qquad\qquad (t \geq 0, \xi \in \Omega_N), \end{aligned} \tag{0.13}$$

where $X^{N,k}_\xi(t)$ is the block average

$$X^{N,k}_\xi(t) := \frac{1}{N^k} \sum_{\eta:\, d(\eta,\xi)\leq k} X^N_\eta(t) \qquad (k = 0, 1, 2 \ldots). \tag{0.14}$$

The system in (0.13) can be interpreted as a model for the time evolution of gene distributions in an infinite population (see [5], [17] for the origin of the model and [11], Chapter 10, for more background). The population is organized in sites, groups, clans, villages etc., where N sites form a group, N groups form a clan, N clans form a village, and so on. The index space Ω_N labels sites by numbering sites within a group by a number $\xi_1 = 0, \ldots, N-1$, numbering groups within a clan by a number $\xi_2 = 0, \ldots, N-1$, and so on. (For example, if the distance between two sites ξ and η is $d(\xi, \eta) = 2$, then ξ and η share the same village and clan but belong to different groups.) The proportion of allele 1 at a given site ξ at some time t is described by $X^N_\xi(t)$. Initially, all proportions are supposed to be θ, and they evolve due to migration and random sampling, as in (0.6). However, the migration is now supposed to cause interaction between sites mutually (see (0.16–0.17) below), instead of between sites and some infinite reservoir. The meaning of the numbers $X^{N,k}_\xi(t)$ is the following:

$X_\xi^{N,0}(t) = X_\xi^N(t)$ is the proportion of individuals of allele 1 at site ξ; $X_\xi^{N,1}(t)$ is the proportion in the group that ξ belongs to; $X_\xi^{N,2}(t)$ is the proportion in the clan that ξ belongs to, and so on. We call the set $\{\eta : d(\eta,\xi) \leq k\}$ the 'k-block' around ξ and the numbers $X_\xi^{N,k}(t)$ the 'k-block averages' around ξ.

The factor $c_k N^{1-k}$ in (0.13) describes the strength of the interaction (due to migration) between a site ξ and the k-block around ξ. The strength of the attraction decays by a factor $1/N$ each time we go up one step in the hierarchy. As we shall see later, precisely this decay will give rise to non-trivial behavior in the limit as $N \to \infty$.

The next theorem follows from [19], Theorem 3.2:

Theorem 0.2 *Let $N \geq 2$, $g \in \mathcal{H}_{Lip}$, $c_k \in [0,\infty)$ $(k \geq 1)$, $\sum_k c_k N^{-k} < \infty$ and $\theta \in [0,1]$. Then the system of SDE's in (0.13) has a unique strong solution satisfying*

$$P[X_\xi^N(t) \in [0,1] \quad \forall\, \xi \in \Omega_N,\ t \geq 0] = 1. \tag{0.15}$$

We need to check that [19], Assumption [B-2]$'$ is satisfied. The drift term in (0.13) can be rewritten as

$$\sum_{k=1}^{\infty} c_k N^{1-k}\left[X_\xi^{N,k}(t) - X_\xi^N(t)\right]dt = \sum_{\eta\in\Omega_N} a_N(\xi,\eta)\left[X_\eta^N(t) - X_\xi^N(t)\right]dt, \tag{0.16}$$

where

$$a_N(\xi,\eta) = \sum_{k=d(\xi,\eta)}^{\infty} c_k N^{1-2k}. \tag{0.17}$$

Hence the drift term is in fact a pair interaction between the different components. A little calculation shows that $\sum_{\eta\in\Omega_N} a_N(\xi,\eta) = \sum_{k=1}^\infty c_k N^{1-k}\ \forall \xi \in \Omega_N$. Condition (0.11) is therefore exactly what is required in [19], Assumption [B-2]$'$.

0.3 The local mean-field limit $N \to \infty$

We shall study the system in (0.13) in the limit as $N \to \infty$. The 1-block average $X_\xi^{N,1}(t)$ is the average of a large number of diffusions that behave independently apart from their linear drift towards block averages.

Let Δt be small and let $\Delta X_\xi^N(t) := X_\xi^N(t+\Delta t) - X_\xi^N(t)$. Let $\mathcal{F}_t$ be the σ-field of events up to time t. Then (0.13) can, in a heuristic way, be rewritten as

$$\begin{aligned} E[\Delta X_\xi^N(t) \mid \mathcal{F}_t] &\cong \textstyle\sum_{k=1}^\infty c_k N^{1-k}\left[X_\xi^{N,k}(t) - X_\xi^N(t)\right]\Delta t \\ E[\Delta X_\xi^N(t)\Delta X_\eta^N(t) \mid \mathcal{F}_t] &\cong \delta_{\xi,\eta} 2g(X_\xi^N(t))\Delta t. \end{aligned} \tag{0.18}$$

It follows that for the 1-block averages we have

$$\begin{aligned} E[\Delta X_\xi^{N,1}(t) \mid \mathcal{F}_t] &\cong \textstyle\sum_{k=2}^\infty c_k N^{1-k}\left[X_\xi^{N,k}(t) - X_\xi^{N,1}(t)\right]\Delta t \\ E[\Delta X_\xi^{N,1}(t)\Delta X_\eta^{N,1}(t) \mid \mathcal{F}_t] &\cong 1_{\{d(\xi,\eta)\leq 1\}} 2N^{-2} \textstyle\sum_{\zeta:\, d(\zeta,\xi)\leq 1} g(X_\zeta^N(t))\Delta t. \end{aligned} \tag{0.19}$$

Note that in the first line the term with $k = 1$ drops out. Note further that in the second line the sum $\sum_{\zeta:\, d(\zeta,\xi)\leq 1}$ is over N terms. Hence both expectations are of order N^{-1}. We are therefore led to believe that the 1-block average $X_\xi^{N,1}(t)$ moves slowly w.r.t. $X_\xi^N(t)$, namely, its time scale is Nt rather than t. For large N this means that $X_\xi^{N,1}(t)$ stays essentially fixed at its initial value θ. Inserting this into (0.13) and neglecting terms of order $1/N$, we see that the single components $X_\xi^N(t)$ satisfy a limiting SDE of the type (0.6) with $c = c_1$. The limit $N \to \infty$ thus corresponds to a 'local mean-field' limit. On the local space scale of 1-blocks, the interaction reduces to a linear drift towards an essentially fixed block average, so that the single components are asymptotically independent (in physics language: the system shows 'propagation of chaos'). This behavior, however, occurs only locally. We shall see later that on larger space scales the interaction still gives rise to nontrivial correlations between components.

A detailed study of the basic diffusion equation (0.6) is the key to understanding the system in (0.13). In particular, the invariant measure of (0.6) plays a key role. The following theorem is generally known (it can be proved using the coupling mentioned in (1.23)).

Theorem 0.3 *For every $g \in \mathcal{H}_{Lip}$, $\theta \in [0,1]$ and $c \in (0,\infty)$, the SDE in (0.6) has a unique equilibrium $\nu_\theta^{g,c}$ and is ergodic, i.e., for any $x \in [0,1]$ the law of X_t given $X_0 = x$ converges weakly to $\nu_\theta^{g,c}$ as $t \to \infty$. The measure $\nu_\theta^{g,c}$ is given by*

$$\begin{array}{ll} \nu_\theta^{g,c}(dx) = \frac{1}{Z_\theta^{g,c}} \frac{1}{g(x)} \exp\left(\int_\theta^x \frac{y-\theta}{g(y)} dy\right) dx & (\theta \in (0,1)) \\ \nu_\theta^{g,c}(dx) = \delta_\theta(dx) & (\theta \in \{0,1\}), \end{array} \tag{0.20}$$

where $Z_\theta^{g,c}$ is a normalization constant depending on g, c and θ.

For $\theta \in (0,1)$, the density of $\nu_\theta^{g,c}$ solves the equation $(c(x-\theta) + \frac{\partial}{\partial x} g(x))\nu_\theta^{g,c}(x) = 0$ (compare (1.25) and (2.17) (ii)).

0.4 The renormalization transformation

The reasoning above indicates that, for large N, the single components $X_\xi^N(t)$ perform a diffusion as in (0.6), with as a stochastic attraction point the 1-block average $X_\xi^{N,1}(t)$. Since the single components reach equilibrium on time scale t (i.e., fast compared to time scale Nt of the block), we expect that at times of order Nt their conditional distribution given the 1-block average is given by

$$P[X_\xi^N(Nt) \in dy \mid X_\xi^{N,1}(Nt) = x] \cong \nu_x^{g,c_1}(dy). \tag{0.21}$$

Now again consider the heuristic formula (0.19). Formula (0.21) suggests that

$$N^{-1} \sum_{\zeta:\, d(\zeta,\xi)\leq 1} g(X_\zeta^N(Nt)) \cong \int_{[0,1]} g(y)\nu^{g,c_1}_{X_\xi^{N,1}(Nt)}(dy). \tag{0.22}$$

This motivates the following definition of our renormalization transformation: for every $g \in \mathcal{H}_{Lip}$, $c \in (0,\infty)$

$$(F_c g)(x) := \int_{[0,1]} g(z)\nu_x^{g,c}(dz) \qquad (x \in [0,1]). \tag{0.23}$$

From [6], Lemma 2.2 it follows that:

Theorem 0.4 *For all $c \in (0,\infty)$: $F_c\mathcal{H}_{Lip} \subset \mathcal{H}_{Lip}$.*

Theorem 0.4 makes it possible to speak about the iterates of F_c, which we shall need below.

0.5 Multiple space-time scale analysis

Combining (0.19) with (0.22) and (0.23), and neglecting higher order terms in N, we find the following conditional expectations for $X_\xi^{N,1}(t)$:

$$\begin{aligned} E[\Delta X_\xi^{N,1}(t) \mid \mathcal{F}_t] &\cong N^{-1}c_2\Big[X_\xi^{N,2}(t) - X_\xi^{N,1}(t)\Big]\Delta t \\ E[\Delta X_\xi^{N,1}\Delta X_\eta^{N,1} \mid \mathcal{F}_t] &\cong N^{-1}1_{\{d(\xi,\eta)\le 1\}}2(F_{c_1}g)(X_\xi^{N,1}(t))\Delta t. \end{aligned} \tag{0.24}$$

Note that $1_{\{d(\xi,\eta)\le 1\}} = 1$ if and only if the 1-block around ξ is the 1-block around η. The conditional expectations above seem to indicate that 1-block averages, when viewed on time scale Nt, behave as diffusions like the single components, *but with the local diffusion rate g replaced by $F_{c_1}g$.* This is precisely what is proved in [7]. In fact, the reasoning can be extended to arbitrary k-blocks. The local diffusion rate is then $(F_{c_k} \circ \cdots \circ F_{c_1})g$. The time scale for the k-blocks turns out to be $N^k t$. Indeed, we must rescale space and time together: each time we go up one step in the hierarchy we have larger blocks moving on a slower time scale.

To be precise, the heuristic formula (0.21) is justified for general k by the following theorem ([7], Theorem 1). Here, for each N, we take $0 = (0,0,\ldots) \in \Omega_N$ as a typical reference point, and we denote weak convergence by $\Longrightarrow$.

Theorem 0.5 *Fix $g \in \mathcal{H}_{Lip}$, $\theta \in [0,1]$, $t > 0$ and $k \ge 0$. Then as $N \to \infty$*

$$(X_0^{N,k}(N^k t), \ldots, X_0^{N,0}(N^k t)) \Longrightarrow (Z_k, \ldots, Z_0), \tag{0.25}$$

where $(Z_k,\ldots,Z_0)$ is a 'backward' time-inhomogeneous Markov chain with transition kernels

$$P[Z_{l-1} \in dy | Z_l = x] = \nu_x^{F^{(l-1)}g,c_l}(dy) \qquad (l = k,\ldots,1), \tag{0.26}$$

and $F^{(k)}g := (F_{c_k}\circ\cdots\circ F_{c_1})g$ is the k-th iterate of the renormalization transformations F_c applied to g ($F^{(0)}g = g$).

The joint distribution of the $(Z_k, \ldots, Z_0)$ above is determined by the 'backward' transition probabilities in (0.26) and the distribution of Z_k. The latter depends on t and can be read off from the next theorem ([7], Theorem 1). Here the $\Longrightarrow$ denotes weak convergence in path space $\mathcal{C}[0, \infty)$.

Theorem 0.6 *Fix $g \in \mathcal{H}_{Lip}$, $\theta \in [0,1]$ and $k \geq 0$. Then as $N \to \infty$*

$$\left(X_0^{N,k}(N^k t)\right)_{t\geq 0} \Longrightarrow \left(Z_\theta^{F^{(k)}g, c_{k+1}}(t)\right)_{t\geq 0}, \tag{0.27}$$

where $(Z_\theta^{g,c}(t))_{t\geq 0}$ is the unique strong solution of the single component SDE on $[0,1]$ given by

$$\begin{aligned} dZ(t) &= c(\theta - Z(t))dt + \sqrt{2g(Z(t))}dB(t) \\ Z(0) &= \theta. \end{aligned} \tag{0.28}$$

For $k = 0$ this result justifies our heuristic belief that the single components follow the basic diffusion equation (0.6), and for $k = 1$ it justifies our formula (0.24). For general $k \geq 1$ it describes the behavior of the k-block averages.

As a side remark, we note that the initial condition $X_\xi^N(0) = \theta$ in (0.13) can be generalized considerably. In [5], section 2, and [7], Remark below equation (1.5), $\{X_\xi^N(0)\}_{\xi \in \Omega_N}$ is taken to be distributed according to a homogeneous ergodic measure μ with $E^\mu(X_\xi^N(0)) = \theta$ for all $\xi \in \Omega_N$. For instance, one can take the $X_\xi^N(0)$ to be i.i.d. with mean θ. In this case, Theorem 0.6 changes, in the sense that the distribution of $Z_\theta^{g,c_1}(0)$ is given by μ rather than δ_θ. The distribution of $Z_\theta^{F^{(k)}g,c_{k+1}}(0)$ for $k \geq 1$ is, however, still δ_θ. In view of this, the model where each component starts in θ is the most natural one.

0.6 Large space-time behavior and universality

Theorems 0.5 and 0.6 describe the behavior of our system in the limit as $N \to \infty$. We next study the system by taking one more limit, namely, we consider k-blocks with $k \to \infty$. This gives rise to two more theorems: Theorem 0.7 describes the behavior of the Markov chain in Theorem 0.5 for large k, while Theorem 0.9 describes the behavior of the renormalized diffusion function in Theorem 0.6 for large k. The translation of these theorems in terms of the infinite system is described in Theorems 0.8 and 0.10.

As a joint function of θ and dx, the equilibrium $\nu_\theta^{g,c}(dx)$ in (0.20) is a continuous probability kernel on $[0,1]$. Let $\mathcal{P}[0,1]$ denote the probability measures on $[0,1]$, equipped with the topology of weak convergence, and let $\mathcal{K}[0,1]$ denote the space of all continuous kernels $K : [0,1] \to \mathcal{P}[0,1]$, equipped with the topology of uniform convergence (see also section 1.3). A kernel K evaluated in a point x is denoted by K_x. Uniform convergence of probablitity kernels implies pointwise convergence, so $K^n \to K$ in the topology on $\mathcal{K}[0,1]$ implies $K_x^n \Rightarrow K_x$ for all $x \in [0,1]$.

We denote the composition of two probability kernels $K_x(dy)$ and $L_x(dy)$ by

$$(KL)_x(dz) := \int_{[0,1]} K_x(dy)L_y(dz). \tag{0.29}$$

By Theorem 0.5, in the limit as $N \to \infty$, the conditional probability of $X_\xi^N(N^k t) \in dy$ given $X_\xi^{N,k}(N^k t) = x$ is given by the kernel

$$K_x^{g,(k)}(dy) := (\nu^{F^{(k-1)}g,c_k} \cdots \nu^{g,c_1})_x(dy), \tag{0.30}$$

with the composition as in (0.29). The following can be found in [1], equation (1.7):

Theorem 0.7 *Fix $g \in \mathcal{H}_{Lip}$. As $k \to \infty$, then in the sense of uniform convergence of probability kernels:*

$$K^{g,(k)} \to K^{(\infty)}, \tag{0.31}$$

where the limiting kernel $K^{(\infty)}$ is universal in g and given by

$$K_\theta^{(\infty)} = (1-\theta)\delta_0 + \theta\delta_1 \qquad (\theta \in [0,1]). \tag{0.32}$$

Note that, for any $k \geq l$, the conditional probability of $X_\xi^{N,l}(N^k t) \in dy$ given $X_\xi^{N,k}(N^k t) = x$ is described by the kernel $\nu^{F^{(k-1)}g,c_k} \cdots \nu^{F^{(l)}g,c_{l+1}}$, which is just the kernel in (0.30) with g replaced by $F^{(l)}g$ and $(c_k)_{k\geq 1}$ replaced by $(c_k)_{k\geq l+1}$. Using Theorem 0.5 and the fact that, with $Z(t)$ as in (0.28), we have $E[Z(t)] = \theta \ \forall t \geq 0$, Theorem 0.7 translates into the following statement about the infinite system:

Theorem 0.8 *Fix $g \in \mathcal{H}_{Lip}$, $\theta \in [0,1]$, $l \geq 0$ and $t > 0$. Then, in the sense of convergence in law:*

$$\lim_{k\to\infty} \lim_{N\to\infty} X_0^{N,l}(N^k t) = Y, \tag{0.33}$$

where the law of Y is given by $\mathcal{L}(Y) = (1-\theta)\delta_0 + \theta\delta_1$.

Thus, the system locally ends up in one of the traps 0 or 1. This behavior is called *clustering* and should be interpreted as saying that, for large N and k, the block averages spend most of their time close to the boundaries of the state space $[0,1]$. Condition (0.10) in fact characterizes the clustering regime for the system in the $N \to \infty$ limit. For finite N, clustering of the system can be related to the recurrence of the random walk with kernel $a_N(\xi,\eta)$ given in (0.17) (see [3]). For a discussion of clustering in the case $g(x) = rx(1-x)$, both for $N \to \infty$ and for finite N, see [7], Theorems 3 and 6.

We next turn to the behavior of $F^{(k)}g$ as $k \to \infty$. Note that since $\nu_\theta^{g,c}$ itself depends on g, the transformation F_c is a non-linear integral transform. As such it is a rather difficult object to study in detail. Nevertheless, [1] gives a complete description of the asymptotic behavior of its iterates. The results show that there is a unique 'fixed shape' $g^* \in \mathcal{H}_{Lip}$ that attracts all orbits after appropriate scaling, as follows:

Theorem 0.9
(a) *Let $g^*(x) = x(1-x)$. The 1-parameter family of functions $g = rg^*$ $(r > 0)$ are fixed shapes under F_c:*

$$F_c(rg^*) = \left(\frac{c}{c+r}\right) rg^*. \tag{0.34}$$

(b) *For all $g \in \mathcal{H}_{Lip}$*

$$\lim_{k\to\infty} \sigma_k F^{(k)} g = g^* \quad \textit{uniformly on } [0,1], \tag{0.35}$$

where $\sigma_k := \sum_{l=1}^{k} c_l^{-1}$.
(c) *Let*

$$\mathcal{H}_1 := \{g \in \mathcal{H}_{Lip} : \liminf_{x\to 0} x^{-2} g(x) > 0 \textit{ and } \liminf_{x\to 1} (1-x)^{-2} g(x) > 0\}. \tag{0.36}$$

Then for all $g \in \mathcal{H}_1$

$$\lim_{k\to\infty} \|\sigma_k F^{(k)} g - g^*\|_{\mathcal{H}_{Lip}} = 0, \tag{0.37}$$

where

$$\|g\|_{\mathcal{H}_{Lip}} := \sup_{x\in(0,1)} \left|\frac{g(x)}{g^*(x)}\right|. \tag{0.38}$$

To be able to state the implications of Theorem 0.9 for the infinite system, we must rescale the time once more, now to compensate not for the large N but for the large k. Indeed, by an easy scaling property of the $Z_\theta^{g,c}$ defined in (0.28), we can rewrite Theorem 0.6 as

$$\left(X_0^{N,k}(\sigma_k N^k t)\right)_{t\geq 0} \Longrightarrow \left(Z_\theta^{\sigma_k F^{(k)} g, \sigma_k c_k}(t)\right)_{t\geq 0} \qquad \text{as } N\to\infty. \tag{0.39}$$

In view of (0.35), the most interesting behavior now occurs when $\sigma_k c_k$ tends to some limit as $k\to\infty$. From Theorem 0.9 (b) we get, by a simple application of [20], Theorem 11.1.4, the following:

Theorem 0.10 *If $\lim_{k\to\infty} \sigma_k c_k = c^* \in [0,\infty)$, then in the sense of weak convergence of the law in path space $\mathcal{C}[0,\infty)$:*

$$\lim_{k\to\infty} \lim_{N\to\infty} \left(X_0^{N,k}(\sigma_k N^k t)\right)_{t\geq 0} = \left(Z_\theta^{g^*,c^*}(t)\right)_{t\geq 0}. \tag{0.40}$$

For example, if $c_k = ab^k$ with $a \in (0,\infty)$ and $b \in (0,1)$, then $\lim_{k\to\infty} \sigma_k c_k = a^2 \frac{b}{1-b}$.

The results in Theorems 0.9 and 0.10 show that our system displays *complete universality* on large space-time scales. For large k (and in the limit as $N\to\infty$) the k-blocks approximately perform the diffusion in (0.28) with diffusion function g^* and with attraction constant c^*, *and this behavior is completely universal in the diffusion function g of the single components.*

Theorem 0.9 (c) is important for the study of how clustering occurs. In fact, under (0.37) *the clustering turns out to be universal in g* (see [7], Corollary at Theorem 5). It turns out that the class $\mathcal{H}_1$ in (0.36) is sharp: if $\limsup_{x\to 0} x^{-2}g(x) = 0$ or $\limsup_{x\to 0}(1-x)^{-2}g(x) = 0$, then $\sigma_k F^{(k)}g$ does not converge in the norm $\|\cdot\|_{\mathcal{H}_{Lip}}$ (see [1]).

1 Results for $d \geq 1$

In this section we present our best results towards extending the model in section 0 to higher dimension. In sections 1.1 and 1.2 we formulate a general program, and specify the particular model that is the subject of the present paper. In section 1.3 we present our theorems on the renormalization transformations F_c ($c \in (0,\infty)$) arising in that model. The theorems are stated in terms of certain classes of functions $\mathcal{H}'$ and $\mathcal{H}''$. These are essentially the largest domains on which we can define our renormalization transformations F_c, resp. the iterates $F^{(k)}$. For the results to make sense, it remains to be shown that these classes are not empty. This task is, with limited success, taken up in section 1.4. In section 1.5 we indicate some of the difficulties that make life hard in $d \geq 2$. Finally, in section 1.6, some of the more urgent open problems are discussed.

Proofs are given in sections 2–4.

1.1 Generalizations to different state spaces

The renormalization techniques described in the last section are not restricted to models with state space $[0,1]$. The construction of more general models could be described in the form of the following program:

1. Choose an open convex domain $D \subset \mathbb{R}^d$ and a class $\mathcal{H}$ of diffusion matrices on $\overline{D}$ (i.e., the equivalents of $[0,1]$ and $\mathcal{H}_{Lip}$ in section 0). Prove (as in Theorem 0.1) that for all $g \in \mathcal{H}$, $\theta \in \overline{D}$, $c \in (0,\infty)$ the martingale problem is well-posed for the differential operator

$$(Af)(x) := \Big(\sum_{i=1}^{d} c(\theta_i - x_i)\tfrac{\partial}{\partial x_i} + \sum_{i,j=1}^{d} g_{ij}(x)\tfrac{\partial^2}{\partial x_i \partial x_j}\Big) f(x). \tag{1.1}$$

2. Prove (as in Theorem 0.3) ergodicity for the diffusion given by A, and define a renormalization transformation F_c by

$$(F_c g)_{ij}(\theta) = \int_{\overline{D}} \nu_\theta^{g,c}(dx) g_{ij}(x), \tag{1.2}$$

where $\nu_\theta^{g,c}$ is the equilibrium associated with (1.1). Show that $\mathcal{H}$ is closed under F_c (as in Theorem 0.4), and show that the iterates $F^{(k)}g$ and the associated kernel $K^{g,(k)}$ describe the multiple space-time scale behavior of the associated infinite system (i.e., prove analogues of Theorems 0.2, 0.5 and 0.6).

3. Investigate the limiting behavior of $K^{g,(k)}_\theta$ and $F^{(k)}g$ as $k \to \infty$ (i.e., try to prove equivalents of Theorems 0.7–0.10).

So far, such a program has only been carried out completely for one-dimensional state spaces, as explained in section 0. For the program to get off the ground, one must at least be able to speak about the iterates $F^{(k)}g$. In practice, this leads to conflicting demands on the class $\mathcal{H}$. When $\mathcal{H}$ is chosen large, it turns out to be difficult to show uniqueness for the martingale problem for A in (1.1), and therefore the program already stops at step 1. On the other hand, when $\mathcal{H}$ is chosen too restrictive, it turns out to be hard to show (in step 2) that $F_c g \in \mathcal{H}$, i.e., we can define $F_c g$ but not its iterates $F^{(k)}g$. At present, these difficulties present a serious obstacle in trying to carry out the program above completely for state spaces in dimensions $d \geq 2$.

In the present paper, we focus on the construction of $F^{(k)}g$ and $K^{g,(k)}$ and the study of their limiting behavior for a certain class $\mathcal{H}$ of 'isotropic' diffusion matrices in dimensions $d \geq 2$. We leave the study of the associated infinite system to be treated in future work. The difficulties mentioned above are dealt with in the following way. We introduce subclasses $\mathcal{H}'$ and $\mathcal{H}''$ that are essentially the largest subsets of $\mathcal{H}$ on which $F_c g$ resp. $F^{(k)}g$ can be defined. (It may be that $\mathcal{H} = \mathcal{H}' = \mathcal{H}''$, but this can at present not be proved.) In section 1.3, we show that on these classes it is possible to carry out step 3 of the above program completely. In particular, we show that there exists a unique fixed shape g^* under F_c that attracts all g under appropriate scaling, and also, that there exists a universal limiting kernel to which all $K^{g,(k)}$ converge. In section 1.4, we investigate under what conditions $F_c g$ and $F^{(k)}g$ can be defined, i.e., we find conditions for $g \in \mathcal{H}'$ and $g \in \mathcal{H}''$. The results in this section are not as conclusive as those in section 1.3, but we can show that many functions are in $\mathcal{H}'$, and at least in one example we can show that $\mathcal{H}''$ is not empty.

1.2 Isotropic models

We consider as state space the closure $\overline{D}$ of an arbitrary open, bounded and convex set $D \subset \mathbb{R}^d$. On $\overline{D}$, we consider a class $\mathcal{H}$ of isotropic diffusion matrices. We say that a diffusion matrix $g_{ij}(x)$ is isotropic if it has the form $g_{ij}(x) = \delta_{ij} g(x)$, where $g : \overline{D} \to [0, \infty)$ is some non-negative function, and $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ otherwise. From the form of the renormalization transformation we see that for an isotropic diffusion matrix

$$(F_c g)_{ij}(\theta) = \int_{\overline{D}} \nu^{g,c}_\theta(dx)\delta_{ij} g(x) = \delta_{ij} \int_{\overline{D}} \nu^{g,c}_\theta(dx) g(x), \tag{1.3}$$

so if g is isotropic, then $F_c g$ is isotropic. On the class of isotropic diffusions, F_c is essentially just a transformation of functions $g : \overline{D} \to [0, \infty)$.

In the special case that $\overline{D} = K_d$, the d-dimensional simplex, we indicate briefly how such isotropic models can arise as continuous limits of discrete models. Consider the random sampling procedure described in section 0.1. Suppose that instead of replacing *pairs* we replace $(d+1)$-*tuples*, in the following manner. After an exponential time, a $(d+1)$-tuple of individuals is selected. If all $d+1$ individuals belong to different types, then they are all replaced by one randomly chosen type. Otherwise nothing happens. A little calculation shows that this procedure gives rise to the following diffusion matrix:

$$g_{ij}(x) = [\delta_{ij}(d+1) - 1]\left(1 - \sum_k x_k\right)\prod_k x_k \qquad (i,j,k = 1,\ldots,d). \tag{1.4}$$

By a simple transformation of the state space, the matrix $\delta_{ij}(d+1) - 1$ can be diagonalized to δ_{ij}. In this way one arrives at an isotropic model with $g_{ij}(x) = \delta_{ij} g(x)$, where g is given by (the transformed function of) $x \mapsto (1 - \sum_k x_k)\prod_k x_k$. More general functions g can be obtained by making the rate of the random sampling process dependent on the state of the system.[2,3]

1.3 Renormalization in $d \geq 1$: Theorems 1.1–1.4

We introduce the following objects:

1. ('state space') $D \subset \mathbb{R}^d$ is a bounded open convex set, $\overline{D}$ is its closure and $\partial D = \overline{D} \backslash D$.

2. ('fixed shape') $g^* : \overline{D} \to \mathbb{R}$ is the unique continuous solution of

$$\begin{aligned} -\tfrac{1}{2}\Delta g^* &= 1 \qquad \text{on } D \\ g^* &= 0 \qquad \text{on } \partial D, \end{aligned} \tag{1.5}$$

   with $\Delta = \sum_{i=1}^d \frac{\partial^2}{\partial {x_i}^2}$ the Laplacian.

3. ('diffusion function') $\mathcal{H}$ is the class of functions $g : \overline{D} \to [0,\infty)$ satisfying

$$\begin{aligned} &\text{(i)} \quad g \leq M g^* \text{ for some } M < \infty \\ &\text{(ii)} \quad g > 0 \text{ on } D \\ &\text{(iii)} \quad g \text{ continuous on } \overline{D}. \end{aligned} \tag{1.6}$$

[2] In section 0.1, the Wright-Fisher diffusion was introduced on K_d. In dimensions $d \geq 2$, this diffusion is non-isotropic. It is not hard to see that it is a fixed shape under F_c. Therefore it is expected that in $d \geq 2$, and on a larger class than only the isotropic diffusions treated in the present paper, the transformation F_c has many different fixed shapes, each with their own domain of attraction.

[3] In $d = 1$, the fixed shape g^* on the simplex appears to be the most natural object when seen as the continuous limit of a discrete model. Comparing the fixed shape g^* that we find in our analysis below with the diffusion matrix in (1.4), it turns out that the formulas coincide in $d = 1, 2$ but, remarkably, not in higher dimensions.

4. ('attraction point') $\theta \in \overline{D}$.

5. ('attraction constant') $c \in (0, \infty)$.

With these ingredients we let our basic diffusion equation be the SDE:

$$dX_t = c(\theta - X_t)dt + \sqrt{2g(X_t)}dB_t, \tag{1.7}$$

where $(B_t)_{t\geq 0}$ is standard d-dimensional Brownian motion. Solutions of (1.7) solve the martingale problem for the operator A with domain $\mathcal{D}(A)$ given by

$$\begin{aligned} (Af)(x) &:= \Big(c(\theta - x) \cdot \nabla + g(x)\Delta\Big) f(x) \\ \mathcal{D}(A) &:= \mathcal{C}^2(\overline{D}), \end{aligned} \tag{1.8}$$

where $\nabla = (\frac{\partial}{\partial x_i}, \ldots, \frac{\partial}{\partial x_d})$ and $\cdot$ denotes inner product. The martingale problem for A is well-posed if and only if, for each initial condition on $\overline{D}$, the SDE (1.7) has a unique weak solution $(X_t)_{t\geq 0}$. In this case, the operator A has a unique extension to a generator of a Feller semigroup, and $(X_t)_{t\geq 0}$ is the associated Feller process.[4]

By a continuous probability kernel on $\overline{D}$ we mean a continuous map $K : \overline{D} \to \mathcal{P}(\overline{D})$, written $x \mapsto K_x$, where $\mathcal{P}(\overline{D})$ is the space of probability measures on $\overline{D}$, equipped with the topology of weak convergence. We equip the space $\mathcal{K}(\overline{D}) := \mathcal{C}(\overline{D}, \mathcal{P}(\overline{D}))$ of probability kernels on $\overline{D}$ with the topology of uniform convergence. (Since $\mathcal{P}(\overline{D})$ is compact and Hausdorff, there is a unique uniform structure defining the topology, and we can unambiguously speak about uniform convergence of $\mathcal{P}(\overline{D})$-valued functions.) There exists a natural identification between continuous probability kernels $K \in \mathcal{K}(\overline{D})$ and continuous positive linear operators $K : \mathcal{C}(\overline{D}) \to \mathcal{C}(\overline{D})$ satisfying $K1 = 1$, the correspondence being given by

$$(Kf)(x) = \int_{\overline{D}} K_x(dy) f(y) \qquad\qquad (f \in \mathcal{C}(\overline{D})). \tag{1.9}$$

In this identification, the composition of two kernels is given by

$$(KL)_x(dy) = \int_{\overline{D}} K_x(dy) L_y(dz). \tag{1.10}$$

The convergence of operators $K_n \to K$ in the topology on $\mathcal{K}(\overline{D})$ is equivalent to the convergence of the functions $K_n f \to Kf$, uniformly on $\overline{D}$ for all $f \in \mathcal{C}(\overline{D})$.

In order to be able to define our renormalization transformation, we introduce a new class $\mathcal{H}'$ of diffusion functions as follows:

3′. $\mathcal{H}'$ is the class of all functions $g \in \mathcal{H}$ such that for all $c \in (0, \infty)$ and $\theta \in \overline{D}$:
(1) The martingale problem associated with the operator A in (1.8) is well-posed.
(2) The diffusion associated with (1.7) has a unique equilibrium $\nu_\theta^{g,c}$.

[4]For a discussion of these facts, see the footnote at Theorem 0.1.

Here, by an equilibrium we mean a stationary distribution of (1.7). As we shall see in section 1.4, these assumptions are satisfied for many $g \in \mathcal{H}$. It turns out that the map $\theta \mapsto \nu_\theta^{g,c}$ is continuous, and so the equilibrium of (1.7) is a continuous probability kernel on $\overline{D}$ as a function of the parameter θ.

Theorem 1.1 *For each $g \in \mathcal{H}'$ and $c \in (0, \infty)$ there exists a continuous probability kernel $\nu^{g,c} \in \mathcal{K}(\overline{D})$ such that, for each $\theta \in \overline{D}$, $\nu_\theta^{g,c}$ is the equilibrium of the diffusion in (1.7).*

For $g \in \mathcal{H}'$ and $c \in (0, \infty)$, we now define our 'renormalization transformation' as

$$(Fg)(\theta) := (\nu^{g,c} g)(\theta) = \int_{\overline{D}} g(x) \nu_\theta^{g,c}(dx). \tag{1.11}$$

In order to speak about the iterates of F_c, we need a subclass of $\mathcal{H}'$ that is closed under the F_c's. For this we may take the largest such subclass, so we define one more class of diffusion functions:

3″. $\mathcal{H}''$ is the union of all $\mathcal{G} \subset \mathcal{H}'$ such that $F_c(\mathcal{G}) \subset \mathcal{G}$ for all $c \in (0, \infty)$.

With these definitions, we have the following result.

Theorem 1.2 *For all $c \in (0, \infty)$: $F_c(\mathcal{H}') \subset \mathcal{H}$.*

It is at present not known if $\mathcal{H} = \mathcal{H}'$, but Theorem 1.2 implies at least that if (!) $\mathcal{H}' = \mathcal{H}$, then $\mathcal{H}'' = \mathcal{H}$.

The next result generalizes Theorem 0.7 (recall the composition of probability kernels defined in (1.10)):

Theorem 1.3 *For $g \in \mathcal{H}''$ and $k \geq 1$, let $K^{g,(k)}$ be given by*

$$K^{g,(k)} := \nu^{F^{(k-1)} g, c_k} \cdots \nu^{g, c_1}, \tag{1.12}$$

where $F^{(k)} g := (F_{c_k} \circ \cdots \circ F_{c_1}) g$ is the k-th iterate of the renormalization transformations F_c applied to g ($F^{(0)} g = g$). If $\sum_k c_k^{-1} = \infty$, then in the sense of uniform convergence of probability kernels:

$$K^{g,(k)} \to K^{(\infty)} \qquad \text{as } k \to \infty, \tag{1.13}$$

where the limiting kernel $K^{(\infty)}$ is universal in g and given by

$$K_\theta^{(\infty)}(dx) = P[B_\tau^\theta \in dx], \tag{1.14}$$

where $(B_t^\theta)_{t \geq 0}$ is Brownian motion starting in θ and $\tau := \inf\{t \geq 0 : B_t^\theta \in \partial D\}$.

The following generalizes Theorem 0.9:

Theorem 1.4
(a) *Let g^* be as in (1.5). If $g^* \in \mathcal{H}'$, then $rg^* \in \mathcal{H}''$ for all $r > 0$. Moreover, the 1-parameter family of functions rg^* $(r > 0)$ are fixed shapes under F_c:*

$$F_c(rg^*) = \left(\frac{c}{c+r}\right) rg^*. \tag{1.15}$$

(b) *If $\sum_k c_k^{-1} = \infty$, then for all $g \in \mathcal{H}''$*

$$\lim_{k\to\infty} \sigma_k F^{(k)} g = g^* \quad \textit{uniformly on } \overline{D}, \tag{1.16}$$

where $\sigma_k := \sum_{l=1}^{k} c_l^{-1}$.
(c) *If, in addition to the assumptions in (b), there exists a $\lambda > 0$ such that $g \geq \lambda g^*$, then*

$$\lim_{k\to\infty} \|\sigma_k F^{(k)} g - g^*\|_{\mathcal{H}} = 0, \tag{1.17}$$

where the norm $\|\cdot\|_{\mathcal{H}}$ is given by

$$\|g\|_{\mathcal{H}} := \sup_{x\in D} \left|\frac{g(x)}{g^*(x)}\right|. \tag{1.18}$$

In $d = 1$, formula (1.17) is in fact known to hold under somewhat weaker conditions on g (see Theorem 0.9 (c)).

1.4 Two renormalization classes: Theorems 1.5–1.10

The results in the last section are useful only after we come up with some examples of functions g in the classes $\mathcal{H}'$ and $\mathcal{H}''$. In this section we try to find sufficient conditions for $g \in \mathcal{H}'$ and for $g \in \mathcal{H}''$.

The following theorem shows that the assumption about ergodicity in the definition of $\mathcal{H}'$ is in most 'neat' cases satisfied.

Theorem 1.5 *Fix $g \in \mathcal{H}$, $\theta \in \overline{D}$ and $c \in (0, \infty)$. Assume that g is locally Hölder continuous (with positive exponent) on D, and that the martingale problem associated with the operator A in (1.8) is well-posed. Then the SDE (1.7) has a unique equilibrium $\nu_\theta^{g,c}$ and is ergodic, i.e., for each initial distribution the law of X_t converges weakly to $\nu_\theta^{g,c}$ as $t \to \infty$.*

Thus, for locally Hölder g, proving that $g \in \mathcal{H}'$ reduces to proving that the martingale problem for A in (1.8) is well-posed. As usual, existence of solutions is no problem:

Theorem 1.6 *For each $g \in \mathcal{H}$, $\theta \in \overline{D}$ and $c \in [0,\infty)$, and for each initial distribution on $\overline{D}$, the SDE (1.7) has a $\overline{D}$-valued, continuous, weak solution $(X_t)_{t\geq 0}$. The law of $(X_t)_{t\geq 0}$ solves the martingale problem for the operator A in (1.8).*

In fact, it seems reasonable to conjecture that uniqueness of the martingale problem, too, holds for all $g \in \mathcal{H}$, and (assuming ergodicity can also be proved), that $\mathcal{H}' = \mathcal{H}$. As we saw in Theorem 1.2, this would imply $\mathcal{H} = \mathcal{H}' = \mathcal{H}''$. However, it is not known whether uniqueness for the martingale problem holds for general $g \in \mathcal{H}$.[5]

In $d = 1$, uniqueness can be proved for many $g \in \mathcal{H}$, and the explicit formula for the equilibrium $\nu_\theta^{g,c}$ in (0.20) can be used to prove that, for all g, $F_c g$ is sufficiently nice. Indeed, the Yamada-Watanabe argument ([14], Proposition 5.2.13) and [1], Remark below Theorem 5, show that:

Theorem 1.7 *Assume that $d = 1$. If $g \in \mathcal{H}$ and $\sqrt{g}$ is Hölder $\frac{1}{2}$-continuous, then $g \in \mathcal{H}'$ and $g \in \mathcal{H}''$.*

In higher dimension, results are much harder to get. The standard theorem for strong uniqueness of multi-dimensional diffusions ([14], Theorem 5.2.9) and Theorem 1.5 give:

Theorem 1.8 *Assume that $d \geq 1$. If $g \in \mathcal{H}$ and $\sqrt{g}$ is Lipschitz, then $g \in \mathcal{H}'$.*

If we restrict ourselves to initial conditions x and attraction points θ that lie within D, then the conditions for strong uniqueness can be weakened. We adopt the following definitions. If $x \in \partial D$, then $n(x) \in \mathbb{R}^d$ is called an (outward) normal to D in x if and only if $|n(x)| = 1$ and $\{y \in \mathbb{R}^d : (y - x) \cdot n(x) \geq 0\} \cap D = \emptyset$. A set $D \subset \mathbb{R}^d$ is called regular if and only if D is open, bounded, convex, and there exists a function $m \in \mathcal{C}^3(\overline{D})$, satisfying $m = 0$ on ∂D and $m < 0$ on D, with the property that for all $x \in \partial D$

$$|\nabla m(x)| = 1. \tag{1.19}$$

Note that, for each $x \in \partial D$, $\nabla m(x)$ is the unique normal to D in x. With these conventions we have the following theorem.

Theorem 1.9 *Let $g \in \mathcal{H}$, $\theta \in D$ and $c \in (0,\infty)$. Assume that D is a finite intersection of regular sets. Let g be Lipschitz on D, and assume that for all $x \in \partial D$, all $x_n \in D$ with $x_n \to x$, and each normal $n(x)$ to D in x:*

$$\limsup_{n\to\infty} \frac{g(x_n)}{|x - x_n|} < c(x - \theta) \cdot n(x). \tag{1.20}$$

Then any solution $(X(t))_{t\geq 0}$ of (1.7) with initial condition $X(0) = x$ ($x \in D$) satisfies

$$P[X(t) \in D \quad \forall t \geq 0] = 1, \tag{1.21}$$

and strong uniqueness holds for the SDE (1.7) with initial condition x.

[5] Even in $d = 1$ this question seems to be open, although for each $g \in \mathcal{H}$ it is known that there exists a unique extension of A to a generator of a Feller semigroup. For this extended operator the martingale problem, of course, is well-posed.

The idea behind this theorem is that, since $\sqrt{g}$ is locally Lipschitz on D, a modification of the standard proof for strong uniqueness shows that solutions of (1.7) are unique up to the first hitting time of the boundary, while condition (1.20) guarantees that this time is infinite. The essential difficulties in proving uniqueness occur when the diffusion hits the boundary in a finite time. Although the conditions on g in Theorem 1.9 are considerably weaker than those in Theorem 1.8, the result is still not very satisfactory for our purposes. Indeed, we want to vary θ, and so if (1.20) is to hold for all $\theta \in D$, then we must have 'sublinear' behavior of g at the boundary:

$$\lim_{n\to\infty} \frac{g(x_n)}{|x - x_n|} = 0 \qquad\qquad (x_n \in D,\ x_n \to x \in \partial D). \tag{1.22}$$

For example, it can be seen that for the fixed shape g^* condition (1.22) is violated. Consequently, Theorem 1.9 cannot even be used to give a satisfactory definition of $(F_c g^*)(\theta)$ for all $\theta \in D$.

Sufficient conditions for $g \in \mathcal{H}''$ are even harder to come by than sufficient conditions for $g \in \mathcal{H}'$. The following special case, however, shows us one example where $F_c g^*$ can be defined in a satisfactory way and where $\mathcal{H}''$ can be shown to be non-empty.

Theorem 1.10 *Let $D = \{x \in \mathbb{R}^d : |x| < 1\}$. Then $g^*(x) = \frac{1}{d}(1 - |x|^2)$ and $g^* \in \mathcal{H}''$.*

This last result is actually the only case in $d \geq 2$ where we are able to prove that $\mathcal{H}''$ is not empty. In view of Theorem 1.4 this is not a very satisfactory result, since we would like $\mathcal{H}''$ to at least contain a neighbourhood of g^* in order for the universality expressed in (1.16) to be meaningful. But nothing better is available at present.

1.5 Difficulties for $d \geq 2$

Higher-dimensional diffusions differ fundamentally from one-dimensional diffusions. In general they are technically much harder to treat. In our situation: let $(X_t^x)_{t\geq 0}$ and $(X_t^y)_{t\geq 0}$ be solutions of (1.7) with initial conditions x resp. y, adapted to the same Brownian motion, and let $g \in \mathcal{H}$ be Lipschitz. In $d = 1$ it can be shown (compare [6], equation (2.47)) that

$$E[|X_t^x - X_t^y|] \leq |x - y| e^{-ct}. \tag{1.23}$$

It is essentially with the help of this coupling that one is able to prove strong uniqueness for solutions of (1.7), convergence to equilibrium, and the property that the class $\mathcal{H}_{Lip}$ in (0.5) is closed under the transformation F_c. In $d \geq 2$, however, (1.23) does not hold. Indeed, let $(S_t)_{t\geq 0}$ be the semigroup associated with the process $(X_t)_{t\geq 0}$, i.e.

$$(S_t f)(x) = E[f(X_t^x)] \qquad (f \in \mathcal{C}(\overline{D})). \tag{1.24}$$

A direct consequence of (1.23) is the following: if f is Lipschitz with constant L, then $S_t f$ is Lipschitz with constant Le^{-ct}. However, in $d \geq 2$ it is possible to show that, for an appropriate g and c, there exist $t > 0$ and Lipschitz f such that the Lipschitz constant of $S_t f$ is strictly larger than the Lipschitz constant of f. Therefore (1.23) cannot hold for these g and c.

Thus, the diffusion (1.7) behaves differently in higher dimension in lacking a good coupling. It also differs in lacking reversibility. By definition, the diffusion in (1.7) is reversible if and only if its equilibrium $\nu_\theta^{g,c}(dx)$ solves the vector equation

$$\int_{\overline{D}} \nu_\theta^{g,c}(dx)[c(\theta - x) + g(x)\nabla]f(x) = 0 \qquad \forall f \in \mathcal{C}^1(\overline{D}). \tag{1.25}$$

Diffusions in $d = 1$ are typically reversible, and we can solve (1.25) explicitly for $\nu_\theta^{g,c}$ to get the formula (0.20). In $d \geq 2$, however, no matter what D, there exists no $g \in \mathcal{H}'$ such that (1.7) is reversible for all (!) $\theta \in \overline{D}$. Related to this is the fact that in general no explicit formula for $\nu_\theta^{g,c}$ is known. Similarly, for general D no explicit formulas are known for the limiting distribution $K_\theta^{(\infty)}$ and for the fixed shape g^*. Since for $d = 1$ the proofs of Theorems 0.7 and 0.9 were based on explicit manipulations with $\nu_\theta^{g,c}$ and g^* (see [1]), the generalization to $d \geq 2$ forces us to use more abstract methods in our proofs. We believe that these methods (in particular the proof of Lemma 2.4 and its use) also give a deeper understanding of the case $d = 1$.

1.6 Open problems

The most urgent open problems concern the question for which functions g it is possible to prove $g \in \mathcal{H}''$ (recall the definitions of $\mathcal{H}'$, $\mathcal{H}''$ and F_c in section 1.3). In particular, one may ask:

1. Is $g^* \in \mathcal{H}'$ for all bounded open convex D?

2. Is $\mathcal{H}' = \mathcal{H}$ for all bounded open convex D?

Since g^* is locally Hölder on D, it is sufficient for question 1. to show that uniqueness holds the martingale problem associated with A in (1.8) (by Theorems 1.5 and 1.6). If the answer to question 1. is affirmative, then at least $g^* \in \mathcal{H}''$ for all D (by Theorem 1.4 (a)). If the answer to question 2. is affirmative, then it implies that $\mathcal{H}'' = \mathcal{H}$, but question 2. certainly represents a hard problem.

In another approach, one may try to show that $\mathcal{H}''$ is not empty by deriving more properties for $F_c g$, given that g is nice. In analogy with the situation in $d = 1$, one may ask:

3. If $g \in \mathcal{H}$ is Lipschitz, then is $F_c g$ also Lipschitz?

4. If $g \in \mathcal{H}$ is Lipschitz, then does it follow that $g \in \mathcal{H}'$?

For question 3. one needs to control the behavior of the equilibrium $\nu_\theta^{g,c}$ as a function of θ. In the absense of an explicit formula, this can be attempted with coupling methods. In fact, the coupling that underlies Theorem 1.8 can be used to show that if $\sqrt{g}$ is Lipschitz with a sufficiently small Lipschitz constant, then $F_c g$ is Lipschitz. However, a better coupling than this one is hard to find in $d \geq 2$, and question 3. is still open. So is question 4., which is a well-known and hard open problem in the field.

2 The renormalization transformation

2.1 Notation

Let $E \subset \mathbb{R}^d$ be open or closed. By $B(E)$ we denote the bounded Borel-measurable real functions on E. For μ a finite measure on E and $f \in B(E)$ we write

$$\langle \mu | f \rangle := \int_E f d\mu. \tag{2.1}$$

The real continuous functions on E are denoted by $\mathcal{C}(E)$, and $\mathcal{C}_b(E)$ is the Banach space of bounded continuous functions with norm $\|f\| := \sup_{x \in E} |f(x)|$. By $\mathcal{C}^n(E)$ we denote the functions $f \in \mathcal{C}(E)$ such that all derivatives up to order n exist on the interior of E and can be extended to functions in $\mathcal{C}(E)$. By definition $\mathcal{C}^\infty(E) := \bigcap_n \mathcal{C}^n(E)$. We sometimes write $f \in \mathcal{C}^n(E)$ when we mean $f|_E \in \mathcal{C}^n(E)$, where $f|_E$ is the restriction of f to E. By $\mathcal{C}^n_c(E)$, $\mathcal{C}^\infty_c(E)$ we denote functions in $\mathcal{C}^n(E)$, $\mathcal{C}^\infty(E)$ that have a compact support in E.

When $X = (X_t)_{t \geq 0}$ is a continuous E-valued stochastic process and $\mathcal{F}^X_t := \sigma(X_s : s \in [0,t])$ is the filtration generated by X, we say that X solves the martingale problem for a linear operator A on $B(E)$ if and only if

$$f(X_t) - \int_0^t (Af)(X_s) ds \tag{2.2}$$

is an $\mathcal{F}^X_t$-martingale for all $f \in \mathcal{D}(A)$, the domain of A. We identify a linear operator A with domain $\mathcal{D}(A)$ with the linear space $\{(f, Af) : f \in \mathcal{D}(A)\}$. Closure always refers to the norm $\|f\|$. We say that a Feller semigroup $(S_t)_{t \geq 0}$ on $B(E)$ is related to X if and only if for all $f \in B(E)$ and $s, t \geq 0$

$$E[f(X_{t+s}) | \mathcal{F}^X_t] = (S_t f)(X_s) \qquad a.s. \tag{2.3}$$

Finally, the notation $\mathcal{A}$ or $\mathcal{A}^{g,c}_\theta$ is used generally (without specification of the domain) for the differential form

$$(\mathcal{A}f)(x) := \Big(c(\theta - x) \cdot \nabla + g(x) \Delta \Big) f(x), \tag{2.4}$$

where $\nabla = (\frac{\partial}{\partial x_i}, \ldots, \frac{\partial}{\partial x_d})$, the $\cdot$ denotes inner product, and $\Delta = \nabla \cdot \nabla = \sum_{i=1}^d \frac{\partial^2}{\partial x_i{}^2}$ is the Laplacian. We write $|x| = \sqrt{x \cdot x}$ for the Euclidian norm.

2.2 Preliminaries

We begin with two lemmas collecting well-known facts. In section 1.3 we already mentioned the following.

Lemma 2.1 *Let $\mathcal{K}(\overline{D})$ be the set of continuous probability kernels on $\overline{D}$, equipped with the topology of uniform convergence, and let $\mathcal{K}'(\overline{D})$ be the space of all positive linear operators $K : \mathcal{C}(\overline{D}) \to \mathcal{C}(\overline{D})$ satisfying $K1 = 1$, equipped with the strong operator topology. Then a homeomorfism between $\mathcal{K}(\overline{D})$ and $\mathcal{K}'(\overline{D})$ is given by*

$$(Kf)(x) = \langle K_x | f \rangle \qquad (f \in \mathcal{C}(\overline{D})). \tag{2.5}$$

In particular, K_n converges to K in the topology on $\mathcal{K}(\overline{D})$ if and only if

$$\|K_n f - Kf\| \to 0 \qquad \forall f \in \mathcal{C}(\overline{D}). \tag{2.6}$$

Proof of Lemma 2.1: Let $K \in \mathcal{K}(\overline{D})$, and for $f \in \mathcal{C}(\overline{D})$ define $(K'f)(x) := \langle K_x | f \rangle$. By the continuity of K, the function $x \mapsto \langle K_x | f \rangle$ is continuous. It is obvious that the map $f \mapsto K'f$ is positive and linear, and therefore continuous, and satisfies $K'1 = 1$. Conversely, by the Riesz-Markov theorem ([16], Theorem IV.14), each such K' defines a probability measure K_x for each $x \in \overline{D}$. Since $K'f \in \mathcal{C}(\overline{D})$ for each $f \in \mathcal{C}(\overline{D})$, the map $x \mapsto K_x$ is continuous in the weak topology. Finally, $\mathcal{P}(\overline{D})$ is continuously imbedded in $\mathcal{C}(\overline{D})^*$, the dual of $\mathcal{C}(\overline{D})$, so $\mathcal{K}(\overline{D})$ is continuously imbedded in $\mathcal{C}(\overline{D}, \mathcal{C}(\overline{D})^*)$, if we equip the latter with the topology of uniform convergence, where the uniform structure on $\mathcal{C}(\overline{D})^*$ is given by the semi-norms $l \mapsto |\langle l | f \rangle|$. The topology of uniform convergence in $\mathcal{C}(\overline{D}, \mathcal{C}(\overline{D})^*)$ is then defined by the semi-norms $(p_f)_{f \in \mathcal{C}(\overline{D})}$ given by

$$p_f(F) = \sup_{x \in \overline{D}} |\langle F(x) | f \rangle| \qquad (F \in \mathcal{C}(\overline{D}, \mathcal{C}(\overline{D})^*)). \tag{2.7}$$

If $K \in \mathcal{K}(\overline{D})$, then $p_f(K) = \sup_{x \in \overline{D}} |\langle K_x | f \rangle| = \|Kf\|$, so uniform convergence of probability kernels corresponds to convergence of the associated operators in the strong operator topology. From now on we identify kernels with linear operators as in (2.5).

Since $\overline{D}$ is convex, the Dirichlet problem on $\overline{D}$ always has a solution. We shall be interested in harmonic functions and functions of constant Laplacian.

Lemma 2.2 **(a)** *For every $\phi \in \mathcal{C}(\partial D)$ there exists a unique $f \in \mathcal{C}(\overline{D}) \cap \mathcal{C}^2(D)$ that solves*

$$\begin{aligned} \Delta f &= 0 && \text{on } D \\ f &= \phi && \text{on } \partial D. \end{aligned} \tag{2.8}$$

The solution is given by

$$f = H\phi, \tag{2.9}$$

where $H \in \mathcal{K}(\overline{D})$ is the probability kernel given by

$$H_x(dy) = P[B^x_\tau) \in dy], \tag{2.10}$$

where $(B^x_t)_{t\geq 0}$ *is Brownian motion starting at* x *and*

$$\tau := \inf\{t \geq 0 : B^x_t \in \partial D\}. \tag{2.11}$$

(b) *There exists a unique* $g^* \in \mathcal{C}(\overline{D}) \cap \mathcal{C}^2(D)$ *that solves*

$$\begin{aligned} -\tfrac{1}{2}\Delta g^* &= 1 \qquad && \text{on } D \\ g^* &= 0 \qquad && \text{on } \partial D. \end{aligned} \tag{2.12}$$

The solution is given (with τ *as in (2.11)) by*

$$g^*(x) = E^x[\tau] \tag{2.13}$$

and satisfies $g^* > 0$ *on* D*. There exists an* $L < \infty$ *such that*

$$g^*(x) \leq L|x-y| \qquad \forall x \in \overline{D},\ y \in \partial D. \tag{2.14}$$

Proof of Lemma 2.2: Formulas (2.9) and (2.10) can be found in [14], Proposition 4.2.7 and Theorems 4.2.12 and 4.2.19. For (2.13) see [14], Problem 4.2.25. The fact that $g^* > 0$ on D can easily be deduced from the representation (2.13), but alternatively one may consult [15], Theorem 2.5. To prove (2.14), we assume without loss of generality that $y = 0$ and $x_1 > 0$ $\forall x \in D$, where for any $x \in \mathbb{R}^d$ we write $x = (x_1, \ldots, x_d)$. Now choose L such that $|x - \tilde{x}| \leq L$ for all $x, \tilde{x} \in \overline{D}$. Define a stopping time $\tilde{\tau}$ by

$$\tilde{\tau} := \inf\{t \geq 0 : B^1_t \in \{0, L\}\}, \tag{2.15}$$

where $B_t = (B^1_t, \ldots, B^d_t)$ is d-dimensional Brownian motion. By [14], Problem 4.2.25, we have

$$g^*(x) = E^x[\tau] \leq E^x[\tilde{\tau}] = x_1(L - x_1) \leq Lx_1 \leq L|x-y|. \tag{2.16}$$

∎

2.3 Proof of Theorem 1.1

Theorem 1.1 follows directly from the following lemma. Formula (2.17) (ii) below will be essential for the rest of this section.

Lemma 2.3 *Fix* $g \in \mathcal{H}'$ *and* $c \in (0, \infty)$*. For any* $\theta \in \overline{D}$*, denote by* $(S_t)_{t\geq 0}$ *the Feller semigroup related to the solution* $(X_t)_{t\geq 0}$ *of the martingale problem associated with* A *in (1.8), and let* G *be the full generator of* $(S_t)_{t\geq 0}$*. Then, for any* $\theta \in \overline{D}$*, the equilibrium* $\nu^{g,c}_\theta$ *of (1.7) is the unique solution of any of the following two equations:*

$$\begin{aligned} &\text{(i)} \quad \langle \nu^{g,c}_\theta | S_t f\rangle = \langle \nu^{g,c}_\theta | f\rangle \qquad && \forall t \geq 0, f \in \mathcal{C}(\overline{D}) \\ &\text{(ii)} \quad \langle \nu^{g,c}_\theta | G f\rangle = 0 \qquad && \forall f \in \mathcal{D}(G). \end{aligned} \tag{2.17}$$

For $\theta \in \partial D$, $\nu^{g,c}_\theta = \delta_\theta$ *and for* $\theta \in D$ *the measure* $\nu^{g,c}_\theta$ *satisfies* $\nu^{g,c}_\theta(D) > 0$*. Furthermore, the map* $\theta \mapsto \nu^{g,c}_\theta$ *is continuous with respect to the topology of weak convergence.*

Proof of Lemma 2.3: For simplicity we drop the superscripts g, c. Relation (2.17) (i) means that $E[f(X_t)]$ is independent of t when $(X_t)_{t\geq 0}$ is the solution of (1.7) with initial condition ν_θ. So (2.17) (i) just says that ν_θ is the unique equilibrium of (1.7), which is by definition true for $g \in \mathcal{H}'$. To prove (2.17) (ii), note that $Gf = \lim_{t\to 0} t^{-1}(S_t f - f)$ for all $f \in \mathcal{D}(G)$, where the limit is in the norm $\|\cdot\|$. So differentiating (2.17) (i), we get (2.17) (ii). To show that (2.17) (ii) determines ν_θ uniquely, note that for all $f \in \mathcal{D}(G)$ it holds that $S_t f \in \mathcal{D}(G)$ $\forall t \geq 0$ and $\frac{\partial}{\partial t} S_t f = G S_t f$, where the differentiation is in the Banach space $\mathcal{C}(\overline{D})$ (see [11], Proposition 1.1.5 (b)). Now, with $\tilde{\nu}_\theta$ a solution of (2.17) (ii), we have

$$\tfrac{\partial}{\partial t}\langle \tilde{\nu}_\theta | S_t f\rangle = \langle \tilde{\nu}_\theta | G S_t f\rangle = 0 \qquad \forall t \geq 0,\ f \in \mathcal{D}(G), \tag{2.18}$$

and this implies (2.17) (i) for $f \in \mathcal{D}(G)$. Since $\mathcal{D}(G)$ is dense in $\mathcal{C}(\overline{D})$, (2.17) (i) holds for general $f \in \mathcal{C}(\overline{D})$ and hence $\tilde{\nu}_\theta = \nu_\theta$.

To see that $\nu_\theta = \delta_\theta$ if $\theta \in \partial D$, note that $X_t \equiv \theta$ solves (1.7), so δ_θ is an equilibrium of (1.7). To see that $\nu_\theta(D) > 0$ for $\theta \in D$, insert $f(x) = |x-\theta|^2$ into (2.17) (ii) to get $c\langle \nu_\theta | f\rangle = d\langle \nu_\theta | g\rangle$ (compare also Lemma 2.4). Now f is strictly bounded away from zero on ∂D, so $\langle \nu_\theta | g\rangle > 0$. Since $g = 0$ on ∂D this implies $\nu_\theta(D) > 0$.

We next show that the probability kernel ν_θ is continuous in θ. For each $\theta \in \overline{D}$ let $(S_t^\theta)_{t\geq 0}$ be the Feller semigroup above and let G^θ be its generator. Let $\theta_n, \theta \in \overline{D}$ with $\theta_n \to \theta$. Using the fact that the martingale problem is well-posed for all θ, we have by [20], Theorem 11.1.4,

$$S_t^{\theta_n} f \to S_t^\theta f \qquad \forall f \in \mathcal{C}(\overline{D}),\ t \geq 0, \tag{2.19}$$

where the convergence is in $\mathcal{C}(\overline{D})$. By [11], Theorem 1.6.1 (c), it follows that for all $f \in \mathcal{D}(G^\theta)$ there exist $f_n \in \mathcal{D}(G^{\theta_n})$ such that

$$G^{\theta_n} f_n \to G^\theta f \qquad \text{as } n \to \infty, \tag{2.20}$$

again in the topology on $\mathcal{C}(\overline{D})$. Now consider the sequence ν_{θ_n}. By compactness, it has a cluster point. For any such cluster point $\tilde{\nu}_\theta$, choose a subsequence such that ν_{θ_n} converges to $\tilde{\nu}_\theta$, and observe that for each $f \in \mathcal{D}(G^\theta)$, with f_n as in (2.20),

$$\begin{aligned} |\langle \tilde{\nu}_\theta | G^\theta f\rangle| &\leq |\langle \tilde{\nu}_\theta | G^\theta f\rangle - \langle \nu_{\theta_n} | G^\theta f\rangle| + |\langle \nu_{\theta_n} | G^\theta f\rangle - \langle \nu_{\theta_n} | G^{\theta_n} f_n\rangle| + |\langle \nu_{\theta_n} | G^{\theta_n} f_n\rangle| \\ &\leq |\langle \tilde{\nu}_\theta | G^\theta f\rangle - \langle \nu_{\theta_n} | G^\theta f\rangle| + \|G^\theta f - G^{\theta_n} f_n\| + 0, \end{aligned} \tag{2.21}$$

where the right-hand side tends to zero as $n \to \infty$. By (2.17) (ii), it follows that $\tilde{\nu}_\theta = \nu_\theta$ for each cluster point $\tilde{\nu}_\theta$ of the ν_{θ_n}, and hence ν_{θ_n} converges to ν_θ. ∎

2.4 Proof of Theorems 1.2–1.4

The proofs of Theorems 1.2–1.4 are based on the following lemma:

Lemma 2.4 *For any $g \in \mathcal{H}'$ and $c \in (0,\infty)$, let $\nu^{g,c} \in \mathcal{K}(\overline{D})$ as in Theorem 1.1. Fix $\lambda \in \mathbb{R}$. Assume that $f \in \mathcal{C}(\overline{D}) \cap \mathcal{C}^2(D)$ satisfies*

$$-\tfrac{1}{2}\Delta f = \lambda \qquad \text{on } D. \tag{2.22}$$

Then

$$\nu^{g,c} f = f - \tfrac{\lambda}{c}\,\nu^{g,c} g. \tag{2.23}$$

Proof of Lemma 2.4: We start with the case $f \in \mathcal{C}^2(\overline{D})$. Let $(T_t^{\theta,c})_{t\geq 0}$ be the Feller semigroup on $\mathcal{C}(\overline{D})$ defined by

$$(T^{\theta,c} f)(x) := f(\theta + e^{-ct}(x-\theta)) \qquad f \in \mathcal{C}(\overline{D}). \tag{2.24}$$

This is the semigroup related to our process in (1.7) when the local diffusion function g is set to zero. If $B_{\theta,c}$ is its full generator, then for every $f \in \mathcal{C}^1(\overline{D})$

$$(B_{\theta,c} f)(x) = c(\theta - x)\cdot \nabla f(x). \tag{2.25}$$

Let us introduce an operator that is in some sense an inverse to $B_{\theta,c}$. Define

$$\begin{aligned} \mathcal{D}(B_{\theta,c}^{-1}) &:= \{f \in \mathcal{C}(\overline{D}) : \int_0^\infty \|T_t^{\theta,c} f\|\,dt < \infty\} \\ B_{\theta,c}^{-1} f &:= -\int_0^\infty T_t^{\theta,c} f\,dt. \end{aligned} \tag{2.26}$$

It follows that

$$B_{\theta,c} B_{\theta,c}^{-1} f = f \qquad \forall f \in \mathcal{D}(B_{\theta,c}^{-1}), \tag{2.27}$$

as can be seen by writing (compare the proof of [11], Proposition 1.1.5 (a))

$$\begin{aligned} B_{\theta,c} B_{\theta,c}^{-1} f &= \lim_{\varepsilon\to 0} -\varepsilon^{-1}(T_\varepsilon^{\theta,c} - 1)\int_0^\infty T_t^{\theta,c} f\,dt \\ &= \lim_{\varepsilon\to 0} \varepsilon^{-1}\int_0^\infty (T_t^{\theta,c} f - T_{t+\varepsilon}^{\theta,c})\,dt \\ &= \lim_{\varepsilon\to 0} \varepsilon^{-1}\left(\int_0^\infty T_t^{\theta,c} f\,dt - \int_\varepsilon^\infty T_t^{\theta,c} f\,dt\right) \\ &= \lim_{\varepsilon\to 0} \varepsilon^{-1}\int_0^\varepsilon T_t^{\theta,c} f\,dt = f. \end{aligned} \tag{2.28}$$

Now let $f \in \mathcal{C}^2(\overline{D})$, $-\frac{1}{2}\Delta f = \lambda$. Then

$$\begin{aligned} & f - f(\theta) \in \mathcal{D}(B_{\theta,c}^{-1}) \\ & B_{\theta,c}^{-1}(f - f(\theta)) \in \mathcal{C}^2(\overline{D}) \\ & \Delta B_{\theta,c}^{-1}(f - f(\theta)) = \tfrac{\lambda}{c}. \end{aligned} \tag{2.29}$$

To see this, substitute the variables $u = e^{-ct}$, $du = -ce^{-ct}dt$ into (2.26) to get

$$B_{\theta,c}^{-1}(f - f(\theta))(x) = -\int_0^1 \tfrac{1}{cu}\Big(f(\theta + u(x-\theta)) - f(\theta)\Big)du. \tag{2.30}$$

Since f is differentiable at θ, the integrand is bounded and it follows that $f - f(\theta) \in \mathcal{D}(B_{\theta,c}^{-1})$. Interchanging differentiation and integration, we get the following expressions for the derivatives of $B_{\theta,c}^{-1}(f - f(\theta))$:

$$\begin{array}{rcl} \frac{\partial}{\partial x_i} B_{\theta,c}^{-1} f(x) & = & -\int_0^1 \frac{1}{c}(\frac{\partial}{\partial x_i} f)(\theta + u(x-\theta))du \\ \frac{\partial^2}{\partial x_i \partial x_j} B_{\theta,c}^{-1} f(x) & = & -\int_0^1 \frac{u}{c}(\frac{\partial^2}{\partial x_i \partial x_j} f)(\theta + u(x-\theta))du. \end{array} \tag{2.31}$$

The interchanging is allowed because the integrands on the right-hand sides are absolutely integrable. In particular, it follows that $\Delta B_{\theta,c}^{-1}(f - f(\theta)) = -\int_0^1 \frac{u}{c}\Delta(f - f(\theta))du = \int_0^1 \frac{u}{c} 2\lambda du = \frac{\lambda}{c}$.

Applying (2.17) (ii) to the function $B_{\theta,c}^{-1}(f - f(\theta)) \in \mathcal{C}^2(\overline{D}) \subset \mathcal{D}(G)$, we get

$$0 = \langle \nu_\theta^{g,c} | (B_{\theta,c} + g\Delta) B_{\theta,c}^{-1}(f - f(\theta))\rangle = \langle \nu_\theta^{g,c} | f - f(\theta)\rangle + \langle \nu_\theta^{g,c} | \tfrac{\lambda}{c} g\rangle, \tag{2.32}$$

which gives (2.23). To extend formula (2.23) to $f \in \mathcal{C}(\overline{D}) \cap \mathcal{C}^2(D)$, pick an $x_0 \in D$ and a sequence $a_n \in (0,1)$ with $a_n \to 1$ as $n \to \infty$. Define functions $f_n \in \mathcal{C}^2(\overline{D})$ by

$$f_n(x) = a_n^{-2} f(x_0 + a_n(x - x_0)). \tag{2.33}$$

Then $-\frac{1}{2}\Delta f_n = \lambda$ for each n and $\|f_n - f\| \to 0$. Letting $n \to \infty$ and using the continuity of $\nu^{g,c}$, we conclude that (2.23) holds for f. ∎

We recall that by the definitions in section 1.3

$$\begin{array}{l} F_c g = \nu^{g,c} g \\ F^{(k)} = F_{c_k} \circ \cdots \circ F_{c_1} \\ K^{g,(k)} = \nu^{F^{(k-1)}g,c_k} \cdots \nu^{g,c_1}, \end{array} \tag{2.34}$$

so that

$$F^{(k)} g = K^{g,(k)} g. \tag{2.35}$$

The following lemma now follows easily by iterating Lemma 2.4.

Lemma 2.5 *Let $g \in \mathcal{H}'$, $c_1, \ldots, c_k \in (0,\infty)$, and let f be as in Lemma 2.4. Define $K^{g,(k)}$ and $F^{(k)}$ as in (2.34), and assume that $F^{(1)}g, \ldots, F^{(k-1)}g \in \mathcal{H}'$. Then*

$$K^{g,(k)} f = f - \lambda \sigma_k F^{(k)} g, \tag{2.36}$$

with $\sigma_k = \sum_{l=1}^k \frac{1}{c_l}$.

We are now ready to prove Theorems 1.2–1.4.

Proof of Theorem 1.2: Since $\nu^{g,c}$ is a continuous probability kernel, it follows from (2.34) that $F_c g \in \mathcal{C}(\overline{D})$. If $\theta \in \partial D$, then $\nu_\theta^{g,c} = \delta_\theta$ by Lemma 2.3 and so $(F_c g)(\theta) = 0$. If $\theta \in D$ then by the same lemma $\nu_\theta^{g,c}(D) > 0$ and so $(F_c g)(\theta) > 0$. Finally, inserting $f = g^*$ into Lemma 2.4, we get $F_c g = \nu^{g,c} g = cg^* - c\nu^{g,c} g^* \leq cg^*$. ∎

Proof of Theorem 1.3: By Lemma 2.1, we must show that $K^{g,(k)} f \to Hf$ as $k \to \infty$ in the norm on $\mathcal{C}(\overline{D})$ for each $f \in \mathcal{C}(\overline{D})$, where H is defined by (2.10). By Lemma 2.5,

$$K^{g,(k)} g^* = g^* - \sigma_k K^{g,(k)} g. \tag{2.37}$$

It follows that $0 \leq \sigma_k K^{g,(k)} g \leq g^*$, and since $\sigma_k \to \infty$ we have $\|K^{g,(k)} g\| \to 0$. Since $g > 0$ on D, this in fact implies that for any $f \in \mathcal{C}(\overline{D})$ with $f \equiv 0$ on ∂D

$$\|K^{g,(k)} f\| \to 0. \tag{2.38}$$

To see why, define $R_n := \{x \in \overline{D} : \exists y \in \partial D,\ |x - y| < \frac{1}{n}\}$. Choose $\phi_n \in \mathcal{C}(\overline{D})$, $0 \leq \phi_n \leq 1$, such that $\phi_n \equiv 0$ on R_{n+1} and $\phi_n \equiv 1$ on $\overline{D} \backslash R_n$. For each n there exists an $M_n < \infty$ such that $\phi_n \leq M_n g$, so $\|K^{g,(k)} \phi_n\| \to 0$ as $k \to \infty$. We may choose a subsequence $n_k \to \infty$ such that

$$\|K^{g,(k)} \phi_{n_k}\| \to 0 \qquad \text{as } k \to \infty. \tag{2.39}$$

Using this, we can estimate for f:

$$\|K^{g,(k)} f\| \leq \|f\| \cdot \|K^{g,(k)} \phi_{n_k}\| + \max_{x \in R_{n_k}} |f(x)|, \tag{2.40}$$

where the right-hand side tends to zero as $k \to \infty$. This proves (2.38).

For any $f \in \mathcal{C}(\overline{D})$ we can now write

$$K^{g,(k)} f = K^{g,(k)}(Hf - (Hf - f)) = Hf - K^{g,(k)}(Hf - f) \to Hf, \tag{2.41}$$

where we use (2.36) and (2.38). ∎

Proof of Theorem 1.4: Pick $g = f = rg^*$ in Lemma 2.4 to get

$$F_c(rg^*) = rg^* - \frac{r}{c} F_c(rg^*), \tag{2.42}$$

which implies Theorem 1.4 (a). To prove Theorem 1.4 (b) we observe that by Lemma 2.5,

$$\sigma_k F^{(k)} g = g^* - K^{g,(k)} g^*. \tag{2.43}$$

By (2.38), $\|K^{g,(k)} g^*\| \to 0$ as $k \to \infty$, and the theorem follows. To prove Theorem 1.4 (c), note that by the reasoning following (2.37),

$$\|K^{g,(k)} g\|_{\mathcal{H}} \to 0 \qquad \text{as } k \to \infty. \tag{2.44}$$

In the special case that $g \geq \lambda g^*$ for some $\lambda > 0$, it also follows that $\|K^{g,(k)} g^*\|_{\mathcal{H}} \to 0$ as $k \to \infty$. Inserting this into (2.43), we see that for such g, the convergence can be strengthened to

$$\|\sigma_k F^{(k)} g - g^*\|_{\mathcal{H}} \to 0 \qquad \text{as } k \to \infty. \tag{2.45}$$

∎

3 Ergodicity: Proof of Theorem 1.5

Theorem 1.5 follows from the following more technical lemma. In this section, we use the symbol ν for the probability measure $\nu_\theta^{g,c}$ (so ν denotes a probability measure, not a probability kernel).

Lemma 3.1 *Fix $g \in \mathcal{H}$, $\theta \in \overline{D}$ and $c \in (0, \infty)$. Assume that g is locally Hölder continuous (with positive exponent) on D and that the martingale problem associated with the operator A in (1.8) is well-posed. Then the SDE (1.7) has a unique equilibrium $\nu \in \mathcal{P}(\overline{D})$. Furthermore, for every $f \in \mathcal{C}(\overline{D})$*

$$\|S_t f - \langle \nu | f \rangle\| \to 0 \qquad \textit{as } t \to \infty. \tag{3.1}$$

If $\theta \in D$, then there exist $t_0, r > 0$ such that (3.1) can be sharpened as follows: For all $f : \overline{D} \to [0, 1]$ measurable

$$\|S_t f - \langle \nu | f \rangle\| \leq e^{-r(t - t_0)}. \tag{3.2}$$

The proof of Lemma 3.1 is long and will keep us busy for the rest of this section. For notational simplicity we treat the case $c = 1$ only. Other c follow trivially by the scaling property $\mathcal{A}_\theta^{\lambda c, \lambda g} = \lambda \mathcal{A}_\theta^{c,g}$.

We start by proving (3.2). To that end we introduce two compact sets $\overline{B} \subset \overline{C} \subset D$, and prove that the expected time for X_t, starting from any point in $\overline{D}$, to reach into B is bounded uniformly in the starting point (Lemma 3.2). On C we then use results from the theory of non-degenerate diffusions to show that the distribution of the process starting from B can be bounded from below in a uniform way (Lemma 3.3). Combining these two results we arrive at Lemma 3.4, which shows that there exists a ν such that (3.2) holds. Once we have shown formula (3.2), it follows that (3.1) holds for $\theta \in D$. The case $\theta \in \partial D$ can then easily be treated separately. We end by showing that ν is the unique equilibrium of (1.7).

Without loss of generality we may assume $\theta = 0$. Choose $\varepsilon > 0$ such that $|x| \leq 2\varepsilon \Rightarrow x \in D$ and define:

$$\begin{aligned} B &:= \{x \in \overline{D} : |x| < \varepsilon\} \\ C &:= \{x \in \overline{D} : |x| < 2\varepsilon\}. \end{aligned} \tag{3.3}$$

Lemma 3.2 *Let B be as in (3.3). Denote by $(X_t^x)_{t\geq 0}$ the process X starting at $X_0 = x$, and define a stopping time τ_B^x by*

$$\tau_B^x := \inf\{t \geq 0 : X_t \in \overline{B}\}. \tag{3.4}$$

Then there exists a constant $T < \infty$ such that

$$\sup_{x\in\overline{D}} E[\tau_B^x] \leq T. \tag{3.5}$$

Proof of Lemma 3.2: Let h_d denote the function

$$h_d(x) := \begin{cases} -\log|x| & (d=2) \\ (d-2)^{-1}|x|^{2-d} & (d \neq 2). \end{cases} \tag{3.6}$$

This function satisfies

$$\begin{aligned} \nabla h_d(x) &= -x|x|^{-d} \\ \Delta h_d(x) &= 0. \end{aligned} \tag{3.7}$$

For $\lambda \geq 0$, we define a function r_λ on $\overline{D}\backslash B$ by

$$r_\lambda(x) := -\log|x| + \lambda h_d(x). \tag{3.8}$$

We shall show that it is possible to choose λ such that $\mathcal{A}r_\lambda \geq 1$, with $\mathcal{A}$ the differential form in (2.4). Indeed, a little calculation shows that

$$\mathcal{A}r_\lambda(x) = 1 + \Big(\lambda + (2-d)g(x)|x|^{d-4}\Big)|x|^{2-d}. \tag{3.9}$$

and so we may choose

$$\begin{array}{ll} \lambda = 0 & (d \leq 2) \\ \lambda = \max_{x\in\overline{D}\backslash B}\ g(x)|x|^{-1} & (d=3) \\ \lambda = \max_{x\in\overline{D}}\ (d-2)g(x)|x|^{d-4} & (d \geq 4). \end{array} \tag{3.10}$$

Next, we can extend r_λ to a function in $\mathcal{C}^2(\overline{D})$, which now has the property (with A the operator in (1.8))

$$(Ar_\lambda)(x) \geq 1 \qquad (x \in \overline{D}\backslash B). \tag{3.11}$$

Abbreviate $\tau = \tau_B^x$ and let $r : [\varepsilon, \infty) \to \mathbb{R}$ be the (decreasing) function such that $r_\lambda(x) = r(|x|)$. The process X solves the martingale problem for A, so for each $x \in \overline{D}\backslash B$ and $t \geq 0$ we have

$$E[\tau \wedge t] \leq E\left[\int_0^{\tau\wedge t}(Ar_\lambda)(X_s)ds\right] = E[r(|X_{\tau\wedge t}|)] - r(|x|) \leq r(\varepsilon) - r(|x|). \tag{3.12}$$

The case $x \in B$ can be added trivially, and letting $t \uparrow \infty$ we find

$$E[\tau_B^x] \leq r(\varepsilon) - \min_{y\in\overline{D}} r(|y|) \qquad \forall x \in \overline{D}, \tag{3.13}$$

which completes the proof. ∎

We have shown that no matter where the process X starts in $\overline{D}$, it reaches into the set $\overline{B}$ in a finite expected time that is uniform in the starting point. We next turn our attention to the process starting in $\overline{B}$. We shall prove:

Lemma 3.3 *Let $(S_t)_{t\geq 0}$ be the Feller semigroup associated with X and let C be as in (3.3). For each $0 < t_1 < t_2$ there exists a non-zero finite measure μ on $\overline{D}$ such that*

$$\inf_{t\in[t_1,t_2]} \inf_{x\in\overline{B}} (S_t f)(x) \geq \langle f|\mu\rangle \qquad (f \geq 0,\ f \in \mathcal{C}(\overline{D})). \tag{3.14}$$

Proof of Lemma 3.3: We shall compare X with the process vanishing at ∂C. To that end, let

$$\tau := \inf\{t \geq 0 : X_t^x \in \overline{D}\backslash C\}. \tag{3.15}$$

Note that for any $f \in \mathcal{C}(\overline{D})$

$$(S_t f)(x) = E[f(X_t^x)] \geq E[f(X_t^x)1_{\{t\leq\tau\}}]. \tag{3.16}$$

The function $(t,x) \mapsto E[f(X_t^x)1_{\{t\leq\tau\}}]$ is the solution of a Cauchy problem on $[0,\infty)\times\overline{C}$ with Dirichlet boundary conditions on ∂C. Since the operator A is uniformly elliptic on $\overline{C}$ and the function g is Hölder continuous on $\overline{C}$, it is known (see [4], volume II, appendix §6, Theorem 0.6 and [12], Corollary 3.7.1) that a fundamental solution to this Cauchy problem exists. In particular, there exists a function $p \in \mathcal{C}((0,\infty)\times\overline{C}\times C)$ with the properties:

$$\begin{array}{ll} E[f(X_t^x)1_{\{t\leq\tau\}}] = \int_C p_t(x|y)f(y)dy & (f \in \mathcal{C}(\overline{C})) \\ p_t(x|y) > 0 & ((t,x,y) \in (0,\infty)\times C\times C). \end{array} \tag{3.17}$$

Note that $p_t(x,\cdot)$ is the probability density of the process vanishing at ∂D. Applying (3.17), we get Lemma 3.3 if we choose for μ the measure on $\overline{C}$ given by

$$\begin{array}{l} \mu(dy) = \mu(y)dy \\ \mu(y) := \min\{p_t(x|y) : t \in [t_1,t_2],\ x \in \overline{B}\}. \end{array} \tag{3.18}$$

∎

Combining Lemmas 3.2 and 3.3 we get:

Lemma 3.4 *For all $\theta \in D$ there exists a $t_0 \in (0,\infty)$ and a non-zero finite measure μ on $\overline{D}$ such that, for all $f \in \mathcal{C}(\overline{D})$, $f \geq 0$,*

$$S_{t_0} f \geq \langle \mu|f\rangle. \tag{3.19}$$

Proof of Lemma 3.4: From Lemma 3.2 we get

$$P[\tau_B^x \leq 2T] \geq \tfrac{1}{2}. \tag{3.20}$$

Let $x \in \overline{D}$, and denote the distribution of $X^x_{\tau_B^x}$ by ρ. Let X_t^ρ be the process X with initial distribution ρ. By Lemma 3.3, there exists a μ such that

$$E[f(X_t^\rho)] \geq \langle f|\mu\rangle \qquad (t \in [T, 3T]). \tag{3.21}$$

By (3.20), (3.21) and the strong Markov property, we have for all $x \in \overline{D}$ and $f \in \mathcal{C}(\overline{D})$

$$\begin{aligned}
(S_{3T}f)(x) &\geq E[f(X_{3T}^x)1_{\{\tau_B^x \leq 2T\}}] \\
&= \int_0^{2T} E[f(X_{3T}^x)|\tau_B^x = s]P[\tau_B^x \in ds] \\
&= \int_0^{2T} E[f(X_{3T-s}^\rho)]P[\tau_B^x \in ds] \\
&\geq \langle f|\mu\rangle \int_0^{2T} P[\tau_B^x \in ds] \\
&\geq \tfrac{1}{2}\langle f|\mu\rangle,
\end{aligned} \tag{3.22}$$

and (3.19) follows if we replace μ by $\frac{1}{2}\mu$ and set $t_0 = 3T$. ∎

We have now completed the preparatory work and are ready for:

Proof of Lemma 3.1: From Lemma 3.4 we get (3.2) with a standard technique. This goes as follows. Fix a measurable $f : \overline{D} \to [0, 1]$ and define, for $t \geq 0$,

$$\begin{aligned}
\nu_t^+ &:= \sup\nolimits_{x\in\overline{D}}(S_t f)(x) \\
\nu_t^- &:= \inf\nolimits_{x\in\overline{D}}(S_t f)(x).
\end{aligned} \tag{3.23}$$

By Lemma 3.4,

$$S_{t+t_0}f = S_{t_0}(\nu_t^+ - (\nu_t^+ - S_t f)) = \nu_t^+ - S_{t_0}(\nu_t^+ - S_t f) \leq \nu_t^+ - \langle \mu|\nu_t^+ - S_t f\rangle. \tag{3.24}$$

A similar argument applies to ν_t^-, and we get

$$\begin{aligned}
\nu_{t+t_0}^+ &\leq \nu_t^+ - \langle \mu|\nu_t^+ - S_t f\rangle \\
\nu_{t+t_0}^- &\geq \nu_t^- + \langle \mu|S_t f - \nu_t^-\rangle.
\end{aligned} \tag{3.25}$$

It follows that

$$\nu_{t+t_0}^+ - \nu_{t+t_0}^- \leq \nu_t^+ - \nu_t^- - \langle \mu|\nu_t^+ - \nu_t^-\rangle = (1 - \langle \mu|1\rangle)(\nu_t^+ - \nu_t^-), \tag{3.26}$$

and by induction that

$$\nu_{t+nt_0}^+ - \nu_{t+nt_0}^- \leq (1 - \langle \mu|1\rangle)^n. \tag{3.27}$$

We thus see that $S_t f$ converges uniformly to a constant. This constant we can formally denote by $\langle\nu|f\rangle$. Formula (3.2) now holds with $r = -t_0^{-1}\log(1-\langle\mu|1\rangle)$.

To complete the proof of (3.2), it is left to show that $f \mapsto \langle\nu|f\rangle$, defined implicitly above for all measurable $f : \overline{D} \to [0,1]$, indeed corresponds to a probability measure. It is sufficient to show that the map is linear, positive, satisfies $\langle\nu|1\rangle = 1$, and is continuous with respect to increasing sequences of functions. The first three properties are easy. We therefore only show continuity.

Let $B_1(\overline{D}) := \{f : \overline{D} \to [0,1] : f \text{ measurable}\}$ and let $f_i, f_\infty \in B_1(\overline{D})$, $f_i \uparrow f_\infty$. Fix any probability measure ρ on $\overline{D}$. As $t \to \infty$, $S_t f$ converges at a rate that is uniform in $f \in B_1(\overline{D})$, so for every $\varepsilon > 0$ there exists a $t > 0$ such that

$$|\langle\rho|S_t f\rangle - \langle\nu|f\rangle| \leq \varepsilon \qquad \forall f \in B_1(\overline{D}). \tag{3.28}$$

There exists an n such that

$$|\langle\rho|S_t f_i\rangle - \langle\rho|S_t f_\infty\rangle| \leq \varepsilon \quad \forall i \geq n, \tag{3.29}$$

and it follows that for every $i \geq n$

$$\begin{aligned} &|\langle\nu|f_i\rangle - \langle\nu|f_\infty\rangle| \\ &\quad \leq |\langle\nu|f_i\rangle - \langle\rho|S_t f_i\rangle| + |\langle\rho|S_t f_i\rangle - \langle\rho|S_t f_\infty\rangle| + |\langle\rho|S_t f_\infty\rangle - \langle\nu|f_\infty\rangle| \leq 3\varepsilon. \end{aligned} \tag{3.30}$$

Note that $\nu \geq \mu$, so $\nu(D) > 0$. This completes the proof of (3.2).

Trivially, (3.2) implies (3.1) for $\theta \in D$. We next turn to the proof of (3.1) for $\theta \in \partial D$. As we shall see, in this case ν turns out to be δ_θ. Let X_t be any solution to the martingale problem associated with A. For $x \in \overline{D}$ write $x = (x_1, \ldots, x_d)$, and write $X_t = (X_t^1, \ldots, X_t^d)$. Without loss of generality we may assume $\theta = 0$ and $x_1 > 0$ $\forall x \in D$. From the martingale problem (2.2) we have for $i = 1, \ldots, d$

$$\begin{aligned} E[X_t^i] &= E[X_0^i] - \int_0^t E[X_s^i]ds \\ E[|X_t|^2] &= E[|X_0|^2] + 2d\int_0^t E[g(X_s)]ds - 2\int_0^t E[|X_s|^2]ds. \end{aligned} \tag{3.31}$$

We see immediately that

$$E[X_t^i] = E[X_0^i]e^{-t}. \tag{3.32}$$

By (1.6), and by (2.14) in Lemma 2.2 (b), there exists a constant $L < \infty$ such that

$$E[g(X_t)] \leq LE[X_t^1]. \tag{3.33}$$

Let

$$M := \sup\{|x-y| : x, y \in \overline{D}\}. \tag{3.34}$$

Then (3.32) and (3.33) imply

$$E[g(X_t)] \leq LMe^{-t}. \tag{3.35}$$

Next, the function $t \mapsto E[|X_t|^2]$ is differentiable and satisfies

$$\frac{\partial}{\partial t} E[|X_t|^2] = 2dE[g(X_t)] - 2E[|X_t|^2]. \tag{3.36}$$

From this it follows that

$$\frac{\partial}{\partial t}\left(E[|X_t|^2]e^{2t}\right) = 2dE[g(X_t)]e^{2t} \tag{3.37}$$

and therefore

$$E[|X_t|^2] \leq e^{-2t}E[|X_0|^2] + e^{-2t}\int_0^\infty 2dE[g(X_s)]e^{2s}ds. \tag{3.38}$$

Here, by (3.35),

$$\begin{aligned} &e^{-2t}\textstyle\int_0^\infty 2dE[g(X_s)]e^{2s}ds = \int_0^\infty 2dE[g(X_s)]e^{2(t-s)}ds \\ &\leq e^{-t}\textstyle\int_0^t 2dE[g(X_s)]ds + \int_t^\infty 2dE[g(X_s)]ds \leq 2dLMe^{-t} + 2dLMe^{-t}. \end{aligned} \tag{3.39}$$

Hence, combining (3.38) and (3.39) we get

$$E[|X_t|^2] \leq e^{-2t}M^2 + 4dLMe^{-t}, \tag{3.40}$$

which tends to zero as $t \to \infty$. Since also $E[X_t^i] \to 0$ as $t \to \infty$, Chebyshev's inequality shows that $(S_t f)(x) = E[f(X_t^x)]$ converges to $\langle \delta_0 | f\rangle = f(0)$ for each $f \in \mathcal{C}(\overline{D})$, and (3.40) shows that this convergence is uniform in the initial value x. This completes the proof of (3.1).

We complete the proof of the theorem by showing that ν is the unique equilibrium of (1.7). This means that we must show (compare (2.17) (i)) that ν is the unique solution of

$$\langle \nu | S_t f\rangle = \langle \nu | f\rangle \qquad \forall t \geq 0, f \in \mathcal{C}(\overline{D}). \tag{3.41}$$

First, for any $x \in \overline{D}$,

$$\langle \nu | S_t f\rangle = \lim_{s\to\infty}(S_s S_t f)(x) = \lim_{s\to\infty}(S_{s+t}f)(x) = \langle \nu | f\rangle, \tag{3.42}$$

which proves that (3.41) holds. Suppose that $\tilde{\nu}$ is another solution. Let $t \to \infty$ in (3.41) and use that $S_t f \to \langle \nu | f\rangle$. By dominated convergence, $\langle \tilde{\nu} | S_t f\rangle \to \langle \nu | f\rangle$. So $\langle \tilde{\nu} | f\rangle = \langle \nu | f\rangle$ for all $f \in \mathcal{C}(\overline{D})$, and hence $\tilde{\nu} = \nu$. ∎

Remark Formula (3.2) actually shows that $\nu(D) = 1$ for $\theta \in D$, whenever it is true that for $x \in D$

$$P[X_t^x \in D] = 1 \qquad \forall t \geq 0. \tag{3.43}$$

Formula (3.43) holds, for example, under the conditions of Theorem 1.9, but no doubt much more generally too.

4 The martingale problem

In this section we prove the theorems about the martingale problem for A mentioned in section 1.4. The proofs of Theorems 1.7 and 1.8 have already been indicated in the text.

4.1 Existence: Proof of Theorem 1.6

We extend the function g to $\mathbb{R}^d$ by putting $g \equiv 0$ on $\mathbb{R}^d \backslash \overline{D}$. Let $\mu \in \mathcal{P}(\overline{D})$. By [11], Theorem 5.3.10, there exists an $\mathbb{R}^d$-valued weak solution to the SDE

$$dX_t = c(\theta - X_t)dt + \sqrt{2\overline{g}(X_t)}dB_t \tag{4.1}$$

with initial distribution $P[X_0 \in dx] = \mu(dx)$. By the same theorem, X solves the martingale problem for the operator $\{(f, \mathcal{A}f) : f \in \mathcal{C}_c^\infty(\mathbb{R}^d)\}$. By [11], Proposition 7.1 from the appendix, there exist $f_n \in \mathcal{C}_c^\infty(\mathbb{R}^d)$ such that $f_n \to f$ and $\mathcal{A}f_n \to \mathcal{A}f$ uniformly on $\mathbb{R}^d$. By [11], Lemma 4.5.1, X now also solves the martingale problem for $\{(f, \mathcal{A}f) : f \in \mathcal{C}_c^2(\mathbb{R}^d)\}$.

Pick $x_i \in \overline{D}$, $R_i \in (0,\infty)$ such that $\overline{D} = \bigcap_i \{x \in \mathbb{R}^d : |x - x_i| \leq R_i\}$. Let $h \in \mathcal{C}^2(\mathbb{R})$, $h \equiv 1$ on $(-\infty, 0]$, $h \equiv 0$ on $[1,\infty)$ and $h' \leq 0$. Define $f_i \in \mathcal{C}_c^2(\mathbb{R}^d)$ by $f_i(x) := h(|x - x_i| - R_i)$. Then $f_i \in \mathcal{C}_c^2(\mathbb{R}^d)$ and it is easy to see that $\mathcal{A}f_i \geq 0$. By the martingale problem,

$$E[f_i(X_t)] = 1 + E[\int_0^t (\mathcal{A}f_i)(X_s)ds] \geq 1, \tag{4.2}$$

which shows that $P[|X_t - x_i| \leq R_i] = 1 \ \forall t \geq 0 \ \forall i$. By the continuity of X, it follows that $P[|X_t - x_i| \leq R_i \ \forall t \geq 0, i] = 1$ and therefore

$$P[X_t \in \overline{D} \ \forall t \geq 0] = 1. \tag{4.3}$$

By Whitney's extension theorem ([11], Corollary 6.3 in the appendix) it now follows that $\mathcal{C}^2(\overline{D}) = \{f|_{\overline{D}} : f \in \mathcal{C}_c^2(\mathbb{R}^d)\}$, and therefore X solves the martingale problem for $A = \{(f, \mathcal{A}f) : f \in \mathcal{C}^2(\overline{D})\}$. ∎

4.2 Strong uniqueness: Proof of Theorem 1.9

For notational simplicity we only consider the case $c = 1$ and $\theta = 0$. Our first aim is to prove (1.21), i.e., we show that the time needed for X_t to reach the boundary ∂D is infinite (Lemma 4.5). For this we construct (in Lemmas 4.3 and 4.4) a function h on D such that $\mathcal{A}h \leq 1$, where $\mathcal{A}$ is the differential form in (2.4), i.e.,

$$(\mathcal{A}f)(x) = \Big(- x \cdot \nabla + g(x)\Delta \Big) f(x). \tag{4.4}$$

With the help of a radial function (Lemma 4.2) the problem is reduced to a one-dimensional problem (Lemma 4.1).

Lemma 4.1 *Let $a, b \in \mathcal{C}[0,1]$ and $a > 0$ on $(0,1]$. Then there exists a unique function $f \in \mathcal{C}^2(0,1]$ such that*

$$\begin{aligned} &f(1) = f'(1) = 0 \\ &b(r)f'(r) + a(r)f''(r) = 1. \end{aligned} \tag{4.5}$$

For all $r \in (0,1)$ this function satisfies

$$\begin{aligned} &f(r) > 0 \\ &f'(r) < 0. \end{aligned} \tag{4.6}$$

Furthermore, if

$$\limsup_{r \to 0} \frac{a(r)}{r} < b(0), \tag{4.7}$$

then

$$\lim_{r \to 0} f(r) = \infty. \tag{4.8}$$

Proof of Lemma 4.1: Let $u \in \mathcal{C}^2(0,1]$ be the unique solution of

$$\begin{aligned} &u(1) = 0 \\ &u'(1) = -1 \\ &b(r)u'(r) + a(r)u''(r) = 0, \end{aligned} \tag{4.9}$$

i.e.,

$$\begin{aligned} u(r) &= -\int_1^r dx \exp(-\int_1^x dy \tfrac{b(y)}{a(y)}) \\ u'(r) &= -\exp(-\int_1^r dx \tfrac{b(x)}{a(x)}) \\ u''(r) &= \tfrac{b(r)}{a(r)} \exp(-\int_1^r dx \tfrac{b(x)}{a(x)}). \end{aligned} \tag{4.10}$$

Note that $u(r) \geq 0$ and $u'(r) < 0$ for all $r \in (0,1]$. From the latter property it follows that u is invertible. Let $u(0) := \lim_{r \to 0} u(r)$ (which is allowed to be ∞). There exists a continuous function $v : [0, u(0)) \to (0, \infty)$ such that

$$v(u(r)) = a(r)(u'(r))^2. \tag{4.11}$$

Let $h \in \mathcal{C}[0, u(0))$ be the unique solution of

$$\begin{aligned} &h(0) = h'(0) = 0 \\ &v(u)h''(u) = 1 \qquad (u \in [0, u(0))), \end{aligned} \tag{4.12}$$

i.e.,

$$h(u) = \int_0^u dp \int_0^p dq \, \frac{1}{v(q)}. \tag{4.13}$$

Note that $h(u) > 0$ and $h'(u) > 0$ for all $u \in (0, u(0))$. We now define $f \in \mathcal{C}^2(0,1]$ by

$$f(r) := h(u(r)). \tag{4.14}$$

It follows that

$$\begin{aligned}
\tfrac{\partial}{\partial r} h(u(r)) &= h'(u(r))u'(r) \\
\tfrac{\partial^2}{\partial r^2} h(u(r)) &= h''(u(r))(u'(r))^2 + h'(u(r))u''(r) \\
b(r)f'(r) + a(r)f''(r) &= \Big(b(r)u'(r) + a(r)u''(r)\Big)h'(u(r)) \\
&\quad + a(r)(u'(r))^2 h''(u(r)) \\
&= v(u(r))h''(u(r)) = 1.
\end{aligned} \tag{4.15}$$

We see that f constructed above is the unique solution of (4.5), and that f satisfies (4.6).

It is left to show that, under the conditions mentioned, f diverges as $r \to 0$. Let L be such that $\limsup_{r\to 0} r^{-1}a(r) < L < b(0)$. It follows that there exists an $\varepsilon > 0$ such that $b(r) > L$ and $a(r) < Lr$ for all $r \in [0, \varepsilon]$. Let $\tilde{f} \in \mathcal{C}(0, \varepsilon]$ be the unique solution of

$$\begin{aligned}
&\tilde{f}(\varepsilon) = 0 \\
&\tilde{f}'(\varepsilon) = 0 \\
&L\tilde{f}'(r) + Lr\tilde{f}''(r) = 1,
\end{aligned} \tag{4.16}$$

i.e.,

$$\begin{aligned}
&(Lr\tilde{f}'(r))' = 1 \\
&Lr\tilde{f}'(r) = r - \varepsilon \\
&\tilde{f}'(r) = \tfrac{1}{L}(1 - \tfrac{\varepsilon}{r}) \\
&\tilde{f}(r) = \tfrac{1}{L}r - \varepsilon \log(r) - \varepsilon + \varepsilon \log(\varepsilon).
\end{aligned} \tag{4.17}$$

It is clear that $\tilde{f}(r) \to \infty$ as $r \to 0$. Furthermore,

$$b(r)\tilde{f}'(r) + a(r)\tilde{f}''(r) < 1. \tag{4.18}$$

On $(0, \varepsilon]$ define $h := f - \tilde{f}$. Then, using (4.6), we get

$$\begin{aligned}
&h(\varepsilon) > 0 \\
&h'(\varepsilon) < 0 \\
&b(r)h'(r) + a(r)h''(r) > 0.
\end{aligned} \tag{4.19}$$

It follows that $h(r) > 0$ for all $r \in (0, \varepsilon]$: if we assume the converse, then h must assume a positive maximum in a point $0 < r < \varepsilon$, which is impossible by (4.19). We thus see that $f > \tilde{f}$, and therefore $f(r) \to \infty$ if $r \to 0$. ∎

Lemma 4.2 *Let D be regular and $0 \in D$. Then there exist a function $r \in \mathcal{C}^2(\overline{D})$ and a constant $K \in (0, \infty)$ with the following properties:*

$$\begin{aligned}
&0 < r(x) \le 1 && (x \in D) \\
&r(x) = 0 && (x \in \partial D) \\
&-x \cdot \nabla r(x) = K && (x \in \partial D).
\end{aligned} \tag{4.20}$$

Proof of Lemma 4.2: Recall the definition of a regular set in section 1.4 and the function m associated with it. The function $x \mapsto x \cdot n(x) = x \cdot \nabla m(x)$ is $\mathcal{C}^2$ and strictly positive on ∂D, so we can find a strictly positive function $\phi \in \mathcal{C}^2(\overline{D})$ such that in an open neighbourhood of ∂D:

$$\phi(x) x \cdot n(x) = 1. \tag{4.21}$$

Define

$$r(x) := -\phi(x) m(x) \qquad (x \in D). \tag{4.22}$$

Then $r \in \mathcal{C}^2(\overline{D})$ and, for all $x \in \partial D$, $\nabla r(x)$ is parallel to $n(x)$ and satisfies $-x \cdot \nabla r(x) = \phi(x) x \cdot \nabla m(x) = 1$. We can multiply r with a constant to get $r \leq 1$. ∎

Lemma 4.3 *Let $D' \supset D$ be regular. For $x \in \partial D'$, let $n(x)$ be the normal to D' in x. For $x \in \overline{D}$, let*

$$l(x) := \inf\{|x - y| : y \in \partial D\}. \tag{4.23}$$

Assume that, for all $x_n \in D$, $x \in \partial D' \cap \partial D$ with $x_n \to x$ as $n \to \infty$,

$$\limsup_{n \to \infty} \frac{g(x_n)}{l(x_n)} < x \cdot n(x). \tag{4.24}$$

Then there exists a function $h \in \mathcal{C}(D')$ such that

$$\begin{array}{ll} 0 \leq h(x) & (x \in D') \\ \lim_{n\to\infty} h(x_n) = \infty & (x_n \to x \in \partial D') \\ (\mathcal{A}h)(x) \leq 1 & (x \in D). \end{array} \tag{4.25}$$

Proof of Lemma 4.3: Extend g by putting $g \equiv 0$ on $D' \backslash D$, so that (4.24) holds for all $x_n \in D'$ with $x_n \to x \in \partial D'$. Let r be as in Lemma 4.2. The idea will be to find a function $f \in \mathcal{C}^2(0, 1]$ such that

$$h(x) := f(r(x)) \tag{4.26}$$

satisfies (4.25).

For any $f \in \mathcal{C}^2(0, 1]$ we have

$$\begin{aligned} \nabla f(r(x)) &= f'(r(x))(\nabla r)(x) \\ \Delta f(r(x)) &= f''(r(x))(\nabla r)(x) \cdot (\nabla r)(x) + f'(r(x))(\Delta r)(x) \\ \mathcal{A} f(r(x)) &= \Big(- x \cdot (\nabla r)(x) + g(x)(\Delta r)(x) \Big) f'(r(x)) \\ &\quad + \Big(g(x) |\nabla r)(x)|^2 \Big) f''(r(x)), \end{aligned} \tag{4.27}$$

where the first two formulas follow from

$$\begin{array}{l} \frac{\partial}{\partial x_i} f(r(x)) = f'(r(x)) \frac{\partial}{\partial x_i} r(x) \\ \sum_i \frac{\partial^2}{\partial {x_i}^2} f(r(x)) = \sum_i \frac{\partial}{\partial x_i} f'(r(x)) \frac{\partial}{\partial x_i} r(x) \\ = \sum_i (f''(r(x)) (\frac{\partial}{\partial x_i} r(x)) (\frac{\partial}{\partial x_i} r(x)) + f'(r(x)) \frac{\partial^2}{\partial {x_i}^2} r(x)). \end{array} \tag{4.28}$$

We want estimates on the two terms in the formula for $\mathcal{A}f(r(x))$. To that aim, we define functions $a, b \in \mathcal{C}[0,1]$ by

$$\begin{array}{l} a(z) := \max\{g(x)|\nabla r(x)|^2 : r(x) = z\} \\ b(z) := \min\{-x \cdot \nabla r(x) + g(x)\Delta r(x) : r(x) = z\}, \end{array} \tag{4.29}$$

We have

$$\begin{array}{l} b(0) = K \\ \displaystyle\limsup_{z\to 0} \frac{a(z)}{z} < K. \end{array} \tag{4.30}$$

Indeed, the first equation is trivial. For the second one, note that

$$\frac{a(z)}{z} = \max\left\{\frac{g(x)|\nabla r(x)|^2}{r(x)} : r(x) = z\right\}, \tag{4.31}$$

where, by (4.24), for each $x_n \in D'$ with $x_n \to x \in \partial D'$

$$\begin{array}{rcl} \displaystyle\limsup_{n\to\infty} \frac{g(x_n)|\nabla r(x_n)|^2}{r(x_n)} & = & \displaystyle\left(\limsup_{n\to\infty} \frac{g(x_n)}{l(x_n)}\right)\left(\lim_{n\to\infty} \frac{l(x_n)}{r(x_n)}|\nabla r(x_n)|^2\right) \\ & < & (x \cdot n(x))|\nabla r(x)| = -x \cdot \nabla r(x) = K. \end{array} \tag{4.32}$$

Here, the last two equalities follow from Lemma 4.2. Using compactness, we arrive at (4.30).

We have thus found functions $a, b \in \mathcal{C}[0,1]$ such that

$$\begin{array}{rcl} -x \cdot \nabla r(x) + g(x)\Delta r(x) & \geq & b(r(x)) \\ g(x)|\nabla r(x)|^2 & \leq & a(r(x)) \\ \limsup_{z\to 0} \frac{a(z)}{z} & < & b(0). \end{array} \tag{4.33}$$

We can change a such that $a > 0$ on $(0,1]$ while (4.33) continues to holds. Applying Lemma 4.1 to these functions a and b, we find a function f satisfying (4.5), (4.6) and (4.8). Since $b(0) > 0$, we see that there exists an $\varepsilon > 0$ such that

$$f''(z) > 0 \qquad (z \in (0, \varepsilon)). \tag{4.34}$$

Combining this with (4.6), (4.27) and (4.33), we see that

$$(\mathcal{A}h)(x) \leq 1 \qquad (x \in D, r(x) < \varepsilon), \tag{4.35}$$

where $h(x) := f(r(x))$ as in (4.26). But $x \mapsto (\mathcal{A}h)(x)$ is continuous on the compact set $\{x \in D : r(x) \geq \varepsilon\}$, so multiplying h by a constant we arrive at a function satisfying (4.25). ∎

Lemma 4.4 *Let D be a finite intersection of regular sets. Assume that for all $x \in \partial D$, all $x_n \in D$ with $x_n \to x$, and each normal $n(x)$ to D in x:*

$$\limsup_{n\to\infty} \frac{g(x_n)}{|x - x_n|} < x \cdot n(x). \tag{4.36}$$

Then there exists a function $h \in \mathcal{C}^2(D)$ such that

$$\begin{array}{l} 0 \leq h(x) \\ (\mathcal{A}h)(x) \leq 1 \end{array} \tag{4.37}$$

and such that $h(x_n) \to \infty$ for all $x_n \to x \in \partial D$.

Proof of Lemma 4.4: Let $D = \bigcap_{i=1}^n D_i$, where the D_i are regular. For each D_i the assumptions in Lemma 4.3 are satisfied. In particular, (4.36) implies (4.24). Let $h_i \in \mathcal{C}^2(D_i)$ be the function constructed in Lemma 4.3. Then $h = \frac{1}{n}\sum_{i=1}^n h_i$ satisfies our requirements. ∎

Lemma 4.5 *Let D and g be as in Lemma 4.4, and let $(X_t^x)_{t\geq 0}$ be a solution to the martingale problem for A with $X_0^x = x \in D$. Then*

$$P[X_t^x \in D \quad \forall t \geq 0] = 1. \tag{4.38}$$

Proof of Lemma 4.5: Let h be the function mentioned in Lemma 4.3. For $H < \infty$ we introduce a stopping time τ_H by

$$\tau_H := \inf\{t \geq 0 : h(X_t^x) = H\}. \tag{4.39}$$

We can extend h outside $\{x \in \overline{D} : h(x) \leq H\}$ to a function in $\mathcal{C}^2(\overline{D})$. From the martingale problem we get

$$E[h(X_{t\wedge\tau_H})] = h(x) + E[\int_0^{t\wedge\tau_H} (\mathcal{A}h)(X_s)ds] \leq h(x) + E[t \wedge \tau_H]. \tag{4.40}$$

Here $E[h(X_{t\wedge\tau_H})] \geq HP[\tau_H \leq t]$ and $E[t\wedge\tau_H] \leq t$, so

$$P[\tau_H \leq t] \leq \frac{h(x)+t}{H}. \tag{4.41}$$

Therefore

$$P[h(X_s^x) < H \quad \forall 0 \leq s \leq t] \geq 1 - \frac{h(x)+t}{H}. \tag{4.42}$$

Letting $H \uparrow \infty$ so that $\{x \in \overline{D} : h(x) < H\} \uparrow D$, we find that

$$P[X_s^x \in D \quad \forall 0 \leq s \leq t] = 1. \tag{4.43}$$

Letting $t \uparrow \infty$ we obtain Lemma 4.5. ∎

Proof of Theorem 1.9: Define $D_n := \{x \in D : g(x) > \frac{1}{n}\}$. Note that if g is Lipschitz on $\overline{D}_n$ with constant L_n, then $\sqrt{g}$ is Lipschitz on $\overline{D}_n$ with constant nL_n. Let $(X_t)_{t\geq 0}$ and $(\tilde{X}_t)_{t\geq 0}$ be solutions of (1.7) with $X_0 = \tilde{X}_0$, adapted to the same Brownian motion. Define stopping times

$$\tau_n := \inf\{t \geq 0 : X_t \in \partial D_n\}, \tag{4.44}$$

and define $\tilde{\tau}_n$ similarly for $(\tilde{X}_t)_{t\geq 0}$. Now follow the proof of Theorem 5.2.5 in [14], to see that the processes X and $\tilde{X}$ are indistinguishable up to time $\tau_n \wedge \tau_n$, where by Lemma 4.5, $\tau_n \wedge \tilde{\tau}_n \uparrow \infty$ as $n \uparrow \infty$. ∎

4.3 Weak uniqueness: Proof of Theorem 1.10

Writing

$$\begin{aligned} -\tfrac{1}{2}\Delta\tfrac{1}{d}(1-|x|^2) &= -\tfrac{1}{2}\tfrac{1}{d}\textstyle\sum_{i=1}^d \frac{\partial^2}{\partial {x_i}^2}(1-\sum_{j=1}^d x_j^2) \\ &= \tfrac{1}{2d}\textstyle\sum_{i=1}^d \frac{\partial}{\partial x_i}(2\sum_{j=1}^d \delta_{ij}x_j) = \frac{1}{2d}\sum_{i=1}^d 2 = 1, \end{aligned} \tag{4.45}$$

we see that $g^*(x) = \frac{1}{d}(1-|x|^2)$ as claimed. We introduce polynomials on $\overline{D}$ in the usual way, namely we define the set of all multi-indices α by

$$\begin{aligned} A &:= \{\alpha \in \mathbb{Z}^d : \alpha_i \geq 0 \ \forall i = 1,\ldots,d\} \\ |\alpha| &:= \textstyle\sum_{i=1}^d \alpha_i. \end{aligned} \tag{4.46}$$

and on $\overline{D}$ we define functions $x \mapsto x^\alpha$ and a space of polynomials of order $\leq n$ by

$$\begin{aligned} x^\alpha &:= \textstyle\prod_{i=1}^d x_i^{\alpha_i} \\ P_n &:= \operatorname{span}\{x^\alpha : |\alpha \leq n\}. \end{aligned} \tag{4.47}$$

Setting $g = rg^*$, $r > 0$, we observe that

$$Ax^\alpha = \Big(c\sum_{i=1}^d (\theta_i - x_i)\tfrac{\partial}{\partial x_i} + \tfrac{r}{d}(1-\sum_{j=1}^d x_j^2)\sum_{i=1}^d \tfrac{\partial^2}{\partial {x_i}^2}\Big)x^\alpha. \tag{4.48}$$

Since $\frac{\partial}{\partial x_i}x^\alpha \in P_{|\alpha|-1}$ and $\frac{\partial^2}{\partial {x_i}^2}x^\alpha \in P_{|\alpha|-2}$, we have

$$Ax^\alpha \in P_{|\alpha|} \qquad \forall \alpha \in A. \tag{4.49}$$

The spaces P_n are finite-dimensional and closed under A, and their union $\bigcup_n P_n$ is dense in $\mathcal{C}(\overline{D})$. Applying [11], Proposition 1.3.5, we see that A is closable and that its closure generates a Feller semigroup on $\mathcal{C}(\overline{D})$. This implies that the martingale problem is well-posed for A ([11], Theorem 4.4.1), and hence $(rg^*)_{r>0} \subset \mathcal{H}'$. But $F_c rg^* = (cr)/(c+r)g^*$, so the family $(rg^*)_{r>0}$ is closed under F_c for all $c \in (0,\infty)$. This implies that $rg^* \in \mathcal{H}''$ for all $r > 0$. ∎

Acknowledgement

We thank Ph. Clément, T. ter Elst, S.N. Ethier, A. Greven, E. Perkins and T. Shiga for valuable comments while this paper was in progress.

References

[1] J.-B. Baillon, Ph. Clément, A. Greven and F. den Hollander, On the attracting orbit of a non-linear transformation arising from renormalization of hierarchically interacting diffusions. Part I: The compact case. Can. J. Math. 47 (1995) 3–27.

[2] J.-B. Baillon, Ph. Clément, A. Greven and F. den Hollander, On the attracting orbit of a non-linear transformation arising from renormalization of hierarchically interacting diffusions. Part II: The non-compact case. J. Funct. Anal. 146-I (1997) 236–298.

[3] J.T. Cox and A. Greven, Ergodic theorems for infinite systems of locally interacting diffusions. Ann. Probab. 22-II (1994) 833–853.

[4] E.B. Dynkin, *Markov Processes*, Die Grundlehren der mathematischen Wissenschaften, Springer, Berlin, 1965.

[5] D.A. Dawson and A. Greven, Multiple time scale analysis of hierarchically interacting systems. In: *Stochastic Processes*, A Festschrift in Honour of Gopinath Kallianpur (eds. S. Cambanis, J.K. Ghosh, R.L. Karandikar and P.K. Sen), Springer, New York, 1993, pp. 41–50.

[6] D.A. Dawson and A. Greven, Multiple time scale analysis of interacting diffusions. Probab. Theory Relat. Fields 95 (1993) 467–508.

[7] D.A. Dawson and A. Greven, Hierarchical models of interacting diffusions: multiple time scale phenomena, phase transition and pattern of cluster-formation. Probab. Theory Relat. Fields 96 (1993) 435–473.

[8] D.A. Dawson and A. Greven, Multiple space-time scale analysis for interacting branching models. Electronic J. of Prob. 1 (1996), paper no. 14, pp. 1–84 (http://math.washington.edu/∼ejpecp/).

[9] D.A. Dawson, A. Greven and J. Vaillancourt, Equilibria and quasi-equilibria for infinite collections of interacting Fleming-Viot processes. Trans. Am. Math. Soc. 347-VII (1995) 2277–2360.

[10] D.A. Dawson and P. March, Resolvent estimates for Fleming-Viot operators and uniqueness of solutions to related martingale problems. J. Funct. Anal. 132-II (1995) 417–472.

[11] S.N. Ethier and T.G. Kurtz, *Markov Processes; characterization and convergence*, John Wiley & Sons, New York, 1986.

[12] A. Friedman, *Partial Differential Equations of Parabolic Type*, Prentice-Hall, Englewood Cliffs, 1964.

[13] A. Klenke, Different clustering regimes in systems of hierarchically interacting diffusions. Ann. Probab. 24-II (1996) 660–697.

[14] I. Karatzas and S.E. Shreve, *Brownian Motion and Stochastic Calculus*, second edition, Springer, New York, 1991.

[15] M.H. Protter and H.F. Weinberger, *Maximum Principles in Differential Equations*, Prentice-Hall, Englewood Cliffs, 1967.

[16] M. Reed and B. Simon, *Methods of Modern Mathematical Physics, I: Functional Analysis* (revised and enlarged edition) Academic Press, San Diego, 1980.

[17] S. Sawyer and J. Felsenstein, Isolation by distance in a hierarchically clustered population. J. Appl. Probab. 20 (1983) 1–10.

[18] T. Shiga, Diffusion processes in population genetics. J. Math. Kyoto Univ. 21 (1981) 133–151.

[19] T. Shiga and A. Shimizu, Infinite dimensional stochastic differential equations and their applications. J. Math. Kyoto Univ. 20 (1980) 395–416.

[20] D.W. Stroock and S.R.S. Varadhan, *Multidimensional Diffusion Processes*, Springer, Berlin, 1979.

Mathematical Institute
University of Nijmegen
Toernooiveld 1
6525 ED Nijmegen
The Netherlands