The Fleming-Viot limit of an interacting spatial population with fast density regulation *

Ankit Gupta[†]

Abstract

We consider population models in which the individuals reproduce, die and also migrate in space. The population size scales according to some parameter N, which can have different interpretations depending on the context. Each individual is assigned a mass of $1/N$ and the total mass in the system is called *population density*. The dynamics has an intrinsic density regulation mechanism that drives the population density towards an equilibrium. We show that under a timescale separation between the *slow* migration mechanism and the *fast* density regulation mechanism, the population dynamics converges to a Fleming-Viot process as the scaling parameter N approaches ∞. We first prove this result for a basic model in which the birth and death rates can only depend on the population density. In this case we obtain a *neutral* Fleming-Viot process. We then extend this model by including position-dependence in the birth and death rates, as well as, offspring dispersal and immigration mechanisms. We show how these extensions add *mutation* and *selection* to the limiting Fleming-Viot process. All the results are proved in a multi-type setting, where there are q types of individuals interacting with each other. We illustrate the usefulness of our convergence result by discussing applications in population genetics and cell biology.

1 Introduction

Density-dependent models are well-known in population biology. In these models, the birth and death rates of individuals may depend on the *density* of the population, where the term density refers to the population size under a suitably chosen normalization. Many models in ecology, epidemiology and immunology can be suitably described by such models (see Thieme [35]). Considering the molecules of chemical species as

[†]Centre de Mathématiquees Appliquées, École Polytechnique, France.
E-mail: gupta@cmap.polytechnique.fr

the individuals in a population, we can also view a chemical reaction network as a density-dependent population model.

Density-dependent models are appealing because one can easily account for interactions among individuals by appropriately specifying the birth and death rates as functions of the population density. For competitive interactions, as in the Lotka-Volterra model (see [28, 38, 30]), the death rate increases with the population density while for cooperative interactions, as in the Allee model (see [1]), the birth rate increases with the population density. Density dependent models are good candidates for modeling natural populations that cannot grow indefinitely due to the limited availability of certain vital resources or due to severe competition at large population sizes. By having the death rate dominate the birth rate at *large* densities, one can ensure that the population density does not go beyond a certain threshold.

For a population having q types of individuals, the population density is a q dimensional vector whose i-th component gives the density of the population of the i-th type. For such a multi-type population, a density-dependent model can be written in the deterministic setting as a system of q ordinary differential equations. If all the trajectories of this system stay within a compact set at all times, then we say that the population dynamics has a *density regulation mechanism*. Such a mechanism is called *equilibrating* if all the trajectories reach a fixed point for this system as time goes to infinity. In such a situation, this fixed point is called the *equilibrium population density*.

In this paper, we will consider population models in which the individuals live in a *geographical region E*, that is a compact metric space. Even though we have a spatial structure, for us the population density will always denote the population size divided by a normalization parameter N. In other words, our notion of population density is *global* in the sense that it carries no information about the distribution of individuals in E. This is unlike other models of spatial populations where the population density is a spatially varying function specifying the local concentration of individuals at each location. The normalization parameter N will be a large positive integer which can have various interpretations depending on the context. In ecological models, N can be taken to be the *carrying capacity* of a habitat, which is the maximum number of individuals that the habitat can support with its resources. In epidemic models, N is usually the total population size, while in chemical reaction networks, N measures the *volume* of the system. In each of these cases, the population size at any time is of order N.

For the moment assume that all the individuals have the same type and that they reproduce, die and also migrate in E. At the time of birth, the offspring gets the same location as its parent. The population consists of approximately N individuals of a *mass* of $1/N$ each. The population density at any time is just the total mass of the individuals that are alive. Suppose that the birth and death rates of the individuals depend on the population density in such a way that they induce a density regulation mechanism which is equilibrating. We also assume that the migration mechanism operates at a timescale that is N times *slower* than the density regulation mechanism. In such a setting, we can view the dynamics of the empirical measure of the population as a measure-valued Markov process parameterized by N. Our goal is to understand how this family of Markov processes behaves as $N \to \infty$. The population dynamics has two timescales separated by N. If we observe the process at the *fast* timescale, then the effect of migration vanishes in the limit and it is uninteresting to consider the population with a spatial structure in this case. Therefore we will observe the dynamics at the *slow* timescale and examine its behaviour in the infinite population limit. Since the density regulation mechanism is *fast*, it will have enough time to re-equilibrate the population density between any two events at the slow timescale. Hence in the limit $N \to \infty$, we would expect the population density to remain equilibrated at all times. We will show

that it is indeed the case. However our main task is to understand the dynamics of the spatial distribution of the population in the infinite population limit. We will prove that in the limit, the spatial distribution of the population evolves according to a Fleming-Viot process which takes values in the space of probability measures over E. This process was introduced in the context of population genetics by Fleming and Viot [17] in 1979 and it has been very well-studied since then. An excellent survey of Fleming-Viot processes is given by Ethier and Kurtz [13]. The model we just described will be called the *basic model*. In this model, the birth and death rates of an individual were density dependent, but independent of the location of the individual. It shall be seen later that the limiting Fleming-Viot process in this case is *neutral*, in the parlance of population genetics. If we add *small* position-dependent terms to the birth and death rates, then in the limit we obtain a Fleming-Viot process with *genetic selection*. Perhaps unsurprisingly, altering the birth rate this way leads to *fecundity selection*, while altering the death rate leads to *viability selection*, in the limiting Fleming-Viot process. We also consider extensions of the basic model by allowing for *offspring dispersal* (offspring is born away from the parent) or *immigration*. Such extensions add extra *mutations* to the limiting Fleming-Viot process.

The results mentioned in the previous paragraph are proved in a multi-type setting. The population has q types of individuals and each type of individual can give birth to an individual of each type. All the individuals are migrating in E according to a type-dependent mechanism. Now we can view the joint dynamics of the empirical measures of the q sub-populations as a Markov process parameterized by N. We make similar assumptions on the dynamics as before. Again in the limit $N \to \infty$, the population density (which is now a q-dimensional vector) stays at an equilibrium at all times. Assuming the irreducibility of an underlying interaction matrix, we show that in the limit all the q sub-populations become *spatially inseparable*. This means that on any patch of E, either there is no mass present or there is mass of each type present in a proportion determined by the equilibrium density. Moreover the spatial distribution of each of the q sub-populations evolves according to a single Fleming-Viot process. This Fleming-Viot process can be seen as describing the limiting dynamics of a *mixed* population, formed by taking a suitable density-dependent convex combination of the q sub-populations.

In ecological models, the individuals need resources to survive and reproduce. Normally in spatial population models, resources are assumed to have a fixed distribution in space. As individuals move, they find the unexploited resources and compete for them *locally* with other individuals present in their neighbourhood. Such a model is different from the models we consider in two ways. Firstly, due to the local nature of the interactions, the density is locally regulated rather than globally regulated as in our models. Secondly, since the discovery of resources is tied to the movement of individuals, it is reasonable to assume that both migration and birth-death mechanisms operate at the same timescale. For such spatial models, Oelschläger [31] has shown in a multi-type setting that the dynamics converges in the infinite population limit to a system of reaction-diffusion partial differential equations. Such equations are in widespread use in biology (see Fife [16]). We now discuss the conditions under which our models can be useful. Consider a situation where the resource is not fixed but rapidly mixing in the whole space. This resource is shared by all the individuals in the population. An individual may deplete the resource locally but its effect is felt globally due to the rapid mixing. This gives rise to global density dependence in a spatial population. If the individuals move very slowly in comparison to their resource consumption mechanism (which is linked to their birth and death mechanisms), then we have a situation in which our models can be used.

This paper is motivated by our earlier work [19] in which we study the phenomenon

of cell polarity using a model considered here. Cell polarity refers to the clustering of molecules on the cell membrane. This clustering is essential to trigger various other cellular processes, such as bud formation [4] or immune response [39]. Therefore understanding how cells establish and maintain polarity is of vital importance. In [2], Altschuler et. al. devised a mathematical model for this phenomenon, by abstracting the mechanisms that are commonly found in cells exhibiting polarity. Their model has a fixed number of molecules that can either reside on the membrane or in the cytosol. These molecules move slowly on the membrane but diffuse rapidly in the cytosol. The dynamics has a *positive feedback* mechanism which allows a membrane molecule to pull a cytosol molecule to its location on the membrane. This mechanism is like a birth process in which a membrane molecule gives birth by exploiting the common resource (cytosol molecules) shared by all the membrane molecules. Since the migration of membrane molecules is *slow* and the mixing of the resource is *fast*, this model can be viewed as a model described in this paper (see Section 3.2 for details). Therefore the results in this paper are applicable and we obtain a Fleming-Viot process in the infinite population limit. In [19] we prove this convergence[1] and use the limiting process to answer some interesting questions about the onset and structure of cell polarity. The model studied in [19] is rather simplistic as all the molecules are assumed to be identical. Most cells that exhibit polarity have molecules of many different types participating in the feedback mechanism and migrating on the membrane in different ways (see [10, 4, 34]). It is natural to ask if the Fleming-Viot convergence is valid in this general framework. The results in this paper show that it is indeed the case as long as certain basic elements of the dynamics are preserved. This ensures that the analysis in [19] can be extended to more complicated (and realistic !) models for cell polarity. We discuss this example further in Section 3.2.

Note that the geographical space E can be considered as the space of *genetic traits*. This casts our models into the setting of population genetics. The spatial migration can be seen as *mutation* that may happen at any time during the life of an individual, while the offspring dispersal mechanism is like *mutation* that can only happen at the time of birth of an offspring. We assume that the reproduction is *clonal* in the absence of mutation. The position-dependent birth and death mechanism is analogous to the *selection* mechanism in population genetics. Hence it is not surprising that spatial migration, offspring dispersal and position-dependent birth and death mechanisms correspond to mutation and selection in the limiting Fleming-Viot process. What is more interesting is that the *sampling mechanism* arises naturally from our models in the infinite population limit. This sampling mechanism is a key feature of the standard models in population genetics, such as the Wright-Fisher model, the Moran model and their variants (see [40, 29, 14]). This mechanism makes the models tractable by keeping the population size constant. It is done by matching the birth of an individual with the death of another individual chosen uniformly from the population. It is obvious that such a mechanism is quite unrealistic, at least for finite populations which are naturally fluctuating. However our Fleming-Viot convergence result shows that one can recover this sampling mechanism in the infinite population limit if the dynamics has an equilibrating density regulation mechanism that acts at a faster timescale than other events. It is well-known that a Fleming-Viot process arises in the infinite population limit of an appropriately scaled version of the Wright-Fisher or the Moran model (see [17] and [13]). Therefore if all the individuals have the same type (that is, $q = 1$) and h_{eq} is the equilibrium population density, then for a large (but finite) value of the scaling parameter N, our models

---

[1]This convergence was proved in [19] using the technique of particle representation described in [9]. This technique cannot be easily extended to the multi-type setting of this paper. Therefore the convergence proof in this paper is vastly different.

will have *roughly* the same dynamical behaviour as a suitably chosen Wright-Fisher or Moran model with the constant population size Nh_{eq}. This insight provides a justification for the assumption of a constant population size in population genetics models. Most of the mathematical literature on population genetics is concerned with two types of questions. In the absence of mutation, one wants to know the probability and the time of *fixation* of a particular genetic trait. The term *fixation* describes the event in which the whole population has the same genetic trait. In the presence of mutation, one attempts to investigate the properties of the stationary distribution, if such a distribution is present. These questions are difficult to answer for finite populations and one typically answers them by studying the limiting Fleming-Viot process. Our discussion shows that for large N, the fixation times and probabilities or the stationary distribution will be approximately the same for our model and the corresponding Wright-Fisher or Moran model. We illustrate this point through an example in Section 3.1.

This paper is organized as follows. In Section 2 we describe the mathematical models that we consider and state our main results. In Section 3 we discuss the aforementioned applications of our results in greater detail. Finally in Section 4 we prove the main results.

Notation

We now introduce some notation that we will use throughout this paper. Let $\mathbb{R}$, $\mathbb{R}_+$, $\mathbb{R}_*$, $\mathbb{N}$ and $\mathbb{N}_0$ denote the sets of all reals, nonnegative reals, positive reals, positive integers and nonnegative integers respectively. For any $a, b \in \mathbb{R}$, their minimum is given by $a \wedge b$.

Let $\|\cdot\|$ and $\langle \cdot, \cdot \rangle$ denote the standard Euclidean norm and inner product in $\mathbb{R}^n$ for any $n \in \mathbb{N}$. Moreover for any $v = (v_1, \ldots, v_n) \in \mathbb{R}^n$, the norms $\|v\|_1$ and $\|v\|_\infty$ are defined as $\|v\|_1 = \sum_{i=1}^n |v_i|$ and $\|v\|_\infty = \max_{1 \le i \le n} |v_i|$. The vectors of all zeros and all ones in $\mathbb{R}^n$ are denoted by $\bar{0}_n$ and $\bar{1}_n$ respectively. Let $\mathbb{M}(n, n)$ be the space of all $n \times n$ matrices with real entries. For any $M \in \mathbb{M}(n, n)$, the entry at the i-th row and the j-th column is indicated by M_{ij}. Its infinity norm is defined as $\|M\|_\infty = \max_{1 \le i \le n} \sum_{j=1}^n |M_{ij}|$ and its transpose and inverse are indicated by M^T and M^{-1} respectively. The symbol I_n refers to the identity matrix in $\mathbb{M}(n, n)$. For any $v = (v_1, \ldots, v_n) \in \mathbb{R}^n$, $\text{Diag}(v)$ refers to the matrix in $\mathbb{M}(n, n)$ whose non-diagonal entries are all 0 and whose diagonal entries are $v_1, \ldots, v_n$. A matrix in $\mathbb{M}(n, n)$ is called *stable* if all its eigenvalues have strictly negative real parts. While multiplying a matrix with a vector we always regard the vector as a column vector.

Let $U \subset \mathbb{R}^n$ and $V \subset \mathbb{R}^m$. Then for any $k \in \mathbb{N}_0$, the class $C^k(U, V)$ refers to the set of all those functions f that are defined on some open set $O \subset \mathbb{R}^n$ containing U such that $f(x) \in V$ for all $x \in U$ and f is k-times continuously differentiable at any $x \in O$.

Let (S, d) be a metric space. Then by $B(S) \, (C(S))$ we refer to the set of all bounded (continuous) real-valued Borel measurable functions. If S is compact, then $C(S) \subset B(S)$ and both $B(S)$ and $C(S)$ are Banach spaces under the sup norm $\|f\|_\infty = \sup_{x \in S} |f(x)|$. Recall that a class of functions in $B(S)$ is called an *algebra* if it is closed under finite sums and products. Let $\mathcal{B}(S)$ be the Borel sigma field on S. By $\mathcal{M}_F(S)$ and $\mathcal{P}(S)$ we denote the space of all finite positive Borel measures and the space of all Borel probability measures respectively. These measure spaces are equipped with the weak topology. For any $f \in B(S)$ and $\mu \in \mathcal{M}_F(S)$ let

$$\langle f, \mu \rangle = \int_E f(x) \mu(dx).$$

The space of cadlag functions (that is, right continuous functions with left limits) from $[0, \infty)$ to S is denoted by $D_S[0, \infty)$ and it is endowed with the Skorohod topology

(for details see Chapter 3, Ethier and Kurtz [12]). The space of continuous functions from $[0, \infty)$ to S is denoted by $C_S[0, \infty)$ and it is endowed with the topology of uniform convergence over compact sets. An operator A on $B(S)$ is a linear mapping that maps any function in its domain $\mathcal{D}(A) \subset B(S)$ to a function in $B(S)$. The notion of the *martingale problem* associated to an operator A is introduced and developed in Chapter 4, Ethier and Kurtz [12]. In this paper, by a solution of the martingale problem for A we mean a measurable stochastic process X with paths in $D_S[0, \infty)$ such that for any $f \in \mathcal{D}(A)$,

$$f(X(t)) - \int_0^t Af(X(s))ds$$

is a martingale with respect to the filtration generated by X. For a given initial distribution $\pi \in \mathcal{P}(S)$, a solution X of the martingale problem for A is a solution of the martingale problem for (A, π) if $\pi = \mathbb{P}X(0)^{-1}$. If such a solution X exists uniquely for all $\pi \in \mathcal{P}(S)$, then we say that the martingale problem for A is well-posed. Additionally, we say that A is the generator of the process X.

Throughout the paper $\Rightarrow$ denotes convergence in distribution.

2 Model descriptions and the main result

Our first task is to describe the models that we consider in the paper. As mentioned in Section 1, we model a population which resides in some compact metric space E and in which the individuals have one of q possible types. We denote these types by elements in the set $Q = \{1, 2, \ldots, q\}$. We identify each individual located at $x \in E$ with the Dirac measure δ_x, concentrated at x. Moreover each individual is assigned a *mass* of $1/N$ where $N \in \mathbb{N}$ is our scaling parameter. For any $i \in Q$, the population of type i individuals can be represented by an atomic measure of the form

$$\mu_i = \frac{1}{N} \sum_{j=1}^{n_i} \delta_{x_j^i},$$

where n_i is the total number of type i individuals and $x_1^i, \ldots, x_{n_i}^i \in E$ are their locations. Define the space of atomic measures scaled by N as

$$\mathcal{M}_{N,a}(E) = \left\{ \frac{1}{N} \sum_{j=1}^{n} \delta_{x_j} : n \in \mathbb{N}_0 \text{ and } x_1, \ldots, x_n \in E \right\}. \tag{2.1}$$

Note that $\mathcal{M}_{N,a}(E) \subset \mathcal{M}_F(E)$, where $\mathcal{M}_F(E)$ is the space of finite positive measures. Let $\mathcal{M}_{N,a}^q(E)$ and $\mathcal{M}_F^q(E)$ be the spaces formed by taking products of q copies of $\mathcal{M}_{N,a}(E)$ and $\mathcal{M}_F(E)$ respectively. Since for each $i \in Q$, the type i population can be represented by a measure $\mu_i \in \mathcal{M}_{N,a}(E)$, the entire population can be represented by a q-tuple of measures $\mu = (\mu_1, \ldots, \mu_q) \in \mathcal{M}_{N,a}^q(E)$.

Let 1_E denote the constant function in $C(E)$ which maps each point in E to 1. Define the *density map* $H : \mathcal{M}_F^q(E) \to \mathbb{R}_+^q$ as the continuous function given by

$$H(\mu_1, \ldots, \mu_q) = (\langle 1_E, \mu_1 \rangle, \ldots, \langle 1_E, \mu_q \rangle) \text{ for any } \mu = (\mu_1, \ldots, \mu_q) \in \mathcal{M}_F^q(E). \tag{2.2}$$

We will refer to $h = H(\mu)$ as the density vector corresponding to $\mu \in \mathcal{M}_F^q(E)$. Note that if the population is represented by a $\mu \in \mathcal{M}_{N,a}^q(E)$ and if $h = (h_1, \ldots, h_q)$ is the corresponding density vector, then h_i is just the total number of type i individuals divided by N. The density vector h contains no information about the distribution of individuals on E.

2.1 The type-dependent migration mechanism

In our models, each individual of type $i \in Q$ will migrate according to an independent E-valued Markov process with generator B_i. We will assume that each operator B_i generates a *Feller* semigroup on $C(E)$ (see Chapter 4 in Ethier and Kurtz [12]). Furthermore we assume that there is an algebra of functions $\mathcal{D}_0 \subset C(E)$ which is dense in $C(E)$, contains 1_E and satisfies

$$\mathcal{D}_0 \subset \mathcal{D}(B_i) \text{ for all } i \in Q. \tag{2.3}$$

The martingale problem corresponding to each B_i is well-posed and any solution is a strong Markov process with sample paths in $D_E[0, \infty)$ (see Theorem 4.2.7 and Corollary 4.2.8 in [12]).

We now formally describe how this type-dependent migration of individuals translates into the evolution of our population in the space $\mathcal{M}_{N,a}^q(E)$. For each $n \in \mathbb{N}$, define a space of atomic probability measures as

$$\mathcal{P}_{n,a} = \left\{ \frac{1}{n} \sum_{j=1}^n \delta_{x_j} : x_1, \dots, x_n \in E \right\}$$

and a class of continuous real-valued functions over $\mathcal{P}(E)$ by

$$\mathcal{C}_0 = \left\{ F(\nu) = \prod_{l=1}^m \langle f_l, \nu \rangle \; : \; f_1, \dots, f_m \in \mathcal{D}_0 \text{ and } m \in \mathbb{N} \right\}. \tag{2.4}$$

Suppose that $\nu = (1/n) \sum_{j=1}^n \delta_{x_j} \in \mathcal{P}_{n,a}$ and $F(\nu) = \prod_{l=1}^m \langle f_l, \nu \rangle \in \mathcal{C}_0$. For positive integers $k \leq m$, let P_k^m be the set of onto functions from $\{1, \dots, m\}$ to $\{1, \dots, k\}$ and for any $p \in P_k^m$ and $l = 1, \dots, k$ let

$$f_l^{(p)}(x) = \prod_{j \in p^{-1}(l)} f_j(x). \tag{2.5}$$

Then we can write

$$F(\nu) = n^{-m} \prod_{l=1}^m \left(\sum_{j=1}^n f_l(x_j) \right) = n^{-m} \sum_{i_1, \dots, i_m = 1}^n \prod_{l=1}^m f_l(x_{i_l})$$

$$= n^{-m} \sum_{k=1}^m \sum_{p \in P_k^m} \sum_{j_1, \dots, j_k = 1}^n \prod_{l=1}^k f_l^{(p)}(x_{j_l}), \tag{2.6}$$

where the last term has summation over distinct choices of $j_1, \dots, j_k \in \{1, \dots, n\}$. For each $i \in Q, n \in \mathbb{N}$ we now define an operator $\mathbf{B}_i^n : \mathcal{D}(\mathbf{B}_i^n) = \mathcal{C}_0 \to B(\mathcal{P}_{n,a})$ by

$$\mathbf{B}_i^n F(\nu) = n^{-m} \sum_{k=1}^m \sum_{p \in P_k^m} \sum_{j_1, \dots, j_k = 1}^n \sum_{l=1}^k B_i f_l^{(p)}(x_{j_l}) \prod_{r=1, r \neq l}^k f_r^{(p)}(x_{j_r}), \tag{2.7}$$

where $F \in \mathcal{C}_0$ is given by (2.6). Observe that any $F \in \mathcal{C}_0$ is bounded and

$$\sup_{n \in \mathbb{N}} \sup_{\nu \in \mathcal{P}_{n,a}} |\mathbf{B}_i^n F(\nu)| < \infty. \tag{2.8}$$

One can easily verify that the martingale problem for each $\mathbf{B}_i^n$ is well-posed. If $\nu_0 = (1/n) \sum_{j=1}^n \delta_{x_j} \in \mathcal{P}_{n,a}$ then the solution of the martingale problem for $(\mathbf{B}_i^n, \delta_{\nu_0})$ is just the empirical measure process of a system of n individuals moving in E according to

independent Markov processes with generators B_i and initial positions $x_1, \ldots, x_n$. For more details see Section 2.2 in Dawson[7].

For any $f_1, \ldots, f_m \in \mathcal{D}_0$ consider a function $\hat{F} : \mathcal{M}_F^q(E) \to \mathbb{R}$ of the form

$$\hat{F}(\mu) = \prod_{j=1}^{m} \left(\sum_{i=1}^{q} c_i(h) \langle f_j, \mu_i \rangle \right), \tag{2.9}$$

where $h = H(\mu)$ and $c : \mathbb{R}_+^q \to \mathbb{R}_+^q$ is a function that satisfies

$$\sup_{h \in \mathbb{R}_+^q} \langle c(h), h \rangle < \infty. \tag{2.10}$$

Such a function $\hat{F}$ is bounded because

$$\sup_{\mu \in \mathcal{M}_F^q(E)} |\hat{F}(\mu)| \leq \left(\max_{l=1,\ldots,m} \|f_l\|_\infty \right)^m \left(\sup_{h \in \mathbb{R}_+^q} \langle c(h), h \rangle \right)^m.$$

Define a class of functions by

$$\mathcal{C}_0^q = \left\{ \hat{F} \in B\left(\mathcal{M}_F^q(E)\right) : \hat{F} \text{ is given by (2.9) and } c \text{ satisfies (2.10)} \right\}. \tag{2.11}$$

If $\mu = (\mu_1, \ldots, \mu_q) \in \mathcal{M}_{N,a}^q(E)$, then for each $i \in Q$, μ_i has the form $\mu_i = (1/N) \sum_{j=1}^{n_i} \delta_{x_j^i}$. Let $\nu_i = (1/n_i) \sum_{j=1}^{n_i} \delta_{x_j^i}$ if $n_i > 0$ and $\nu_i = \delta_{x_0}$ if $n_i = 0$, where $x_0 \in E$ is arbitrary. Let P_k^m be as before. Pick a $\hat{F} \in \mathcal{C}_0^q$ of the form (2.9). For any $p \in P_k^m$ and $l = 1, \ldots, k$ define $F_l^{(p)} \in \mathcal{C}_0$ by $F_l^{(p)}(\nu) = \langle f_l^{(p)}, \nu \rangle$, where $f_l^{(p)}$ is given by (2.5). We can write the function $\hat{F}$ as

$$\hat{F}(\mu) = \sum_{i_1,\ldots,i_m=1}^{q} \prod_{j=1}^{m} c_{i_j}(h) \langle f_j, \mu_{i_j} \rangle = \sum_{i_1,\ldots,i_m=1}^{q} \prod_{j=1}^{m} (c_{i_j}(h) h_{i_j}) \langle f_j, \nu_{i_j} \rangle$$

$$= \sum_{k=1}^{m} \sum_{p \in P_k^m} \sum_{l_1,\ldots,l_k=1}^{q} (c_{l_1}(h) h_{l_1})^{|p^{-1}(1)|} \ldots (c_{l_k}(h) h_{l_k})^{|p^{-1}(k)|} \prod_{r=1}^{k} F_r^{(p)}(\nu_{l_r}),$$

where the last term has summation over distinct choices of $l_1, \ldots, l_k \in \{1, \ldots, q\}$. Let $\mathbf{B}^N : \mathcal{D}(\mathbf{B}^N) = \mathcal{C}_0^q \to B\left(\mathcal{M}_{N,a}^q(E)\right)$ be the operator whose action on any $\hat{F} \in \mathcal{C}_0^q$ written in this form is given by

$$\mathbf{B}^N \hat{F}(\mu) = \sum_{k=1}^{m} \sum_{p \in P_k^m} \sum_{l_1,\ldots,l_k=1}^{q} (c_{l_1}(h) h_{l_1})^{|p^{-1}(1)|} \ldots (c_{l_k}(h) h_{l_k})^{|p^{-1}(k)|}$$

$$\times \sum_{r=1}^{k} \mathbf{B}_{l_r}^{n_{l_r}} F_r^{(p)}(\nu_{l_r}) \prod_{j=1,j\neq r}^{k} F_j^{(p)}(\nu_{l_j}), \tag{2.12}$$

where for any $n \in \mathbb{N}, i \in Q$, the operator $\mathbf{B}_i^n$ is given by (2.7). For convenience $\mathbf{B}_i^0$ is defined to be the identity map on $\mathcal{C}_0$. The function $\mathbf{B}^N \hat{F}$ is bounded due to (2.8) and (2.10).

The martingale problem for $\mathbf{B}^N$ is well-posed because the martingale problem for $\mathbf{B}_i^n$ is well-posed for each $i \in Q, n \in \mathbb{N}$. To see this suppose that $\mu_0 = (\mu_{0,1}, \ldots, \mu_{0,q}) \in \mathcal{M}_{N,a}^q(E)$ and for each $i \in Q$, $\mu_{0,i} \in \mathcal{M}_{N,a}(E)$ has the form $\mu_{0,i} = (1/N) \sum_{j=1}^{n_i} \delta_{x_j^i}$. If $n_i > 0$ let $\nu_{0,i} = (1/n_i) \sum_{j=1}^{n_i} \delta_{x_j^i}$ and if $n_i = 0$ let $\nu_{0,i} = \delta_{x_0}$ for some arbitrary $x_0 \in E$. For each $i \in Q$, let $\{\nu_i(t) : t \geq 0\}$ be the unique solution of the martingale problem for $(\mathbf{B}_i^{n_i}, \delta_{\nu_{0,i}})$. Then the process $\{\mu(t) : t \geq 0\}$ given by

$$\mu(t) = \left(\frac{n_1}{N} \nu_1(t), \ldots, \frac{n_q}{N} \nu_q(t) \right),$$

is the unique solution to the martingale problem for $(\mathbf{B}^N, \delta_{\mu_0})$.

2.2 The density regulation mechanism

We mentioned in Section 1 that in our models, the birth and death rates of individuals depend on the population density in such a way that they induce an equilibrating density regulation mechanism. We now describe this mechanism in our multi-type setting.

For each $i, j \in Q$, let β_{ij} be a bounded function in $C^2(\mathbb{R}^q_+, \mathbb{R}_+)$ and for each $i \in Q$ let ρ_i be any function in $C^2(\mathbb{R}^q_+, \mathbb{R}_+)$. For any $h \in \mathbb{R}^q_+$, let the matrix $A(h) \in \mathbb{M}(q, q)$ and the vector $\theta(h) \in \mathbb{R}^q$ be given by

$$A_{ij}(h) = \left\{ \begin{array}{cc} \beta_{ji}(h) & \text{if } i \neq j \\ \beta_{ii}(h) - \rho_i(h) & \text{otherwise} \end{array} \right\}, \tag{2.13}$$

and

$$\theta(h) = A(h)h. \tag{2.14}$$

Consider θ to be a vector field over $\mathbb{R}^q_+$. Observe that if $h \in \mathbb{R}^q_+$ is such that $h_i = 0$ then $\theta_i(h) \geq 0$. This shows that any solution to the initial value problem

$$\frac{dh}{dt} = \theta(h), \quad h(0) = h_0 \in \mathbb{R}^q_+ \tag{2.15}$$

stays inside $\mathbb{R}^q_+$ for all positive times for which it is defined. Standard existence and uniqueness theorems imply that for any $h_0 \in \mathbb{R}^q_+$, there is a solution $h(t)$ of (2.15) defined on some maximal time interval $[0, a)$. Moreover if $a < \infty$ then $\|h(t)\|_1 \to \infty$ as $t \to a_-$. Since β_{ij} is bounded for each $i, j \in Q$, there is a positive constant C_θ such that

$$\sum_{i=1}^q \theta_i(h) \leq \sum_{i,j=1}^q \beta_{ji}(h)h_j \leq C_\theta \|h\|_1 \text{ for all } h \in \mathbb{R}^q_+ \tag{2.16}$$

and hence

$$\frac{d\|h(t)\|_1}{dt} = \sum_{i=1}^q \frac{dh_i(t)}{dt} = \sum_{i=1}^q \theta_i(h(t)) \leq C_\theta \|h(t)\|_1. \tag{2.17}$$

Therefore using Gronwall's inequality and (2.17) we obtain that $\|h(t)\|_1 \leq \|h(0)\|_1 e^{C_\theta t}$ for all $t \in [0, a)$ and so $\|h(t)\|_1$ cannot go to ∞ as $t \to a_-$. Thus $a = \infty$ and this shows that for any $h_0 \in \mathbb{R}^q_+$, the initial value problem (2.15) has a unique solution which is defined for all $t \geq 0$.

Let $\psi_\theta : \mathbb{R}^q_+ \times \mathbb{R}_+ \to \mathbb{R}^q_+$ be the *flow* associated to the vector field θ. This means that ψ_θ satisfies

$$\psi_\theta(x, t) = x + \int_0^t \theta(\psi_\theta(x, s))ds \text{ for all } (x, t) \in \mathbb{R}^q_+ \times \mathbb{R}_+. \tag{2.18}$$

This flow is well-defined because of the arguments given in the preceding paragraph. In fact since θ is in $C^2(\mathbb{R}^q_+, \mathbb{R}^q)$, the map ψ_θ is in $C^2(\mathbb{R}^q_+ \times \mathbb{R}_+, \mathbb{R}^q_+)$. This map also satisfies the semigroup property

$$\psi_\theta(x, t + s) = \psi_\theta(\psi_\theta(x, t), s) \text{ for all } x \in \mathbb{R}^q_+ \text{ and } s, t \in \mathbb{R}_+. \tag{2.19}$$

We will say that a set $U \subset \mathbb{R}^q_+$ is ψ_θ-invariant if for all $t \geq 0$, $\psi_\theta(U, t) \subset U$ where $\psi_\theta(U, t) = \{\psi_\theta(x, t) : x \in U\}$. Before we proceed, we need to make some more assumptions.

Assumption 2.1. *(A) There exists a vector $h_{\text{eq}} \in \mathbb{R}^q_+$, $h_{\text{eq}} \neq \bar{0}_q$ such that $\theta(h_{\text{eq}}) = A(h_{\text{eq}})h_{\text{eq}} = \bar{0}_q$.*

(B) The Jacobian matrix $[J\theta(h_{eq})]$ is stable.

(C) The matrix $A(h_{eq})$ is irreducible, that is, there does not exist a permutation matrix $P \in \mathbb{M}(q,q)$ such that the matrix $PA(h_{eq})P^{-1}$ is block upper-triangular.

(D) For each $i,j \in Q$, the functions ρ_i and β_{ij} are analytic at h_{eq}, that is, they agree with their Taylor series expansion in a neighbourhood of h_{eq}.

Part (A) says that there is a nonzero vector $h_{eq} \in \mathbb{R}_+^q$ which is a fixed point for the flow ψ_θ. Part (B) implies that this fixed point h_{eq} is *asymptotically stable* for this flow. The significance of part (C) will become clear later in this section. Part (D) is a technical condition that we require to prove our main result.

We define the region of attraction of the fixed point h_{eq} for the flow ψ_θ by

$$U_{eq} = \left\{ h \in \mathbb{R}_+^q : \lim_{t \to \infty} \psi_\theta(h,t) = h_{eq} \right\}. \tag{2.20}$$

Part (B) of Assumption 2.1 and Lemma 3.2 in Khalil [24], ensure that U_{eq} is a ψ_θ-invariant open set in $\mathbb{R}_+^q$. Note that U_{eq} may not be an open set in $\mathbb{R}^q$.

We are now ready to describe the density regulation mechanism. Suppose that when the population density vector is $h \in \mathbb{R}_+^q$, then for each $i,j \in Q$, an individual of type i gives birth to an individual of type j at rate $\beta_{ij}(h)$ and an individual of type i dies at rate $\rho_i(h)$. At the time of birth, the offspring is placed at the same location in E as its parent. Note that the birth and death rates do not depend on the location of the individuals. If the scaling parameter is $N \in \mathbb{N}$ and the mass of each individual is $1/N$, then we can view this density-dependent population dynamics as a Markov process over the state space $\mathcal{M}_{N,a}^q(E)$ with generator $\mathbf{R}^N : \mathcal{D}(\mathbf{R}^N) = B(\mathcal{M}_{N,a}^q(E)) \to B(\mathcal{M}_{N,a}^q(E))$ defined as follows. For any $F \in B(\mathcal{M}_{N,a}^q(E))$ and any $\mu \in \mathcal{M}_{N,a}^q(E)$ with $h = H(\mu)$

$$\mathbf{R}^N F(\mu) = \sum_{i,j \in Q} N \int_E \beta_{ij}(h) \left(F\left(\mu + \frac{1}{N}\delta_x^j\right) - F(\mu) \right) \mu_i(dx) \tag{2.21}$$
$$+ \sum_{i \in Q} N \int_E \rho_i(h) \left(F\left(\mu - \frac{1}{N}\delta_x^i\right) - F(\mu) \right) \mu_i(dx),$$

where for any $\mu = (\mu_1, \ldots, \mu_q) \in \mathcal{M}_{N,a}^q(E)$, $j \in Q$ and $x \in E$

$$\mu \pm \frac{1}{N}\delta_x^j = \left(\mu_1, \ldots, \mu_{j-1}, \mu_j \pm \frac{1}{N}\delta_x, \mu_{j+1}, \ldots, \mu_q \right).$$

Concrete results on the well-posedness of the martingale problem for $\mathbf{R}^N$ will come later. First let us understand how this operator $\mathbf{R}^N$ drives the population density to an equilibrium value. Let $\{\mu^N(t) : t \geq 0\}$ be a $\mathcal{M}_{N,a}^q(E)$-valued Markov process with generator $\mathbf{R}^N$ and let $\{h^N(t) : t \geq 0\}$ be the corresponding density process defined by $h^N(t) = H(\mu^N(t))$. Assume that $h^N(0) \to h_0$ a.s. as $N \to \infty$ and h_0 is some vector in U_{eq}. From Theorem 11.2.1 in [12], one can conclude that as $N \to \infty$, $\{h^N(t) : t \geq 0\}$ converges in the Skorohod topology in $D_{\mathbb{R}^q}[0,\infty)$ to the process $\{\psi_\theta(h_0,t) : t \geq 0\}$. Since $h_0 \in U_{eq}$, $\psi_\theta(h_0,t) \to h_{eq}$ as $t \to \infty$ which indicates that for a large N, the density process $h^N(\cdot)$ gets closer and closer to h_{eq} with time. This shows, at least informally, that the operator $\mathbf{R}^N$ drives the population density towards h_{eq}. Henceforth we shall refer to the vector h_{eq} as the *equilibrium population density*. Of course, this discussion totally ignores how $\mathbf{R}^N$ affects the spatial configuration of the population. Our main goal in this paper is to discover how the spatial distribution of the population evolves when the density is equilibrated at a faster timescale in comparison to the other mechanisms.

For any $h \in \mathbb{R}_+^q$, the matrix $A(h)$ given by (2.13) signifies how the various types of individuals interact when the population density is h. Part (C) of Assumption 2.1 means that at the equilibrium population density, all the types of individuals are communicating with each other by influencing each other's birth and death rates.

2.3 Mathematical Models

We now describe the models that we consider in this paper. We start with a basic model which only has spatial migration along with density regulation. We then extend this model by adding other features such as position dependence in the birth and death rates, offspring dispersal and immigration. All the models will be parameterized by the scaling parameter N, with $1/N$ being the mass of each individual in the population. We will describe each model by specifying the generator of the associated Markov process. The well-posedness of the martingale problems corresponding to these generators is given by Proposition 2.4. Our main results are presented in Section 2.4.

2.3.1 Basic Model

In this model, the individuals migrate according to the type-dependent migration mechanism specified in Section 2.1 and their birth and death rates regulate the population density as described in Section 2.2. If the scaling parameter is N, then at any time we represent the population as a measure in $\mathcal{M}_{N,a}^q(E)$. The population evolves due to the following events.

- Each individual of type $i \in Q$ migrates in E according to an independent Markov process with generator B_i.
- When the population density vector is $h \in \mathbb{R}_+^q$, each individual of type $i \in Q$ gives birth to an individual of type $j \in Q$ at rate $N\beta_{ij}(h)$. At the time of birth, the offspring is placed at the same location as its parent.
- When the population density vector is $h \in \mathbb{R}_+^q$, each individual of type $i \in Q$ dies at rate $N\rho_i(h)$.

This population dynamics can be viewed as a $\mathcal{M}_{N,a}^q(E)$-valued Markov process whose generator $\mathbf{A}_0^N : \mathcal{D}(\mathbf{A}_0^N) = \mathcal{C}_0^q \to B(\mathcal{M}_{N,a}^q(E))$ is given by

$$\mathbf{A}_0^N \hat{F}(\mu) = \mathbf{B}^N \hat{F}(\mu) + N^2 \sum_{i,j \in Q} \int_E \beta_{ij}(h) \left(\hat{F}\left(\mu + \frac{1}{N}\delta_x^j\right) - \hat{F}(\mu) \right) \mu_i(dx) \qquad (2.22)$$
$$+ N^2 \sum_{i \in Q} \int_E \rho_i(h) \left(\hat{F}\left(\mu - \frac{1}{N}\delta_x^i\right) - \hat{F}(\mu) \right) \mu_i(dx),$$

for any $\hat{F} \in \mathcal{C}_0^q$. Here $h = H(\mu)$ is the density vector corresponding to μ.

2.3.2 Model with position dependence in the birth and death rates

In the above model, the birth and death rates of individuals do not depend on their location. One may want to consider models in ecology where some spatial locations are more advantageous for reproduction or some locations are more hazardous for survival. If we think of E as the space of genetic traits, then one may consider models in which the trait of an individual influences its chances of reproduction or survival. To capture such situations we now introduce a model in which the birth and death rates of an individual can vary with its position. However we will assume that this position dependent variation is *small*, in the sense that even though an individual's birth and death rate is

of order N (as in Section 2.3.1), the spatial variation in these rates is of order 1. We now make the model precise.

For each $i, j \in Q$, let b_{ij}^s, d_i^s be bounded continuous functions from $E \times \mathbb{R}_+^q$ to $\mathbb{R}_+$. These functions determine the *spatial* variation in the birth and death rates. The migration of individuals is like in the basic model. However now the birth and death mechanism changes as follows.

- When the population density vector is $h \in \mathbb{R}_+^q$, an individual of type $i \in Q$ located at $x \in E$, gives birth to an individual of type $j \in Q$ at rate $b_{ij}^s(x, h) + N\beta_{ij}(h)$. At the time of birth, the offspring is placed at the same location as its parent.
- When the population density vector is $h \in \mathbb{R}_+^q$, an individual of type $i \in Q$ located at $x \in E$, dies at rate $d_i^s(x, h) + N\rho_i(h)$.

The evolution of our population under this dynamics can be viewed as a $\mathcal{M}_{N,a}^q(E)$-valued Markov process with generator $\mathbf{A}_1^N : \mathcal{D}(\mathbf{A}_1^N) = \mathcal{C}_0^q \to B(\mathcal{M}_{N,a}^q(E))$ defined for any $\hat{F} \in \mathcal{C}_0^q$ by

$$
\mathbf{A}_1^N \hat{F}(\mu) = \mathbf{B}^N \hat{F}(\mu) + \sum_{i,j \in Q} N \int_E \left(b_{ij}^s(x, h) + N\beta_{ij}(h) \right) \left(\hat{F}\left(\mu + \frac{1}{N}\delta_x^j \right) - \hat{F}(\mu) \right) \mu_i(dx)
$$

$$
\tag{2.23}
$$

$$
+ \sum_{i \in Q} N \int_E (d_i^s(x, h) + N\rho_i(h)) \left(\hat{F}\left(\mu - \frac{1}{N}\delta_x^i \right) - \hat{F}(\mu) \right) \mu_i(dx),
$$

where $h = H(\mu)$ is the density vector corresponding to μ.

2.3.3 Model with offspring dispersal at birth

In the basic model we described in Section 2.3.1, the offspring is placed at the same location as its parent at the time of birth. However we may want to construct models where this restriction needs to be relaxed. For example, while modeling plant populations, one may wish to account for the spreading of seeds due to wind and other factors. Also in models for population genetics, where E is the space of genetic traits, offsprings may be born with a different trait than their parents due to mutations. To consider such situations we now present a model in which an offspring may be born away from its parent. We allow this offspring dispersal to either be *rare* (happens with probability proportional to $1/N$) or *small* (the offspring is placed at a distance proportional to $1/N$ from the parent). We handle both these cases in a unified way.

For each $i, j \in Q$ and $N \in \mathbb{N}$, let ϑ_{ij}^N be a function from E to $\mathcal{P}(E)$ and let p_{ij}^N be a function from E to $[0, 1]$. The individuals migrate and die in the same way as described in the basic model (Section 2.3.1). The birth rates are also the same as in the basic model. However when an individual of any type $i \in Q$ located at $x \in E$, gives birth to an individual of type j, the location of the offspring is x with probability $(1 - p_{ij}^N(x))$ and distributed according to $\vartheta_{ij}^N(x, \cdot)$ with probability $p_{ij}^N(x)$.

To pass to the limit $N \to \infty$, we need an assumption on p_{ij}^N and ϑ_{ij}^N which is stated below.

Assumption 2.2. *For each $i, j \in Q$ we assume that there is an operator $C_{ij} : \mathcal{D}(C_{ij}) \to C(E)$, whose domain is taken to be the same as $\mathcal{D}_0$ (see (2.3)) for convenience, such that for every $f \in \mathcal{D}_0$.*

$$
\lim_{N \to \infty} \sup_{x \in E} \left| N p_{ij}^N(x) \int_E (f(y) - f(x)) \vartheta_{ij}^N(x, dy) - C_{ij} f(x) \right| = 0.
$$

The evolution of our population under the dynamics described above can be viewed as a $\mathcal{M}_{N,a}^q(E)$-valued Markov process with generator $\mathbf{A}_2^N : \mathcal{D}(\mathbf{A}_2^N) = \mathcal{C}_0^q \to B(\mathcal{M}_{N,a}^q(E))$ defined by its action on any $\hat{F} \in \mathcal{C}_0^q$ by

$$\mathbf{A}_2^N \hat{F}(\mu) = \mathbf{B}^N \hat{F}(\mu) + N^2 \sum_{i \in Q} \int_E \rho_i(h) \left(\hat{F}\left(\mu - \frac{1}{N}\delta_x^i \right) - \hat{F}(\mu) \right) \mu_i(dx) \quad (2.24)$$

$$+ N^2 \sum_{i,j \in Q} \int_E \beta_{ij}(h) \left(1 - p_{ij}^N(x) \right) \left(\hat{F}\left(\mu + \frac{1}{N}\delta_x^j \right) - \hat{F}(\mu) \right) \mu_i(dx)$$

$$+ N^2 \sum_{i,j \in Q} \int_E \beta_{ij}(h) p_{ij}^N(x) \left[\int_E \left(\hat{F}\left(\mu + \frac{1}{N}\delta_y^j \right) - \hat{F}(\mu) \right) \vartheta_{ij}^N(x, dy) \right] \mu_i(dx),$$

where $h = H(\mu)$ is the density vector corresponding to μ.

2.3.4 Model with immigration

Consider a population whose dynamics is as described in the basic model (Section 2.3.1). In addition, suppose that the individuals of each type are immigrating to E at a certain density dependent rate and settling down according to some distribution on E. In this section we model this situation. Such a model can help us understand the effects of immigration on the population demography.

For each $i \in Q$, let $\kappa_i : \mathbb{R}_+^q \to \mathbb{R}_+$ be a continuous function satisfying

$$\kappa_i(h) \leq C(1 + \|h\|_1) \text{ for all } h \in \mathbb{R}_+^q, \quad (2.25)$$

for some $C > 0$. The individuals migrate, reproduce and die as in the basic model. Moreover, when the population density vector is $h \in \mathbb{R}_+^q$, the individuals of each type $i \in Q$ arrive in the population at rate $N\kappa_i(h)$ and their initial location is given by the distribution $\Theta_i \in \mathcal{P}(E)$.

The evolution of our population under this dynamics can be viewed as a $\mathcal{M}_{N,a}^q(E)$-valued Markov process with generator $\mathbf{A}_3^N : \mathcal{D}(\mathbf{A}_3^N) = \mathcal{C}_0^q \to B(\mathcal{M}_{N,a}^q(E))$ defined by its action on any $\hat{F} \in \mathcal{C}_0^q$ as

$$\mathbf{A}_3^N \hat{F}(\mu) = \mathbf{B}^N \hat{F}(\mu) + N \sum_{i \in Q} \kappa_i(h) \int_E \left(\hat{F}\left(\mu + \frac{1}{N}\delta_x^i \right) - \hat{F}(\mu) \right) \Theta_i(dx) \quad (2.26)$$

$$+ N^2 \sum_{i,j \in Q} \int_E \beta_{ij}(h) \left(\hat{F}\left(\mu + \frac{1}{N}\delta_x^j \right) - \hat{F}(\mu) \right) \mu_i(dx)$$

$$+ N^2 \sum_{i \in Q} \int_E \rho_i(h) \left(\hat{F}\left(\mu - \frac{1}{N}\delta_x^i \right) - \hat{F}(\mu) \right) \mu_i(dx),$$

where $h = H(\mu)$ is the density vector corresponding to μ.

Remark 2.3. For each $l \in \{0, 1, 2, 3\}$ define the operators $\mathbf{G}_l^N : \mathcal{D}(\mathbf{G}_l^N) = B(\mathcal{M}_F^q(E)) \to B(\mathcal{M}_F^q(E))$ as follows. For any $F \in B\left(\mathcal{M}_F^q(E) \right)$ let

$$\mathbf{G}_0^N F(\mu) = 0, \quad (2.27)$$

$$\mathbf{G}_1^N F(\mu) = N \left[\sum_{i,j \in Q} \int_E b_{ij}^s(x, h) \left(F\left(\mu + \frac{1}{N}\delta_x^j \right) - F(\mu) \right) \mu_i(dx) \quad (2.28) \right.$$

$$\left. + \sum_{i \in Q} \int_E d_i^s(x, h) \left(F\left(\mu - \frac{1}{N}\delta_x^i \right) - F(\mu) \right) \mu_i(dx) \right],$$

$$\mathbf{G}_2^N F(\mu) = N^2 \sum_{i,j \in Q} \beta_{ij}(h) \int_E \left[\int_E \left(F\left(\mu + \frac{1}{N}\delta_y^j\right) - F\left(\mu + \frac{1}{N}\delta_x^j\right) \right) \vartheta_{ij}^N(x, dy) \right]$$
$$\times p_{ij}^N(x)\mu_i(dx) \tag{2.29}$$

$$\text{and } \mathbf{G}_3^N F(\mu) = N \sum_{i \in Q} \kappa_i(h) \int_E \left(F\left(\mu + \frac{1}{N}\delta_x^i\right) - F(\mu) \right) \Theta_i(dx), \tag{2.30}$$

where $h = H(\mu)$. Then for each $l \in \{0, 1, 2, 3\}$ and $\hat{F} \in \mathcal{C}_0^q$ we can write

$$\mathbf{A}_l^N \hat{F} = \mathbf{B}^N \hat{F} + N\mathbf{R}^N \hat{F} + \mathbf{G}_l^N \hat{F}. \tag{2.31}$$

This form makes the timescale separation clear between the *fast* density regulation mechanism ($N\mathbf{R}^N$) and the *slow* migration ($\mathbf{B}^N$), position-dependent birth and death ($\mathbf{G}_1^N$), offspring dispersal ($\mathbf{G}_2^N$) and immigration ($\mathbf{G}_3^N$) mechanisms. $\qquad \square$

2.4 The main results

The main results of this paper are concerned with the limiting behaviour of the dynamics under the models described in Section 2.3. Before we present these results we must first verify that all the models in Section 2.3 can be represented by a suitable Markov process. This is established by the following proposition which will be proved in Section 4.1.

Proposition 2.4. *For each $l \in \{0, 1, 2, 3\}$ and $N \in \mathbb{N}$, the $D_{\mathcal{M}_{N,a}^q(E)}[0, \infty)$ martingale problem for $\mathbf{A}_l^N$ is well-posed.*

We now begin analyzing how a sequence of Markov processes with generators $\mathbf{A}_l^N$ behave as $N \to \infty$. The next proposition exhibits some important properties about the limiting dynamics. The proof of this proposition is given in Section 4.2. Recall the definition of U_{eq} from (2.20).

Proposition 2.5. *Fix a $l \in \{0, 1, 2, 3\}$. For each $N \in \mathbb{N}$, let $\{\mu^N(t) : t \geq 0\}$ be a solution of the martingale problem for $\mathbf{A}_l^N$ and let $\{h^N(t) = H(\mu^N(t)) : t \geq 0\}$ be the corresponding density process. Assume that there is a compact set $K_0 \subset U_{\text{eq}}$ such that $h^N(0) \in K_0$ a.s. for all $N \in \mathbb{N}$. Let t_N be a sequence of positive numbers satisfying $t_N \to 0$ and $Nt_N \to \infty$ as $N \to \infty$. Then we have:*

(A) For all $T > 0$

$$\sup_{t \in [0,T]} \left\| h^N(t + t_N) - h_{\text{eq}} \right\|_1 \Rightarrow 0 \text{ as } N \to \infty.$$

(B) For all $T > 0$, $f \in C(E)$ and $i, j \in Q$

$$\sup_{t \in [0,T]} \left| h_j^N(t + t_N)\langle f, \mu_i^N(t + t_N)\rangle - h_i^N(t + t_N)\langle f, \mu_j^N(t + t_N)\rangle \right| \Rightarrow 0 \text{ as } N \to \infty.$$

Let the processes $\{\mu^N(t) : t \geq 0\}$ and $\{h^N(t) : t \geq 0\}$ be as in the above proposition. Part (A) of this proposition implies that the process $h^N(\cdot + t_N) \Rightarrow h_{\text{eq}}$ as $N \to \infty$ in $D_{\mathbb{R}^q}[0, \infty)$. In other words, for large N, the density process is constantly near the equilibrium population density h_{eq} (after a small time shift t_N). This emphasizes the point that we made in Section 1. The density regulation mechanism operating at a faster timescale than our timescale of observation, keeps the population density equilibrated at all times. Note that the process $\{\mu^N(t) = (\mu_1^N(t), \ldots, \mu_q^N(t)) : t \geq 0\}$ is $\mathcal{M}_F^q(E)$-valued and it keeps track of how the populations corresponding to all the q-types are evolving in the space E. Part (B) of the above proposition shows that in the limit, all the q

sub-populations are spatially *fused* (in proportions determined by the density vector). Hence their spatial evolution can be studied together by using a single $\mathcal{P}(E)$-valued process. This kind of model reduction result is quite common in stochastic reaction networks with multiple timescales (see [3] and [22]), where one can often equilibrate the concentrations of the *fast* chemical species and derive a reduced model for the dynamics of the *slow* species. In our case, the dynamics at the fast timescale equilibrates the population density as well as the relative abundances of all the q sub-populations at each location on E. These relative abundances equilibrate because the birth-death interaction matrix $A(h)$ (see (2.13)) is irreducible at the equilibrium density h_{eq} (see part (C) of Assumption 2.1).

The proof of Proposition 2.5 will exploit the form (2.31) of the operator $\mathbf{A}_l^N$. A brief outline of the proof is as follows. We will define a $\mathbb{R}_+^q \times \mathbb{R}^{q-1}$ valued process $\{X^N(t) : t \geq 0\}$ by

$$X^N(t) = (h_1^N(t), \ldots, h_q^N(t), Y_1^N(t), \ldots, Y_{q-1}^N(t)), \tag{2.32}$$

where each $Y_i^N(t)$ is a density-dependent linear combination of terms like $(h_j^N(t)\langle f, \mu_i^N(t)\rangle - h_i^N(t)\langle f, \mu_j^N(t)\rangle)$ for some choice of $f \in C(E)$. Next we will show that X^N is a semimartingale which satisfies an equation of the form

$$X^N(t) = X^N(0) + N \int_0^t F(X^N(s))ds + Z^N(t), \tag{2.33}$$

where $\{Z^N : N \in \mathbb{N}\}$ is a sequence of $\mathbb{R}^{2q-1}$-semimartingales which is tight in the space $D_{\mathbb{R}^{2q-1}}[0, \infty)$. This clearly indicates that for large values of N, the drift term of the form $NF(X^N(\cdot))$ completely overwhelms the effect of the semimartingale Z^N. Equations like (2.33) were studied by Katzenberger in [23] in a much more general setting. He showed that under certain conditions, the sequence of semimartingales $\{X^N : N \in \mathbb{N}\}$ converges in distribution to a semimartingale X as $N \to \infty$. Moreover X only takes values in a set Γ which is an invariant manifold for the deterministic flow induced by F. In our case, this set Γ only consists of one point $x_{eq} = (h_{eq}, \bar{0}_{q-1})$ and this enables us to prove Proposition 2.5. The details are given in Section 4.2.

We mentioned before that in the limit $N \to \infty$, the spatial evolution of all the q sub-populations is governed by a single $\mathcal{P}(E)$-valued process. Our next result, Theorem 2.6, shows that this $\mathcal{P}(E)$-valued process is in fact a Fleming-Viot process that can be characterized by its generator. Before we state Theorem 2.6 we first need to introduce several objects. The existence and properties of some of these objects will be studied in the appendix.

Recall the equilibrium population density vector $h_{eq} = (h_{eq,1}, \ldots, h_{eq,q})$ from Section 2.2. It can be verified that this vector has strictly positive components (see part (A) of Lemma A.1). Moreover part (C) of Lemma A.1 shows that there is a unique vector $v_{eq} = (v_{eq,1}, \ldots, v_{eq,q}) \in \mathbb{R}_*^q$ such that

$$v_{eq}A(h_{eq}) = \bar{0}_q \quad \text{and} \quad \langle v_{eq}, h_{eq}\rangle = \sum_{i=1}^q v_{eq,i}h_{eq,i} = 1. \tag{2.34}$$

Observe that $\mathcal{D}_0 \subset C(E)$ satisfies (2.3). Define an operator $B_{avg} : \mathcal{D}(B_{avg}) = \mathcal{D}_0 \to C(E)$ by

$$B_{avg}f = \sum_{i \in Q} v_{eq,i}h_{eq,i}B_i f \quad \text{for} \ f \in \mathcal{D}_0. \tag{2.35}$$

From (2.34) we can see that the operator B_{avg} is a convex combination of the operators $\{B_i : i \in Q\}$. Let γ_{smpl} be the positive constant given by

$$\gamma_{\text{smpl}} = \sum_{i \in Q} v_{\text{eq},i}^2 h_{\text{eq},i} \rho_i(h_{\text{eq}}). \tag{2.36}$$

For each $i, j \in Q$, let the functions b_{ij}^s, d_i^s be as in Section 2.3.2. Define $b_{\text{avg}}^s, d_{\text{avg}}^s \in C(E)$ as

$$b_{\text{avg}}^s(x) = \sum_{i,j \in Q} b_{ij}^s(x, h_{\text{eq}}) v_{\text{eq},j} h_{\text{eq},i} \ \text{ and } \ d_{\text{avg}}^s(x) = \sum_{i \in Q} d_i^s(x, h_{\text{eq}}) v_{\text{eq},i} h_{\text{eq},i} \text{ for } x \in E.$$

For each $i, j \in Q$, let the operator C_{ij} be as in Assumption 2.2. Define the operator $C_{\text{avg}} : \mathcal{D}(C_{\text{avg}}) = \mathcal{D}_0 \to C(E)$ by

$$C_{\text{avg}} f = \sum_{i,j \in Q} \beta_{ij}(h_{\text{eq}}) v_{\text{eq},j} h_{\text{eq},i} C_{ij} f \text{ for } f \in \mathcal{D}_0. \tag{2.37}$$

For each $i \in Q$, let κ_i, Θ_i be as in Section 2.3.4. Define the operator $I_{\text{avg}} : \mathcal{D}(I_{\text{avg}}) = B(\mathcal{P}(E)) \to B(\mathcal{P}(E))$ by

$$I_{\text{avg}} f(x) = \sum_{i \in Q} \kappa_i(h_{\text{eq}}) v_{\text{eq},i} \int_E (f(y) - f(x)) \Theta_i(dy) \text{ for } f \in B(\mathcal{P}(E)). \tag{2.38}$$

We now define the operators $\mathbf{A}_0, \mathbf{A}_1, \mathbf{A}_2$ and $\mathbf{A}_3$ with domain $\mathcal{C}_0$ (see (2.4)) as below. For any $F(\nu) = \prod_{l=1}^m \langle f_l, \nu \rangle \in \mathcal{C}_0$ let

$$\mathbf{A}_0 F(\nu) = \sum_{l=1}^m \langle B_{\text{avg}} f_l, \nu \rangle \prod_{j \neq l} \langle f_j, \nu \rangle \tag{2.39}$$

$$+ \gamma_{\text{smpl}} \sum_{1 \leq l \neq k \leq m} (\langle f_l f_k, \nu \rangle - \langle f_l, \nu \rangle \langle f_k, \nu \rangle) \prod_{j \neq l,k} \langle f_j, \nu \rangle,$$

$$\mathbf{A}_1 F(\nu) = \mathbf{A}_0 F(\nu) + \sum_{l=1}^m \left(\langle b_{\text{avg}}^s f_l, \nu \rangle - \langle b_{\text{avg}}^s, \nu \rangle \langle f_l, \nu \rangle \right) \prod_{j \neq l} \langle f_j, \nu \rangle \tag{2.40}$$

$$+ \sum_{l=1}^m \left(\langle d_{\text{avg}}^s, \nu \rangle \langle f_l, \nu \rangle - \langle d_{\text{avg}}^s f_l, \nu \rangle \right) \prod_{j \neq l} \langle f_j, \nu \rangle,$$

$$\mathbf{A}_2 F(\nu) = \mathbf{A}_0 F(\nu) + \sum_{l=1}^m \langle C_{\text{avg}} f_l, \nu \rangle \prod_{j \neq l} \langle f_j, \nu \rangle \tag{2.41}$$

$$\text{and } \ \mathbf{A}_3 F(\nu) = \mathbf{A}_0 F(\nu) + \sum_{l=1}^m \langle I_{\text{avg}} f_l, \nu \rangle \prod_{j \neq l} \langle f_j, \nu \rangle. \tag{2.42}$$

We will assume that the operators B_{avg}, $(B_{\text{avg}} + C_{\text{avg}})$ and $(B_{\text{avg}} + I_{\text{avg}})$ generate Feller semigroups on $C(E)$. The well-posedness of the martingale problems corresponding to $\mathbf{A}_0, \mathbf{A}_1, \mathbf{A}_2$ and $\mathbf{A}_3$ follows from Theorem 3.2 in [13]. In fact, any solution will have sample paths in $C_{\mathcal{P}(E)}[0, \infty)$. The operator $\mathbf{A}_0$ is the generator of a *neutral* Fleming-Viot process on E with *mutation* operator B_{avg} and *sampling* rate $2\gamma_{\text{smpl}}$. The operators $\mathbf{A}_2$ and $\mathbf{A}_3$ generate a similar Fleming-Viot process with the *mutation* operator changed to $(B_{\text{avg}} + C_{\text{avg}})$ and $(B_{\text{avg}} + I_{\text{avg}})$ respectively. The operator $\mathbf{A}_1$ also generates a similar Fleming-Viot process, but with *selection*. The last two terms in its definition correspond to *fecundity selection* (with intensity function b_{avg}^s) and *viability selection* (with intensity function d_{avg}^s). See Donnelly and Kurtz [8] for more details. We now formally state the main result of our paper. The proof is given in Section 4.4.

Theorem 2.6. *Fix a $l \in \{0, 1, 2, 3\}$ and let $\{\mu^N(t) : t \geq 0\}$ be a solution to the martingale problem for $\mathbf{A}_l^N$. Suppose that $\mu^N(0) \Rightarrow \mu(0)$ as $N \to \infty$ and $H(\mu(0)) \in U_{\mathrm{eq}}$ a.s. where U_{eq} is given by (2.20). Let t_N be a sequence as in Proposition 2.5. Define another process $\{\hat{\mu}^N(t) : t \geq 0\}$ by*

$$\hat{\mu}^N(t) = \mu^N(t + t_N) \text{ for } t \geq 0. \tag{2.43}$$

Then there exists a distribution $\pi \in \mathcal{P}(\mathcal{P}(E))$ such that $\hat{\mu}^N \Rightarrow h_{\mathrm{eq}}\nu$ in $D_{\mathcal{M}_F^q(E)}[0, \infty)$ as $N \to \infty$ and $\{\nu(t) : t \geq 0\}$ is a Fleming-Viot process with type space E, generator $\mathbf{A}_l$ and initial distribution π.

Remark 2.7. *The initial distribution π of the process $\{\nu(t) : t \geq 0\}$ is related to the distribution of $\mu(0)$. This relation is stated in Remark 4.10.*

Remark 2.8. *In Section 2.3 we first defined a basic model and then constructed auxiliary models by adding other mechanisms, one at a time. These other mechanisms are position dependent birth and death, offspring dispersal and immigration. One can consider models in which more than one of these mechanisms are simultaneously added to the basic model. The proof will demonstrate that the generator of the limiting Fleming-Viot process is then obtained by adding the correct term corresponding to each of these additional mechanisms to the operator $\mathbf{A}_0$. This correct term can be seen from the definitions of $\mathbf{A}_1$, $\mathbf{A}_2$ and $\mathbf{A}_3$. For example, one can have the basic model along with position dependent birth and death (Section 2.3.2) and offspring dispersal (Section 2.3.3). Then the limiting Fleming-Viot process has the generator given by*

$$\mathbf{A}F(\nu) = \mathbf{A}_0 F(\nu) + \sum_{l=1}^m \left(\langle b_{\mathrm{avg}}^s f_l, \nu \rangle - \langle b_{\mathrm{avg}}^s, \nu \rangle \langle f_l, \nu \rangle \right) \prod_{j \neq l} \langle f_j, \nu \rangle$$

$$+ \sum_{l=1}^m \left(\langle d_{\mathrm{avg}}^s, \nu \rangle \langle f_l, \nu \rangle - \langle d_{\mathrm{avg}}^s f_l, \nu \rangle \right) \prod_{j \neq l} \langle f_j, \nu \rangle + \sum_{l=1}^m \langle C_{\mathrm{avg}} f_l, \nu \rangle \prod_{j \neq l} \langle f_j, \nu \rangle,$$

for any $F(\nu) = \prod_{l=1}^m \langle f_l, \nu \rangle \in \mathcal{C}_0$.

We now give a heuristic explanation of why the dynamics under the models described in Section 2.3 converges to a Fleming-Viot process. Note that part (A) of Proposition 2.5 says that the population density is *pinned* to a constant value h_{eq} in the limit. Therefore any addition of new mass in the population must be concurrently offset by an equal reduction of existing mass and vice versa. Furthermore, when the mass is reduced or added to keep the balance, this reduction or addition happens at locations that are chosen more or less uniformly from the current empirical measure of the population. This is because the birth and deaths rates of individuals are dominated by a term which is density dependent but location independent. This argument offers some intuition as to why the fast birth-death terms (that form part of the operator $N\mathbf{R}^N$) give rise to the *sampling* term in the limit (the second term in $\mathbf{A}_0$). It also shows why the position dependent birth and death terms in $\mathbf{A}_1^N$ become *selection* terms in $\mathbf{A}_1$ and the offspring dispersal (immigration) term in $\mathbf{A}_2^N$ ($\mathbf{A}_3^N$) becomes a *mutation* term in $\mathbf{A}_2$ ($\mathbf{A}_3$). Since the position of an individual in E can also be seen as its genetic trait, one can interpret the migration on E as genetic mutation. Hence it is not surprising that the migration operators appear as part of the mutation operator in the limiting process. Part (B) of Proposition 2.5 says that in the limit, all the q sub-populations become spatially inseparable. This causes all the mechanisms in the limiting process to appear in an *averaged* form.

Let $\{\mu^N(t) = (\mu_1^N(t), \ldots, \mu_q^N(t)) : t \geq 0\}$ be a Markov process with generator $\mathbf{A}_l^N$, for some $l \in \{0, 1, 2, 3\}$. It is difficult to prove the convergence of this process directly

because the density regulation mechanism acts on it at the fast timescale. This can be seen by splitting the operator $\mathbf{A}_i^N$ according to (2.31) and noting that $N\mathbf{R}^N$ becomes unbounded as $N \to \infty$. To pass to the limit we consider another measure-valued process $\{\nu^N(t) : t \geq 0\}$ that is constructed by suitably combining the various components of $\{\mu^N(t) : t \geq 0\}$. In particular

$$\nu^N(t) = \sum_{i=1}^{q} \Lambda_i(h^N(t))\mu_i^N(t) \ \text{ for } t \geq 0, \tag{2.44}$$

where $h^N(t) = H(\mu^N(t))$ and $\Lambda = (\Lambda_1, \ldots, \Lambda_q)$ is a function from $\mathbb{R}_+^q$ to $\mathbb{R}_+^q$ which satisfies certain conditions. These conditions are chosen to ensure that $\{\nu^N(t) : t \geq 0\}$ is a $\mathcal{P}(E)$-valued process whose dynamics is such that the density regulation mechanism acts at the slow timescale. Such a function Λ can be shown to exist by proving that a certain system of coupled partial differential equations has a solution with some desired properties. This is done in Section 4.3. We will then show that as $N \to \infty$ we have $\nu^N \Rightarrow \nu$ where $\{\nu(t) : t \geq 0\}$ is the Fleming-Viot process specified by Theorem 2.6. This convergence along with Proposition 2.5 allow us to prove Theorem 2.6. The details of the proof are given in Section 4.4.

The discussion in the preceding paragraph also shows that intuitively we can think of the limiting Fleming-Viot process as describing the spatial evolution of a *mixed* population formed by taking a suitable density-dependent linear combination of all the q sub-populations. This is reminiscent of the notion of *virtual* species (formed by linearly combining several chemical species), that are needed in the specification of the reduced models in chemical reaction networks with multiple timescales (see [5]).

3 Applications

In this section we discuss the applications mentioned in Section 1 in greater detail. Note that a Fleming-Viot process usually has continuous paths (see [13]). Hence Theorem 2.6 can be seen as a *diffusion approximation* result which shows that a stochastic process with jumps can be approximated by a process with continuous paths. Such results provide a justification for drawing inferences about the original process (with jumps) by analyzing a more tractable process with continuous paths.

To demonstrate the usefulness of Theorem 2.6 we present two examples. In the first example we consider a population genetics model having logistic interactions along with *rare mutation* and *weak selection*. The words *rare* and *weak* indicate that the mutation and selection events occur at a slower timescale than other events. The difference between this model and a standard population genetics model (Wright-Fisher or Moran) is that the population size is not fixed but fluctuating due to the logistic interactions. Theorem 2.6 guarantees that by taking the infinite population limit in a suitable way, we obtain a Fleming-Viot process. In many cases this limiting process is well-studied and using its properties one can estimate fixation probabilities, fixation times and the stationary distribution for the finite population model. Our second example sheds light on the phenomenon of cell polarity which refers to the spatial crowding of molecules on the cell membrane. We draw upon our work in [19] to show that Fleming-Viot convergence can help us understand how cells establish and maintain polarity. In [19] we only consider a very simple model, but the results in this paper ensure that the same analysis holds for a general class of models.

3.1 Logistic model for population genetics

The logistic growth model is very popular in ecology. It was proposed by Verhulst [37] in 1838 to describe the growth of a population in the presence of competition for

resources. In this model each individual reproduces at rate β and dies at a rate $\rho P/N$, where P is the current population size and N is the *carrying capacity* of the habitat. In the deterministic setting, the population size (P) evolves as a function of time (t) according to the ordinary differential equation

$$\frac{dP}{dt} = \beta P - \rho \frac{P^2}{N}.$$

Let $h(t) = P(t)/N$ be the *population density* at time t. Then the above differential equation becomes

$$\frac{dh}{dt} = \beta h - \rho h^2. \tag{3.1}$$

It is immediate that if $h(0) > 0$ then $h(t) \to h_{\mathrm{eq}} := \beta/\rho$ as $t \to \infty$.

We now construct a population genetics model that has logistic interactions along with *rare mutation* and *weak selection*. Suppose the compact metric space E is the set of all the genetic traits that an individual can have. Each individual is given a mass of $1/N$, with N being the carrying capacity as before. The population at time t can be represented by the measure

$$\bar{\mu}^N(t) = \frac{1}{N} \sum_{i=1}^{n^N(t)} \delta_{x_i}, \tag{3.2}$$

where $n^N(t)$ is the number of individuals at time t and $x_1, x_2, \cdots \in E$ are their genetic traits. Let b^s be a continuous function from E to $\mathbb{R}_+$. When the population density (total mass) is h, an individual with trait $x \in E$ gives birth at rate $(\beta + b^s(x)/N)$ and dies at rate ρh. Its offspring has the same trait x with probability $(1 - p(x)/N)$. However with probability $p(x)/N$, the offspring is a *mutant* and its trait is chosen according to the distribution $\vartheta(x, \cdot) \in \mathcal{P}(E)$. The process $\{\bar{\mu}^N(t) : t \geq 0\}$ can be viewed as a Markov process with state space $\mathcal{M}_{N,a}(E)$ (see (2.1)). The timescale at which we have described the dynamics is such that the mutation and selection events will vanish in the limit $N \to \infty$. Therefore to study their effects, we must observe the process at the timescale which is N times slower. Let

$$\mu^N(t) = \bar{\mu}^N(Nt) \text{ for } t \geq 0.$$

The dynamics of $\{\mu^N(t) : t \geq 0\}$ has fast density regulation along with position-dependent birth (see Section 2.3.2) and offspring dispersal mechanism (see Section 2.3.3). Assuming that $\mu^N(0) \Rightarrow \mu(0)$ as $N \to \infty$ and $\langle 1_E, \mu(0) \rangle > 0$ a.s. Theorem 2.6 gives us the following. If t_N is a sequence satisfying $t_N \to 0$ and $Nt_N \to \infty$, then the process $\mu^N(\cdot + t_N) \Rightarrow h_{\mathrm{eq}}\nu$ as $N \to \infty$, where $h_{\mathrm{eq}} = \beta/\rho$ and $\{\nu(t) : t \geq 0\}$ is a Fleming-Viot process with generator given by

$$\mathbf{A}F(\nu) = \beta \sum_{l=1}^{m} \left[\int_E p(x) \left(\int_E (f_l(y) - f_l(x))\vartheta(x, dy) \right) \nu(dx) \right] \prod_{j \neq l} \langle f_j, \nu \rangle \tag{3.3}$$

$$+ \sum_{l=1}^{m} \left(\langle b^s f_l, \nu \rangle - \langle b^s, \nu \rangle \langle f_l, \nu \rangle \right) \prod_{j \neq l} \langle f_j, \nu \rangle$$

$$+ \rho \sum_{1 \leq l \neq k \leq m} \left(\langle f_l f_k, \nu \rangle - \langle f_l, \nu \rangle \langle f_k, \nu \rangle \right) \prod_{j \neq l,k} \langle f_j, \nu \rangle$$

for any $F(\nu) = \prod_{l=1}^{m} \langle f_l, \nu \rangle$ where $f_1, \ldots, f_m \in B(E)$. This is of course the generator of a Fleming-Viot process with mutation and fecundity selection. We now present a couple of cases where this process is well-studied.

Suppose that $E = \{1, \ldots, K\}$ for some $K \in \mathbb{N}$. Then for any $t \geq 0$ we can express $\nu(t) \in \mathcal{P}(E)$ as the K-tuple $(\nu_1(t), \ldots, \nu_K(t))$, where $\nu_i(t)$ is the proportion of individuals having genetic trait $i \in E$. This representation allows us to view $\{\nu(t) : t \geq 0\}$ as a process over the K-simplex

$$\Delta_K = \left\{ (x_1, \ldots, x_K) : \ x_i \geq 0 \text{ and } \sum_{i=1}^{K} x_i = 1 \right\}.$$

For all $i, j \in E$ set $\theta_{ij} = \beta p(i) \vartheta(i, \{j\})$ and $\alpha_i = b^s(i)$. Then $\{\nu(t) : t \geq 0\}$ is a *diffusion* process over Δ_K with generator given by

$$\mathbf{A}f(\nu) = \rho \sum_{i,j \in E} \nu_i(\delta_{ij} - \nu_j) \frac{\partial^2 f(\nu)}{\partial \nu_i \partial \nu_j} + \sum_{j \in E} \left(\sum_{i \in E} (\theta_{ij} \nu_i + \nu_j \alpha_i (\delta_{ij} - \nu_i)) \right) \frac{\partial f(\nu)}{\partial \nu_j}$$

where $f \in C^2(\mathbb{R}_+^K, \mathbb{R})$, $\nu = (\nu_1, \ldots, \nu_K)$ and δ_{ij} is the Kronecker delta function. This is the generator of the *Wright-Fisher* diffusion process [25]. Many explicit results about the fixation probabilities, fixation times and the stationary distribution can be found in [14].

Let us return to the situation where E is a general compact metric space and the dynamics evolves according to (3.3). Assume that for all $x \in E$ we have $b^s(x) = 0$, $p(x) = 1$ and $\vartheta(x, \cdot) = \vartheta_0(\cdot)$, for some non-atomic probability measure $\vartheta_0 \in \mathcal{P}(E)$. The resulting Fleming-Viot process $\{\nu(t) : t \geq 0\}$ arises as a reformulation of the infinitely-many-neutral-alleles model due to Kimura and Crow [26] (see [12] and Section 9.2 in [13] for more details). In this case, $\nu(t)$ can be written as a countable sum $\sum_{i=1}^{\infty} a_i \delta_{x_i}$ for any $t > 0$ (see Theorem 7.2 in [13]). This means that at time t, a_i fraction of the population is located at x_i. Arranging these a_i-s in descending order we can extract a process over the ordered infinite simplex

$$\hat{\Delta}_\infty = \left\{ (x_1, x_2, \ldots) : \ x_1 \geq x_2 \cdots \geq 0 \text{ and } \sum_{i=1}^{\infty} x_i = 1 \right\}.$$

This extracted process is a diffusion process over $\hat{\Delta}_\infty$ whose various properties are presented in [11]. Furthermore in [13] it is shown that the Fleming-Viot process $\{\nu(t) : t \geq 0\}$ is ergodic and its unique stationary distribution $\Pi \in \mathcal{P}(\mathcal{P}(E))$ is given by

$$\Pi(S) = \mathbb{P}\left(\sum_{i=1}^{\infty} \phi_i \delta_{\xi_i} \in S \right) \text{ for all } S \in \mathcal{B}(\mathcal{P}(E)), \tag{3.4}$$

where the infinite vector $(\phi_1, \phi_2, \ldots)$ has the Poisson-Dirichlet distribution with parameter $\beta/2\rho$ and $\xi_1, \xi_2, \ldots$ are i.i.d. with distribution ϑ_0, independent of $(\phi_1, \phi_2, \ldots)$. The Poisson-Dirichlet distribution was introduced and studied by Kingman [27] in 1975.

The results mentioned in the last two paragraphs indicate the behaviour of the *evolutionary* dynamics under our original model for large values of N.

3.2 Cell polarity

Cell polarity is an important phenomenon and understanding the mechanisms responsible for it is a matter of fundamental concern for biologists. It is widely accepted that polarity is established in 3 stages (see [10, 2, 33]), which can be described as follows:

1. An unpolarized cell receives a spatial cue that may be *intrinsic* (coming from inside the cell) or *extrinsic* (coming from the extracellular environment).

2. This cue is interpreted by the membrane-bound receptor molecules.

3. The feedback network inside the cell is activated, which amplifies the weak initial signal into a robust signal that can direct the molecules towards the clustering site.

The feedback network has two components : *positive feedback* which enables the membrane molecules to pull the cytosol molecules to their location on the membrane, and *negative feedback* that pushes the membrane molecules into the cytosol. Positive feedback is responsible for the localized recruitment of molecules on the membrane while negative feedback helps in regulating the population size on the membrane. The molecules diffuse slowly on the membrane but rapidly in the cytosol.

Even though the feedback mechanism may bring the molecules together on the membrane, any clusters that form may not persist due to spatial diffusion. This caused some biologists to propose that other additional mechanisms are needed to generate spatial asymmetry (see [18, 20]), but these mechanisms are not always found in cells that exhibit polarity. Hence it is important to investigate if the feedback mechanism can alone counter spatial diffusion to establish cell polarity. For this purpose, Altschuler et. al. [2] formulated a simple model based on the mechanisms mentioned above. We now describe their model. Consider the cell to be a sphere of radius R in $\mathbb{R}^3$. The whole cell has N molecules which may be present on the membrane or in the cytosol. The following four mechanisms change the configuration of molecules in the cell.

1. *Association mechanism:* Each molecule in the cytosol can move to a uniformly chosen location on the membrane at rate k_{on}.

2. *Positive feedback:* Each molecule on the membrane pulls another molecule from the cytosol to its location at rate $k_{\mathrm{fb}} \times$ (fraction of molecules in the cytosol).

3. *Negative feedback:* Each molecule on the membrane is pushed into the cytosol at rate k_{off}.

4. *Spatial migration:* Each membrane molecule is constantly diffusing on the membrane according to an independent Brownian motion with diffusion rate D.

The association mechanism provides the initial spatial cue to trigger cluster formation. In [2], this spatial cue is intrinsic because the authors are concerned with *spontaneous* cell polarity, which means that polarity is established without any extracellular influence. Hence the association mechanism acts uniformly on the membrane. When one wants to consider polarity that is established in response to a chemical gradient (see [39]) then a molecule *associating* itself to the membrane must choose its location according to some distribution that encodes the gradient information. We mentioned in Section 1, that the positive feedback mechanism is like a birth process, where the pulled cytosol molecule is the *offspring* of the recruiting membrane molecule. This introduces genealogical relationships between the membrane molecules. A set of membrane molecules are said to belong to a *clan* if they have a common ancestor. Note that when the diffusion rate (D) is small, we would expect the clan members to be huddled together.

The analysis of the above model in [2] gives some interesting results. When the dynamics is described deterministically, using a reaction-diffusion partial differential equation, then the model fails to capture cell polarity. However in the stochastic setting, the model does predict the formation of clusters in certain parameter regimes, when the number of molecules (N) is small. This result is proved by showing that the number of clans on the membrane drops to 1 at certain times. For small D, one would observe a cluster at these times. However the frequency of these events is proportional to N^{-1},

which indicates that polarity cannot occur in the large population limit $N \to \infty$, unless other mechanisms are present.

In [19] we rigorously study this model under a different scaling of parameters. We multiply $k_{\rm fb}$ and $k_{\rm off}$ by N, leaving $k_{\rm on}$ and D unchanged. We keep track of the locations of the membrane molecules as well as their *clan identities*. A clan identity is a number in $[0,1]$ which is passed unaltered from the parent molecule to the offspring. A molecule that associates itself on the membrane is assigned a uniformly chosen clan identity in $[0,1]$. At any time, the molecules on the membrane that have the same clan identity should have a common ancestor and hence they must belong to the same clan. Note that here the number of types (q) is equal to 1 and the population density is the same as the fraction of cell molecules that are on the membrane. Let $E = \hat{E} \times [0,1]$, where $\hat{E}$ is the membrane (sphere of radius R in $\mathbb{R}^3$). When the number of molecules is N, the population dynamics is described by a $\mathcal{M}_{N,a}(E)$-valued process $\{\mu^N(t) : t \geq 0\}$ as before. For any $h \in \mathbb{R}_+$ let $\beta(h) = k_{\rm fb}(1-h)$, $\rho(h) = k_{\rm off}$ and $\kappa(h) = k_{\rm on}(1-h)$. Let $\Theta \in \mathcal{P}(E)$ be the uniform distribution on E and let the spatial migration operator B (see Section 2.1) be $(D/2)\Delta$, where Δ denotes the Laplace-Beltrami operator on the sphere $\hat{E}$. For any $f : \hat{E} \times [0,1] \to \mathbb{R}$, Δ acts on f only as the function of the first coordinate. The operator $(D/2)\Delta$ is just the generator of the Brownian motion on $\hat{E}$ with diffusion rate D. With this notation one can verify that this model is a special case of the model in Section 2.3.4. The association mechanism is analogous to immigration while the feedback mechanism gives rise to the density regulation mechanism. Theorem 2.6 (see also Theorem 2.3 in [19]) shows that as $N \to \infty$ we have $\mu^N \Rightarrow h_{\rm eq}\nu$ where

$$h_{\rm eq} = 1 - \frac{k_{\rm off}}{k_{\rm fb}}$$

and $\{\nu(t) : t \geq 0\}$ is a $\mathcal{P}(E)$-valued Fleming-Viot process with generator

$$\mathbf{A}F(\nu) = \frac{D}{2} \sum_{l=1}^{m} \langle \Delta f_l, \nu \rangle \prod_{j \neq l} \langle f_j, \nu \rangle + \frac{k_{\rm off}}{h_{\rm eq}} \sum_{1 \leq l \neq k \leq m} (\langle f_l f_k, \nu \rangle - \langle f_l, \nu \rangle \langle f_k, \nu \rangle) \prod_{j \neq l,k} \langle f_j, \nu \rangle$$
$$+ k_{\rm on} \frac{(1 - h_{\rm eq})}{h_{\rm eq}} \sum_{l=1}^{m} \left(\int_E \int_E (f_l(y) - f_l(x)) \Theta(dy)\nu(dx) \right) \prod_{j \neq l} \langle f_j, \nu \rangle$$

for any $F(\nu) = \prod_{l=1}^{m} \langle f_l, \nu \rangle \in \mathcal{C}_0$. It can be shown that this Fleming-Viot process is ergodic and has a unique stationary distribution in $\mathcal{P}(\mathcal{P}(E))$ (see Section 5 in [13] and Proposition 2.5 in [19]). To study the evolution of the clan sizes we define a $\mathcal{P}([0,1])$-valued process $\{\nu_c(t) : t \geq 0\}$ by

$$\nu_c(t, S) = \nu(t, \hat{E} \times S), \ S \in \mathcal{B}([0,1]).$$

This is a Fleming-Viot process that describes the infinitely-many-neutral-alleles model (recall the discussion in Section 3.1). Therefore for any $t > 0$, we can write $\nu_c(t) = \sum_{i=1}^{\infty} a_i \delta_{x_i}$, which means that a_i fraction of the population has clan identity x_i. At stationarity, the clan sizes (arranged in descending order) are distributed according to the Poisson-Dirichlet distribution with parameter $\alpha = k_{\rm on}/k_{\rm fb}$. Properties of the Poisson-Dirichlet distribution (see [15]) tell us that for any small $\epsilon > 0$, there is a positive probability of the largest clan having size greater than $(1 - \epsilon)$. Furthermore one can show that at stationarity the molecules in each clan are concentrated on a circular patch on the membrane. The square of the radius of this patch can be approximately computed as (see Theorem 2.7 in [19])

$$\frac{2D}{\left(\frac{(k_{\rm on} + k_{\rm fb})k_{\rm off}}{(k_{\rm fb} - k_{\rm off})} + \frac{D}{R^2} \right)}.$$

The last two assertions imply that if D is small in comparison to R^2, then at stationarity there is a positive probability that most of the membrane molecules are in one clan and that clan is spread over a small area on the membrane. Due to ergodicity this event will occur infinitely often in any trajectory of the process $\{\nu(t) : t \geq 0\}$. Whenever this event happens we can expect the cell to be *polarized*. Therefore the limiting process exhibits recurring cell polarity. In [19] we discuss how the frequency of observing polarity depends on various model parameters.

The above analysis shows that if the feedback mechanism is strong enough, it can counter spatial diffusion to generate cell polarity. However this conclusion is based on a highly simplified model. As mentioned in Section 1, most cells that exhibit polarity have complicated feedback circuits, with molecules of several types pulling each other on and off the membrane at various type-dependent rates. These different types of molecules may also have their own migration and association mechanisms. It would be interesting to know if the above analysis can be extended to general multi-type models for cell polarity. The results in this paper show that this can indeed be done as long as the feedback mechanism satisfies the assumptions in Section 2.2, and acts at a faster timescale than the association and migration mechanisms. In this case, Theorem 2.6 guarantees convergence to a Fleming-Viot process and this limiting process can then be analyzed in the same way as in [19]. This enables us to draw similar conclusions about the onset of cell polarity in this multi-type setting.

4 Proofs

4.1 Well-posedness of the martingale problems for $\mathbf{A}_l^N$

Recall the definitions of the operators $\mathbf{A}_0^N, \mathbf{A}_1^N, \mathbf{A}_2^N$ and $\mathbf{A}_3^N$ from (2.22), (2.23), (2.24) and (2.26). In this section we prove Proposition 2.4 which says that the martingale problem for these operators is well-posed in the space $D_{\mathcal{M}_{N,a}^q(E)}[0,\infty)$. Pick a $l \in \{0,1,2,3\}$. If we do not allow the dynamics under $\mathbf{A}_l^N$ to leave a compact set of $\mathcal{M}_{N,a}^q(E)$, then $\mathbf{A}_l^N$ can be viewed as a bounded perturbation of the migration operator $\mathbf{B}^N$ (given by (2.12)). The well-posedness of the corresponding martingale problem is immediate from the well-posedness of the martingale problem for $\mathbf{B}^N$ (see Chapter 4 in Ethier and Kurtz [12]). In our case, the dynamics under $\mathbf{A}_l^N$ may exit any compact set of $\mathcal{M}_{N,a}^q(E)$. However we can still argue the well-posedness of the corresponding martingale problem by showing that this exit time tends to infinity as the compact set gets bigger and bigger in size. We now make these ideas precise.

Lemma 4.1. *Fix a $l \in \{0,1,2,3\}$, $N \in \mathbb{N}$ and $\pi \in \mathcal{P}\left(\mathcal{M}_{N,a}^q(E)\right)$. For each $k \in \mathbb{N}$ let $\{\mu_k(t) : t \geq 0\}$ be a $\mathcal{M}_{N,a}^q(E)$-valued process with initial distribution π. Define a stopping time*

$$\tau_k = \inf\left\{t \geq 0 : \|H(\mu_k(t-))\|_1 \geq k \text{ or } \|H(\mu_k(t))\|_1 \geq k\right\}, \tag{4.1}$$

where H is the density map (2.2). Suppose that for each $k \in \mathbb{N}$ and $\hat{F} \in \mathcal{C}_0^q$

$$\hat{F}(\mu_k(t \wedge \tau_k)) - \hat{F}(\mu_k(0)) - \int_0^{t \wedge \tau_k} \mathbf{A}_l^N \hat{F}(\mu_k(s))ds$$

is a martingale. Then for any $t \geq 0$

$$\lim_{k \to \infty} \mathbb{P}\left(\tau_k \leq t\right) = 0.$$

Proof. Let $\{h_k(t) = H(\mu_k(t)) : t \geq 0\}$ be the density process corresponding to μ_k and let $c_k : \mathbb{R}_+^q \to \mathbb{R}_+^q$ be the function defined by

$$c_k(h) = (c_{k,1}(h), \ldots, c_{k,q}(h)) = \begin{cases} \bar{1}_q & \text{if } \|h\|_1 < 2k \\ \bar{0}_q & \text{otherwise .} \end{cases} \tag{4.2}$$

Pick any $\epsilon \in (0, 1)$. Since for each $k \in \mathbb{N}$ the distribution of $\mu_k(0)$ is π there must exist a $k_\epsilon > 0$ such that

$$\mathbb{P}\left(\|h_k(0)\|_1 > k_\epsilon\right) < \epsilon \text{ for all } k \in \mathbb{N}.$$

If k satisfies $k_\epsilon \leq k\epsilon$ then

$$\mathbb{E}\left(\langle c_k(h_k(0)), h_k(0)\rangle\right) \leq k_\epsilon \mathbb{P}\left(\|h_k(0)\|_1 \leq k_\epsilon\right) + 2k\mathbb{P}\left(\|h_k(0)\|_1 > k_\epsilon\right) \leq 3\epsilon k. \tag{4.3}$$

For each $i \in Q$ and $k \in \mathbb{N}$, define a function $\hat{F}_i^k \in \mathcal{C}_0^q$ by $\hat{F}_i^k(\mu) = c_{k,i}(h)h_i$ where $h = H(\mu)$. From (2.31) we know that for any $\mu \in \mathcal{M}_{N,a}^q(E)$

$$\mathbf{A}_l^N \hat{F}_i^k(\mu) = \mathbf{B}^N \hat{F}_i^k(\mu) + N\mathbf{R}^N \hat{F}_i^k(\mu) + \mathbf{G}_l^N \hat{F}_i^k(\mu).$$

One can easily verify that $\mathbf{B}^N \hat{F}_i^k(\mu) = 0$, $\mathbf{R}^N \hat{F}_i^k(\mu) = c_{k,i}(h)\theta_i(h)$, $\mathbf{G}_0^N \hat{F}_i^k(\mu) = 0$, $\mathbf{G}_2^N \hat{F}_i^k(\mu) = 0$, $\mathbf{G}_3^N \hat{F}_i^k(\mu) = c_{k,i}(h)\kappa_i(h)$ and finally

$$\mathbf{G}_1^N \hat{F}_i^k(\mu) = c_{k,i}(h)\left(\sum_{j \in Q} \int_E b_{ji}^s(x, h)\mu_j(dx) - \int_E d_i^s(x, h)\mu_i(dx)\right).$$

Note that for each $i, j \in Q$, the functions b_{ji}^s, d_i^s are bounded, while the functions θ_i and κ_i satisfy (2.16) and (2.25). This implies that there exists a positive constant C (depending on N and l) such that

$$\mathbf{A}_l^N \hat{F}_i^k(\mu) \leq C\left(1 + \langle c_k(h), h\rangle\right) \text{ for all } \mu \in \mathcal{M}_{N,a}^q(E). \tag{4.4}$$

By the assumption stated in the statement of this lemma we can say that

$$\hat{F}_i^k(\mu_k(t \wedge \tau_k)) - \hat{F}_i^k(\mu_k(0)) - \int_0^{t \wedge \tau_k} \mathbf{A}_l^N \hat{F}_i^k(\mu_k(s))ds$$

is a martingale starting at 0. Taking expectations we get

$$\mathbb{E}\left(\hat{F}_i^k(\mu_k(t \wedge \tau_k))\right) = \mathbb{E}\left(\hat{F}_i^k(\mu_k(0))\right) + \mathbb{E}\left(\int_0^{t \wedge \tau_k} \mathbf{A}_l^N \hat{F}_i^k(\mu_k(s))ds\right). \tag{4.5}$$

Let $\hat{F}^k : \mathcal{M}_F^q(E) \to \mathbb{R}$ be given by $\hat{F}^k(\mu) = \sum_{i \in Q} \hat{F}_i^k(\mu) = \langle c_k(h), h\rangle$. Then summing over $i \in Q$ in (4.5) and using (4.4) we arrive at

$$\mathbb{E}\left(\hat{F}^k(\mu_k(t \wedge \tau_k))\right) \leq \mathbb{E}\left(\hat{F}^k(\mu_k(0))\right) + Cq\int_0^t \left[1 + \mathbb{E}\left(\hat{F}^k(\mu_k(s \wedge \tau_k))\right)\right]ds.$$

From (4.3) and Gronwall's inequality, for $k \geq k_\epsilon/\epsilon$ we obtain

$$\mathbb{E}\left(\hat{F}^k(\mu_k(t \wedge \tau_k))\right) \leq (3k\epsilon + Cqt)e^{Cqt}.$$

Then by Markov's inequality

$$\begin{aligned} \lim_{k \to \infty} \mathbb{P}\left(\hat{F}^k(\mu_k(t \wedge \tau_k)) \geq k\right) &\leq \lim_{k \to \infty} \frac{\mathbb{E}\left(\hat{F}^k(\mu_k(t \wedge \tau_k))\right)}{k} \\ &\leq \lim_{k \to \infty} \frac{(3k\epsilon + Cqt)e^{Cqt}}{k} \\ &\leq 3\epsilon e^{Cqt}. \end{aligned} \tag{4.6}$$

Observe that

$$\mathbb{P}\left(\tau_k \leq t\right) = \mathbb{P}\left(\|h_k(t \wedge \tau_k)\|_1 \geq k\right)$$
$$= \mathbb{P}\left(\hat{F}^k(\mu_k(t \wedge \tau_k)) \geq k\right) + \mathbb{P}\left(\|h_k(t \wedge \tau_k)\|_1 \geq 2k\right).$$

For large k, the second probability on the right is 0 because of the following reason. The process h_k has jumps of size $1/N$ and hence the definition of τ_k (see (4.1)) implies that $\|h_k(t \wedge \tau_k)\|_1 \leq k + (1/N) < 2k$. Therefore using (4.6) we get

$$\lim_{k \to \infty} \mathbb{P}\left(\tau_k \leq t\right) = \lim_{k \to \infty} \mathbb{P}\left(\hat{F}^k(\mu_k(t \wedge \tau_k)) \geq k\right) \leq 3\epsilon e^{Cqt}.$$

Letting $\epsilon \to 0$ proves the lemma. $\qquad\square$

Proof of Proposition 2.4. Fix a $N \in \mathbb{N}$ and a $l \in \{0, 1, 2, 3\}$. The space $\mathcal{M}^q_{N,a}(E)$ is complete and separable and for each $k \in \mathbb{N}$ the set

$$U_k = \left\{\mu \in \mathcal{M}^q_{N,a}(E) : \|H(\mu)\|_1 < k\right\} \tag{4.7}$$

is open with a compact closure in $\mathcal{M}^q_F(E)$. Define an operator $\mathbf{L}_k : \mathcal{D}(\mathbf{L}_k) = \mathcal{C}^q_0 \to B\left(\mathcal{M}^q_{N,a}(E)\right)$ by

$$\mathbf{L}_k \hat{F}(\mu) = \mathbf{B}^N \hat{F}(\mu) + \mathbb{1}_{U_k}(\mu)\left(N\mathbf{R}^N \hat{F}(\mu) + \mathbf{G}^N_l \hat{F}(\mu)\right),$$

for any $\hat{F} \in \mathcal{C}^q_0$. The operator $\mathbf{L}_k$ can be seen as a bounded perturbation of the operator $\mathbf{B}^N$. We argued in Section 2.1 that the martingale problem for $\mathbf{B}^N$ is well-posed. From Theorem 4.10.3 in Ethier and Kurtz [12], the martingale problem for $\mathbf{L}_k$ is well-posed for each $k \in \mathbb{N}$. Pick a $\pi \in \mathcal{P}\left(\mathcal{M}^q_{N,a}(E)\right)$ and let $\{\mu_k(t) : t \geq 0\}$ be the unique solution to the martingale problem for $(\mathbf{L}_k, \pi)$. Define a stopping time by

$$\tau_k = \inf\{t \geq 0 : \mu_k(t) \notin U_k \text{ or } \mu_k(t-) \notin U_k\}.$$

Then for any $\hat{F} \in \mathcal{C}^q_0$

$$\hat{F}(\mu_k(t)) - \hat{F}(\mu_k(0)) - \int_0^t \mathbf{L}_k \hat{F}(\mu_k(s)) ds$$

is a martingale. From (2.31) one can see that if $\mu \in U_k$ then $\mathbf{A}^N_l \hat{F}(\mu) = \mathbf{L}_k \hat{F}(\mu)$. Using the optional sampling theorem we get that

$$\hat{F}(\mu_k(t \wedge \tau_k)) - \hat{F}(\mu_k(0)) - \int_0^{t \wedge \tau_k} \mathbf{L}_k \hat{F}(\mu_k(s)) ds$$
$$= \hat{F}(\mu_k(t \wedge \tau_k)) - \hat{F}(\mu_k(0)) - \int_0^{t \wedge \tau_k} \mathbf{A}^N_l \hat{F}(\mu_k(s)) ds$$

is a martingale. Lemma 4.1 ensures that for any $t \geq 0$,

$$\lim_{k \to \infty} \mathbb{P}(\tau_k \leq t) = 0.$$

From Theorem 4.6.3 in Ethier and Kurtz [12] we can conclude that there exists a unique solution to the martingale problem for $(\mathbf{A}^N_l, \pi)$. $\qquad\square$

4.2 Properties of the limiting process

The goal of this section is to prove Proposition 2.5 which gives important insights into the limiting behaviour of the dynamics under the models described in Section 2.3. As mentioned in Section 2.4 our proof of Proposition 2.5 will rely on the work of Katzenberger [23] which studies how semimartingales behave when they are driven by a fast drift that forces them to stay on a stable manifold. Before we can use the framework in [23] we need to prove some preliminary results. We start by recalling a tightness condition for semimartingales.

Condition 4.2. *Let $\{Z^N : N \in \mathbb{N}\}$ be a sequence of $\{\mathcal{F}_t^N\}$-semimartingales with paths in $D_{\mathbb{R}^d}[0, \infty)$. Assume that for all $T > 0$*

$$\sup_{0 \le t \le T} \left\| Z^N(t) - Z^N(t-) \right\| \Rightarrow 0 \text{ as } N \to \infty. \tag{4.8}$$

Moreover assume that each semimartingale Z^N can be written as

$$Z^N(t) = M^N(t) + \int_0^t A^N(s) ds$$

where M^N is a square integrable $\{\mathcal{F}_t^N\}$-martingale and A^N is a $\{\mathcal{F}_t^N\}$ adapted process satisfying

$$\sup_{N \in \mathbb{N}} \mathbb{E} \left([M^N]_t + \int_0^t \left\| A^N(s) \right\| ds \right) < \infty$$

for each $t \ge 0$, where $[M^N]_t$ is the quadratic variation of the martingale M^N.

Remark 4.3. *If a sequence of semimartingales $\{Z^N : N \in \mathbb{N}\}$ satisfies Condition 4.2 then this sequence is tight in $D_{\mathbb{R}^d}[0, \infty)$ in the Skorohod topology (see Corollary 2.3.3 in Joffe and Metivier [21]) and any limit point Z is a semimartingale with continuous sample paths.*

Lemma 4.4. *Pick a $l \in \{0, 1, 2, 3\}$, $i \in Q$ and $f \in \mathcal{D}_0$. For each $N \in \mathbb{N}$, let $\{\mu^N(t) : t \ge 0\}$ be a solution of the martingale problem for $\mathbf{A}_l^N$. Define a real-valued process $\{Z^N(t) : t \ge 0\}$ by*

$$Z^N(t) = \langle f, \mu_i^N(t) \rangle - \langle f, \mu_i^N(0) \rangle$$
$$- N \int_0^t \left[\sum_{j \in Q} \beta_{ji}(h^N(s)) \langle f, \mu_j^N(s) \rangle - \rho_i(h^N(s)) \langle f, \mu_i^N(s) \rangle \right] ds,$$

where $h^N(t) = H(\mu^N(t))$. Then Z^N is a semimartingale with respect to the filtration generated by $\{\mu^N(t) : t \ge 0\}$. For any compact $K \subset \mathbb{R}_+^q$ define

$$\lambda^N(K) = \inf \left\{ t \ge 0 : h^N(t-) \notin \overset{\circ}{K} \text{ or } h^N(t) \notin \overset{\circ}{K} \right\}, \tag{4.9}$$

where $\overset{\circ}{K}$ denotes the interior of the set K. If Z_K^N is the semimartingale given by $Z_K^N(t) = Z^N(t \wedge \lambda^N(K))$ for $t \ge 0$, then the sequence of semimartingales $\{Z_K^N : N \in \mathbb{N}\}$ satisfies Condition 4.2.

Proof. Let $i \in Q$ and $f \in \mathcal{D}_0$ be fixed. For each $N \in \mathbb{N}$ and $l \in \{0, 1, 2, 3\}$ define a

function $a_l^N : \mathcal{M}_F^q(E) \to \mathbb{R}$ by

$$a_0^N(\mu) = \langle B_i f, \mu_i \rangle,$$

$$a_1^N(\mu) = \langle B_i f, \mu_i \rangle + \sum_{j \in Q} \int_E b_{ji}^s(x,h) f(x) \mu_j(dx) - \int_E d_i^s(x,h) f(x) \mu_i(dx),$$

$$a_2^N(\mu) = \langle B_i f, \mu_i \rangle + N \sum_{j \in Q} \int_E \beta_{ji}(h) p_{ji}^N(x) \left[\int_E (f(y) - f(x)) \vartheta_{ji}^N(x,dy) \right] \mu_j(dx)$$

and $a_3^N(\mu) = \langle B_i f, \mu_i \rangle + \kappa_i(h) \int_E f(x) \Theta_i(dx),$

where $h = H(\mu)$. Let $U_k^N \subset \mathcal{M}_{N,a}^q(E)$ be given by

$$U_k^N = \left\{ \mu \in \mathcal{M}_{N,a}^q(E) : \|H(\mu)\|_1 < 2k \right\}. \tag{4.10}$$

Then for each $k \in \mathbb{N}$ and $l \in \{0,1,2,3\}$ we have

$$\sup_{N \in \mathbb{N}} \sup_{\mu \in U_k^N} a_l^N(\mu) < \infty. \tag{4.11}$$

To see this note that for any $N \in \mathbb{N}$, $U_k^N \subset U_k := \{\mu \in \mathcal{M}_F^q(E) : \|H(\mu)\|_1 \leq 2k\}$ and U_k is a compact subset of $\mathcal{M}_F^q(E)$. For $l \in \{0,1,3\}$, a_l^N is a continuous function which does not depend on N and hence we get (4.11) simply by observing that

$$\sup_{N \in \mathbb{N}} \sup_{\mu \in U_k^N} a_l^N(\mu) \leq \sup_{\mu \in U_k} a_l^N(\mu) < \infty. \tag{4.12}$$

Similarly if we define a continuous function $\hat{a}_2 : \mathcal{M}_F^q(E) \to \mathbb{R}$ by

$$\hat{a}_2(\mu) = \langle B_i f, \mu_i \rangle + \sum_{j \in Q} \beta_{ji}(h) \langle C_{ji} f, \mu_j \rangle$$

then we also have

$$\sup_{N \in \mathbb{N}} \sup_{\mu \in U_k^N} \hat{a}_2(\mu) \leq \sup_{\mu \in U_k} \hat{a}_2(\mu) < \infty. \tag{4.13}$$

Here C_{ji}'s are the operators satisfying Assumption 2.2. This assumption also implies that

$$\sup_{N \in \mathbb{N}} \sup_{\mu \in U_k^N} \left| a_2^N(\mu) - \hat{a}_2(\mu) \right|$$

$$\leq \sup_{N \in \mathbb{N}} \sup_{\mu \in U_k^N} \sum_{j \in Q} \beta_{ji}(h) \int_E \left(N p_{ji}^N(x) \int_E (f(y) - f(x)) \vartheta_{ji}^N(x,dy) - C_{ji} f(x) \right) \mu_j(dx)$$

$$\leq \sup_{N \in \mathbb{N}} \sup_{\mu \in U_k^N} \sum_{j \in Q} \beta_{ji}(h) h_j \sup_{x \in E} \left(N p_{ji}^N(x) \int_E (f(y) - f(x)) \vartheta_{ji}^N(x,dy) - C_{ji} f(x) \right)$$

$$< \infty.$$

This bound along with (4.13) and the triangle inequality shows (4.11) for $l = 2$.

Let $c_k : \mathbb{R}_+^q \to \mathbb{R}_+^q$ be given by (4.2). Define $\hat{F}_k : \mathcal{M}_F^q(E) \to \mathbb{R}$ by

$$\hat{F}_k(\mu) = c_{k,i}(h) \langle f, \mu_i \rangle,$$

where $h = H(\mu)$. One can verify that for any $\mu \in U_k^N$

$$\left(\mathbf{B}^N + \mathbf{G}_l^N\right)\hat{F}_k(\mu) = a_l^N(\mu) \tag{4.14}$$

$$\text{and } \mathbf{R}^N \hat{F}_k(\mu) = \sum_{j \in Q} \beta_{ji}(h)\langle f, \mu_j\rangle - \rho_i(h)\langle f, \mu_i\rangle. \tag{4.15}$$

Suppose $\{\mu^N(t) : t \geq 0\}$ solves the martingale problem for $\mathbf{A}_l^N$ and $\{h^N(t) = H(\mu^N(t)) : t \geq 0\}$ is the corresponding density process. Define another process $\{m^N(t) : t \geq 0\}$ by

$$m^N(t) = Z^N(t) - \int_0^t a_l^N(\mu^N(s))ds. \tag{4.16}$$

For any $k \in \mathbb{N}$ let

$$\tau_k^N = \inf\{t \geq 0 : \|h^N(t-)\|_1 \geq k \text{ or } \|h^N(t)\|_1 \geq k\}. \tag{4.17}$$

From Lemma 4.1 we can conclude that for any fixed N, the stopping times τ_k^N converge to ∞ a.s. as $k \to \infty$. Observe that $\hat{F}_k$ belongs to the class $\mathcal{C}_0^q = \mathcal{D}(\mathbf{A}_l^N)$. Hence

$$\hat{F}_k(\mu^N(t)) - \hat{F}_k(\mu^N(0)) - \int_0^t \mathbf{A}_l^N \hat{F}_k(\mu^N(s))ds$$

is a martingale. From (2.31) and the optional sampling theorem we get that

$$m_k^N(t) = \hat{F}_k(\mu^N(t \wedge \tau_k^N)) - \hat{F}_k(\mu^N(0)) - \int_0^{t \wedge \tau_k^N} \mathbf{A}_l^N \hat{F}_k(\mu^N(s))ds \tag{4.18}$$

$$= \hat{F}_k(\mu^N(t \wedge \tau_k^N)) - \hat{F}_k(\mu^N(0)) - N \int_0^{t \wedge \tau_k^N} \mathbf{R}^N \hat{F}_k(\mu^N(s))ds$$

$$- \int_0^{t \wedge \tau_k^N} \left(\mathbf{B}^N + \mathbf{G}_l^N\right)\hat{F}_k(\mu^N(s))ds$$

is also a martingale. If the set $(0, t \wedge \tau_k^N]$ is non-empty then for any $s \in (0, t \wedge \tau_k^N]$, we have $c_k(h^N(s)) = \bar{1}_q$ and therefore $\hat{F}_k(\mu^N(s)) = \langle f, \mu_i^N(s)\rangle$. If the set $(0, t \wedge \tau_k^N]$ is empty then $t \wedge \tau_k^N = 0$ and in this case $m_k^N(t) = 0$. From (4.14) and (4.15) we see that for all $t \geq 0$, $m_k^N(t) = m^N(t \wedge \tau_k^N)$, where m^N is defined by (4.16). But m_k^N is a martingale and for a fixed N, $\tau_k^N \to \infty$ a.s. as $k \to \infty$. Therefore we can conclude that m^N is local martingale and hence Z^N is a semimartingale.

Let $\hat{F}_k^2 : \mathcal{M}_F^q(E) \to \mathbb{R}$ be given by $\hat{F}_k^2(\mu) = (\hat{F}_k(\mu))^2$. Note that for any $\mu \in \mathcal{M}_F(E)$

$$\left\langle f, \mu \pm \frac{1}{N}\delta_x\right\rangle^2 - \langle f, \mu\rangle^2 = \pm 2\langle f, \mu\rangle\frac{f(x)}{N} + \frac{f^2(x)}{N^2}.$$

Using this one can verify that if $\mu \in U_k^N$ and $h = H(\mu)$ then we have

$$N\left(\mathbf{R}^N \hat{F}_k^2(\mu) - 2\hat{F}_k(\mu)\mathbf{R}^N \hat{F}_k(\mu)\right) = \sum_{j \in Q} \beta_{ji}(h)\langle f^2, \mu_j\rangle + \rho_i(h)\langle f^2, \mu_i\rangle,$$

$$\mathbf{G}_1^N \hat{F}_k^2(\mu) - 2\hat{F}_k(\mu)\mathbf{G}_1^N \hat{F}_k(\mu)$$
$$= \frac{1}{N}\sum_{j \in Q}\int_E b_{ji}^s(x, h)f^2(x)\mu_j(dx) + \int_E d_i^s(x, h)f^2(x)\mu_i(dx),$$

$$\mathbf{G}_2^N \hat{F}_k^2(\mu) - 2\hat{F}_k(\mu)\mathbf{G}_2^N \hat{F}_k(\mu)$$
$$= \sum_{j \in Q}\int_E \beta_{ji}(h)p_{ji}^N(x)\left(\int_E (f^2(y) - f^2(x))\vartheta_{ji}^N(x, dy)\right)\mu_j(dx)$$

$$\text{and } \mathbf{G}_3^N \hat{F}_k^2(\mu) - 2\hat{F}_k(\mu)\mathbf{G}_3^N \hat{F}_k(\mu) = \frac{1}{N}\kappa_i(h)\int_E f^2(x)\Theta_i(dx).$$

Also for any $\mu \in U_k^N$

$$\mathbf{B}^N \hat{F}_k^2(\mu) - 2\hat{F}_k(\mu)\mathbf{B}^N \hat{F}_k(\mu) = \frac{1}{N}\langle B_i f^2 - 2f B_i f, \mu_i \rangle.$$

Using (2.31) and the above expressions we can show in a manner similar to (4.11) that for any $k \in \mathbb{N}$ we have

$$\sup_{N \in \mathbb{N}} \sup_{\mu \in U_k^N} |\mathbf{A}_l^N \hat{F}_k^2(\mu) - 2\hat{F}_k(\mu)\mathbf{A}_l^N \hat{F}_k(\mu)| < \infty. \tag{4.19}$$

The function $\hat{F}_k^2$ is also in $\mathcal{C}_0^q = \mathcal{D}(\mathbf{A}_l^N)$. Therefore if m_k^N is the martingale given by (4.18) then

$$\left(m_k^N(t)\right)^2 - \int_0^{t \wedge \tau_k} \left(\mathbf{A}_l^N \hat{F}_k^2(\mu^N(s)) - 2\hat{F}_k(\mu^N(s))\mathbf{A}_l^N \hat{F}_k(\mu^N(s))\right) ds$$

is also a martingale. Therefore the expected quadratic variation of m_k^N can be computed as

$$\mathbb{E}\left([m_k^N]_t\right) = \mathbb{E}\left((m_k^N(t))^2\right) = \mathbb{E}\left(\int_0^{t \wedge \tau_k^N} \left(\mathbf{A}_l^N \hat{F}_k^2(\mu^N(s)) - 2\hat{F}_k(\mu^N(s))\mathbf{A}_l^N \hat{F}_k(\mu^N(s))\right) ds\right). \tag{4.20}$$

For any fixed $k \in \mathbb{N}$, the sequence of semimartingales $\{Z^N(\cdot \wedge \tau_k^N) : N \in \mathbb{N}\}$ satisfy (4.8) because the discontinuities of μ^N are of size proportional to $1/N$. From (4.16) we can see that the semimartingale $Z^N(\cdot \wedge \tau_k^N)$ can be decomposed as

$$Z^N(t \wedge \tau_k^N) = m^N(t \wedge \tau_k^N) + \int_0^{t \wedge \tau_k^N} a_l^N(\mu^N(s))ds = m_k^N(t) + \int_0^t \mathbb{1}_{\{s \leq \tau_k^N\}} a_l^N(\mu^N(s))ds.$$

For any $0 < s \leq \tau_k^N$, $\mu^N(s) \in U_k^N$. Using (4.20), (4.19) and (4.11) we can see that

$$\sup_{N \in \mathbb{N}} \mathbb{E}\left([m_k^N]_t + \int_0^t \mathbb{1}_{\{s \leq \tau_k^N\}} a_l^N(\mu^N(s))ds\right) < \infty. \tag{4.21}$$

Therefore for any $k \in \mathbb{N}$, the sequence of semimartingales $\{Z^N(\cdot \wedge \tau_k^N)\}$ satisfies Condition 4.2. For any compact set $K \subset \mathbb{R}_+^q$, there exists a k such that $K \subset \{h \in \mathbb{R}_+^q : \|h\|_1 < k\}$. If the stopping time λ_K^N is defined by (4.9) then $\lambda_K^N \leq \tau_k^N$ a.s. where τ_k^N is given by (4.17). Hence it is immediate that if Z_K^N is the semimartingale defined by $Z_K^N(\cdot) = Z^N(\cdot \wedge \lambda^N(K))$ then the sequence of semimartingales $\{Z_K^N : N \in \mathbb{N}\}$ will also satisfy Condition 4.2. $\qquad \square$

Proof of Proposition 2.5. Let $\{\mathcal{F}_t^N\}$ be the filtration generated by the process $\{\mu^N(t) : t \geq 0\}$. For any compact set $K \subset U_{\text{eq}}$ let $\lambda^N(K)$ be given by (4.9). In this proof a sequence of $\{\mathcal{F}_t^N\}$-semimartingales $\{Z^N : N \in \mathbb{N}\}$ with paths in $D_{\mathbb{R}^d}[0, \infty)$ will be called *well-behaved* if for each compact $K \subset U_{\text{eq}}$, the sequence of semimartingales $\{Z^N(\cdot \wedge \lambda^N(K))\}$ satisfies Condition 4.2.

Using Lemma 4.4 with $f = 1_E$, (where 1_E is as in (2.2)) for each $i \in Q$ we obtain a well-behaved $\mathbb{R}$-valued semimartingale $Z_i^{N,1}$ such that

$$\langle 1_E, \mu_i^N(t) \rangle = \langle 1_E, \mu_i^N(0) \rangle + N \int_0^t \left[\sum_{j \in Q} \beta_{ji}(h^N(s))\langle 1_E, \mu_j^N(s) \rangle - \rho_i(h^N(s))\langle 1_E, \mu_i^N(s) \rangle\right] ds$$

$$+ Z_i^{N,1}(t).$$

Recall the definition of the matrix $A(h)$ from (2.13). The above expression is the same as

$$h_i^N(t) = h_i^N(0)\rangle + N \int_0^t \left[\sum_{j \in Q} \beta_{ji}(h^N(s)) h_j^N(s) - \rho_i(h^N(s)) h_i^N(s) \right] ds + Z_i^{N,1}(t)$$

$$= h_i^N(0) + N \int_0^t \left[\sum_{j \in Q} A_{ij}(h^N(s)) h_j^N(s) \right] ds + Z_i^{N,1}(t).$$

If we let $Z^{N,1}$ to be the $\mathbb{R}^q$-valued semimartingale given by

$$Z^{N,1}(t) = (Z_1^{N,1}(t), \ldots, Z_q^{N,1}(t)) \text{ for } t \geq 0,$$

then $Z^{N,1}$ is also a well-behaved semimartingale. The semimartingale h^N satisfies

$$h^N(t) = h^N(0) + N \int_0^t A(h^N(s)) h^N(s) ds + Z^{N,1}(t)$$

$$= h^N(0) + N \int_0^t \theta(h^N(s)) ds + Z^{N,1}(t), \tag{4.22}$$

where the last equality holds due to definition (2.14). Now fix a $f \in \mathcal{D}_0$. From Lemma 4.4, for each $i \in Q$ there is a well-behaved semimartingale $Z_i^{N,f}$ such that

$$\langle f, \mu_i^N(t) \rangle = \langle f, \mu_i^N(0) \rangle + N \left(\int_0^t \sum_{j \in Q} \beta_{ji}(h^N(s)) \langle f, \mu_j^N(s) \rangle ds - \int_0^t \rho_i(h^N(s)) \langle f, \mu_i^N(s) \rangle ds \right)$$

$$+ Z_i^{N,f}(t)$$

$$= \langle f, \mu_i^N(0) \rangle + N \int_0^t \left[\sum_{j \in Q} A_{ij}(h^N(s)) \langle f, \mu_j^N(s) \rangle \right] ds + Z_i^{N,f}(t).$$

Using the integration by parts formula for semimartingales, for each $i, j \in Q$ we can write

$$h_i^N(t) \langle f, \mu_j^N(t) \rangle = h_i^N(0) \langle f, \mu_j^N(0) \rangle + N \left(\int_0^t h_i^N(s) \left[\sum_{k \in Q} A_{jk}(h^N(s)) \langle f, \mu_k^N(s) \rangle \right] ds \right.$$

$$\left. + \int_0^t \langle f, \mu_j^N(s) \rangle \left[\sum_{k \in Q} A_{ik}(h^N(s)) h_k^N(s) \right] ds \right) + Z_{ij}^N(t), \tag{4.23}$$

where Z_{ij}^N is another well-behaved semimartingale given by

$$Z_{ij}^N(t) = \int_0^t h_i^N(s) Z_j^{N,f}(s) ds + \int_0^t \langle f, \mu_j^N(s) \rangle Z_i^{N,1}(s) ds + [Z_i^{N,1}, Z_j^{N,f}]_t.$$

The last term in the above equation is the cross-variation term between $Z_i^{N,1}$ and $Z_j^{N,f}$. Now for each $i \in Q$ define the semimartingale Y_i^N by

$$Y_i^N(t) = \langle f, \mu_i^N(t) \rangle \left(\sum_{j \in Q} h_j^N(t) \right) - h_i^N(t) \left(\sum_{j \in Q} \langle f, \mu_j^N(t) \rangle \right). \tag{4.24}$$

The semimartingale Y_i^N is just a linear combination of the semimartingales of the form $h_j^N(t)\langle f, \mu_i^N(t)\rangle$. Using (4.23) we can write

$$Y_i^N(t) = Y_i^N(0) + N \int_0^t L_i^N(s)ds + Z_i^{N,2}(t), \tag{4.25}$$

where $Z_i^{N,2}$ is a well-behaved semimartingale and for any $t \geq 0$

$$\begin{aligned}
L_i^N(t) &= \sum_{j,k \in Q} \Big[A_{ik}(h^N(t))h_j^N(t)\langle f, \mu_k^N(t)\rangle + A_{jk}(h^N(t))h_k^N(t)\langle f, \mu_i^N(t)\rangle \\
&\qquad - A_{jk}(h^N(t))h_i^N(t)\langle f, \mu_k^N(t)\rangle - A_{ik}(h^N(t))h_k^N(t)\langle f, \mu_j^N(t)\rangle \Big] \\
&= \sum_{k \in Q} A_{ik}(h^N(t))Y_k^N(t) + \langle f, \mu_i^N(t)\rangle \left[\sum_{j,k \in Q} A_{jk}(h^N(t))h_k^N(t) \right] \\
&\quad - h_i^N(t) \left[\sum_{j,k \in Q} A_{jk}(h^N(t))\langle f, \mu_k^N(t)\rangle \right].
\end{aligned}$$

Define a matrix $G(h)$ for each $h \in \mathbb{R}_+^q$ as follows

$$G(h) = \begin{cases} A(\bar{0}_q) & \text{if } h = \bar{0}_q \\ A(h) + \left(\frac{\langle \bar{1}_q, \theta(h)\rangle}{\langle \bar{1}_q, h\rangle} I_q - \frac{h}{\langle \bar{1}_q, h\rangle} \bar{1}_q^T A(h) \right) & \text{otherwise.} \end{cases}$$

Note that if $h^N(t) \neq \bar{0}_q$ then

$$\langle f, \mu_i^N(t)\rangle = \frac{Y_i^N(t) + h_i^N(t)\left(\sum_{j \in Q}\langle f, \mu_j^N(t)\rangle \right)}{\langle \bar{1}_q, h^N(t)\rangle} \text{ and}$$

$$\sum_{j,k \in Q} A_{jk}(h^N(t))\langle f, \mu_k^N(t)\rangle$$

$$= \frac{\sum_{j,k \in Q} A_{jk}(h^N(t))Y_k^N(t) + \left(\sum_{j \in Q}\langle f, \mu_j^N(t)\rangle \right)\left(\sum_{j,k \in Q} A_{jk}(h^N(t))h_k^N(t) \right)}{\langle \bar{1}_q, h^N(t)\rangle}.$$

This allows us to write

$$L_i^N(t) = \sum_{j \in Q} G_{ij}(h^N(t))Y_j^N(t)$$

and hence from (4.25)

$$Y_i^N(t) = Y_i^N(0) + N \int_0^t \left[\sum_{j \in Q} G_{ij}(h^N(s))Y_j^N(s) \right] ds + Z_i^{N,2}(t). \tag{4.26}$$

Let Y^N and $Z^{N,2}$ be the $\mathbb{R}^{q-1}$-valued semimartingales given by

$$Y^N(t) = (Y_1^N(t), \ldots, Y_{q-1}^N(t)) \text{ and } Z^{N,2}(t) = (Z_1^{N,2}(t), \ldots, Z_{q-1}^{N,2}(t)) \text{ for } t \geq 0.$$

For each $h \in \mathbb{R}_+^q$ let $\bar{G}(h) \in \mathbb{M}(q-1, q-1)$ be the matrix defined by

$$\bar{G}_{ij}(h) = G_{ij}(h) - G_{iq}(h) \text{ for all } i, j \in \{1, \ldots, q-1\}.$$

Observe that $\sum_{i \in Q} Y_i^N(t) = 0$ and hence $Y_q^N(t) = -\sum_{i=1}^{q-1} Y_i^N(t)$. From (4.26) we get

$$Y^N(t) = Y^N(0) + N \int_0^t \bar{G}(h^N(s))Y^N(s)ds + Z^{N,2}(t). \tag{4.27}$$

From part (E) of Lemma A.1 the matrix $\bar{G}(h_{\text{eq}})$ is stable, that is, all its eigenvalues have strictly negative real parts.

We now define a $\mathbb{R}^q_+ \times \mathbb{R}^{q-1}$-valued semimartingale X^N by

$$X^N(t) = (h^N(t), Y^N(t)) = (h^N_1(t), \ldots, h^N_q(t), Y^N_1(t), \ldots, Y^N_{q-1}(t)) \text{ for } t \geq 0. \qquad (4.28)$$

From (4.22) and (4.27) we can see that X^N satisfies

$$X^N(t) = X^N(0) + N \int_0^t F(X^N(s))ds + Z^N(t) \qquad (4.29)$$

where $Z^N(t) = (Z^{N,1}(t), Z^{N,2}(t))$ is a well-behaved semimartingale and $F : \mathbb{R}^q_+ \times \mathbb{R}^{q-1} \to \mathbb{R}^{2q-1}$ is the function given by

$$F(x) = (\theta(h), \bar{G}(h)y) \text{ for } x = (h, y) \in \mathbb{R}^q_+ \times \mathbb{R}^{q-1}. \qquad (4.30)$$

Let $x_{\text{eq}} = (h_{\text{eq}}, \bar{0}_{q-1})$. Then $F(x_{\text{eq}}) = \bar{0}_{2q-1}$ and the Jacobian matrix $[JF(x_{\text{eq}})] \in \mathbb{M}(2q-1, 2q-1)$ has the block lower-triangular form

$$[JF(x_{\text{eq}})] = \left[\begin{array}{cc} [J\theta(h_{\text{eq}})] & O_{q,q-1} \\ C & \bar{G}(h_{\text{eq}}) \end{array} \right],$$

where C is some $(q-1) \times q$ matrix in $\mathbb{M}(q-1, q)$ and $O_{q,q-1}$ is the $q \times (q-1)$ matrix of zeroes. We mentioned above that the matrix $\bar{G}(h_{\text{eq}})$ is stable and part (B) of Assumption 2.1 says that the matrix $[J\theta(h_{\text{eq}})]$ is also stable. Due to the block triangular form, the matrix $[JF(x_{\text{eq}})]$ is stable as well.

Pick any $h_0 \in U_{\text{eq}}$ and $y_0 \in \mathbb{R}^{q-1}$. Let $h(t) = \psi_\theta(h_0, t)$ for all $t \geq 0$, where ψ_θ is the *flow* defined in Section 2.2. The set $U_{\text{eq}} \subset \mathbb{R}^q_+$ in ψ_θ-invariant and hence $h(t) \in U_{\text{eq}}$ for all $t \geq 0$. Let $y(t)$ be the unique solution of the initial value problem

$$\frac{dy}{dt} = \bar{G}(h(t))y, \quad y(0) = y_0. \qquad (4.31)$$

Since the above differential equation is linear in the y variable, the solution $y(t)$ is defined for all $t \geq 0$. Moreover $h(t) \to h_{\text{eq}}$ and $\bar{G}(h(t)) \to \bar{G}(h_{\text{eq}})$ as $t \to \infty$. The matrix $\bar{G}(h_{\text{eq}})$ is stable and therefore $y(t) \to \bar{0}_{q-1}$ as $t \to \infty$. Let $U(x_{\text{eq}}) = U_{\text{eq}} \times \mathbb{R}^{q-1}$. For each $x_0 = (h_0, y_0) \in U(x_{\text{eq}})$ and $t \geq 0$, let $\psi_F(x_0, t) = (h(t), y(t))$ with $h(t) = \psi_\theta(h_0, t)$ and $y(t)$ being the solution of (4.31).

The mapping $\psi_F : U(x_{\text{eq}}) \times \mathbb{R}_+ \to U(x_{\text{eq}})$ is the flow of the vector field F on $U(x_{\text{eq}})$. For all $x \in U(x_{\text{eq}})$ and $t \geq 0$, ψ_F satisfies

$$\psi_F(x, t) = x + \int_0^t F(\psi_F(x, s))ds. \qquad (4.32)$$

From the discussion in the preceding paragraph we can conclude that

$$\lim_{t \to \infty} \psi_F(x, t) = x_{\text{eq}} \text{ for all } x \in U(x_{\text{eq}}). \qquad (4.33)$$

We have assumed in this proposition that there is a compact set $K_0 \subset U_{\text{eq}}$ such that $h^N(0) \in K_0$ a.s. for each $N \in \mathbb{N}$. Note that any function $f \in \mathcal{D}_0$ is bounded. The definition of the semimartingale Y^N guarantees that there is a compact set $K_1 \subset \mathbb{R}^{q-1}$ such that $X^N(0) = (h^N(0), Y^N(0)) \in K_0 \times K_1 \subset U(x_{\text{eq}})$ a.s. for each $N \in \mathbb{N}$.

Now consider the equation (4.29). For large values of N, the semimartingale X^N is driven by a large drift term of the form $NF(X^N(\cdot))$. The vector x_{eq} is a stable fixed point for this drift term and $U(x_{\text{eq}})$ is its region of attraction. If we start in this region

of attraction, then this drift is very forceful. It completely overwhelms the effect of the well-behaved semimartingale Z^N and drives X^N to the stable fixed point x_{eq}. Moreover as N gets large, the trajectories of X^N start looking more and more like the trajectories of the deterministic flow ψ_F with time compressed by a factor of N. These ideas are made precise in a much more general setting by Katzenberger [23]. We use Theorem 6.3 in [23] to deduce that for any $T > 0$

$$\sup_{t \in [0,T]} \left\| X^N(t) - \psi_F(X^N(0), Nt) \right\| \Rightarrow 0 \text{ as } N \to \infty, \tag{4.34}$$

where $\|\cdot\|$ is the standard Euclidean norm. From the definition of X^N and ψ_F it is also clear that for any $T > 0$

$$\sup_{t \in [0,T]} \left\| h^N(t) - \psi_\theta(h^N(0), Nt) \right\| \Rightarrow 0 \text{ as } N \to \infty. \tag{4.35}$$

From now on, for any $x \in \mathbb{R}^n$ and $\epsilon > 0$, let $B_\epsilon^n(x)$ denote the open ball in $\mathbb{R}^n$ centered at x with radius ϵ. We have already argued that the Jacobian matrix of F at x_{eq} is stable. From a simple linearization argument (see for example the proof of Theorem 3.7 in Khalil [24]) we can see that there exists a $\delta_0 > 0$ such that the open ball $B_{\delta_0}^{2q-1}(x_{\text{eq}}) \subset U(x_{\text{eq}})$ and for every $\delta \in (0, \delta_0)$ there exists a ψ_F-invariant open set W_δ whose closure $\bar{W}_\delta$ is contained in $B_\delta^{2q-1}(x_{\text{eq}})$. From (4.33) we obtain

$$K_0 \times K_1 \subset \bigcup_{t \geq 0} \left\{ x \in U(x_{\text{eq}}) : \psi_F(x,t) \subset W_\delta \right\}.$$

As W_δ is ψ_F-invariant, the open sets on the right are getting bigger and bigger as t increases. Compactness of $K_0 \times K_1$ implies that there exists a $t_\delta > 0$ such that $\psi_F(x,t) \in W_\delta$ for all $x \in K_0 \times K_1$ and $t \geq t_\delta$. This immediately gives us

$$\sup_{t \geq t_\delta} \sup_{x \in K_0 \times K_1} \left\| \psi_F(x,t) - x_{\text{eq}} \right\| \leq \delta$$

and letting $\delta \to 0$ we obtain

$$\limsup_{t \to \infty} \sup_{x \in K_0 \times K_1} \left\| \psi_F(x,t) - x_{\text{eq}} \right\| = 0. \tag{4.36}$$

Now let t_N be any sequence satisfying the conditions of this proposition. Then for any $T > 0$

$$\sup_{t \in [0,T]} \left\| X^N(t + t_N) - x_{\text{eq}} \right\|$$

$$\leq \sup_{t \in [0,T]} \left\| X^N(t + t_N) - \psi_F(X^N(0), N(t + t_N)) \right\| + \sup_{t \in [0,T]} \left\| \psi_F(X^N(0), Nt + Nt_N) - x_{\text{eq}} \right\|$$

$$\leq \sup_{t \in [0,T]} \left\| X^N(t + t_N) - \psi_F(X^N(0), N(t + t_N)) \right\| + \sup_{t \geq Nt_N} \sup_{x \in K_0 \times K_1} \left\| \psi_F(x,t) - x_{\text{eq}} \right\|,$$

where the second inequality is true because $X^N(0) \in K_0 \times K_1$ a.s. for each $N \in \mathbb{N}$. From (4.34) and (4.36) we can see that as $N \to \infty$

$$\sup_{t \in [0,T]} \left\| X^N(t + t_N) - x_{\text{eq}} \right\| \Rightarrow 0$$

which of course implies that

$$\sup_{t \in [0,T]} \left\| h^N(t + t_N) - h_{\text{eq}} \right\| \Rightarrow 0 \text{ and } \sup_{t \in [0,T]} \left\| Y^N(t + t_N) \right\| \Rightarrow 0. \tag{4.37}$$

This proves part (A) of the proposition since the norms $\|\cdot\|$ and $\|\cdot\|_1$ are equivalent in $\mathbb{R}^q$. From the definition of Y_i^N we can check that for each $i, j \in Q$ and $t \geq 0$

$$h_j^N(t)Y_i^N(t) - h_i^N(t)Y_j^N(t) = \left(\sum_{l \in Q} h_l^N(t)\right) \left(h_j^N(t)\langle f, \mu_i^N(t)\rangle - h_i^N(t)\langle f, \mu_j^N(t)\rangle\right).$$

The limits (4.37) immediately give us part (B) of the proposition for any $f \in \mathcal{D}_0$. As $\mathcal{D}_0$ is dense in $C(E)$, part (B) holds for any $f \in C(E)$. $\qquad\square$

The following lemma will be useful in proving Theorem 2.6.

Lemma 4.5. *Let the notation and assumptions be the same as in Proposition 2.5. Then there is a compact set $K \subset U_{\mathrm{eq}}$ such that for all $T > 0$*

$$\lim_{N \to \infty} \mathbb{P}\left(h^N(t) \notin K \text{ for any } t \in [0, T]\right) = 0.$$

Proof. Since the Jacobian matrix $[J\theta(h_{\mathrm{eq}})]$ is stable (part (B) of Assumption 2.1), by a linearization argument similar to the one referred in the proof above, we can find an $\epsilon > 0$ such that the open ball $B_\epsilon^q(h_{\mathrm{eq}}) \subset U_{\mathrm{eq}}$ and there exists a ψ_θ-invariant open set U_ϵ such that its closure $\bar{U}_\epsilon \subset B_\epsilon^q(h_{\mathrm{eq}})$. One can argue as before that since K_0 is a compact set, there exists a t_ϵ such that for all $t \geq t_\epsilon$ and $x \in K_0$, $\psi_\theta(x, t) \in U_\epsilon$. If we define $\hat{K}_0$ as

$$\hat{K}_0 = \left\{h \in U_{\mathrm{eq}} : \psi_\theta(h, t_\epsilon) \subset \bar{U}_\epsilon\right\}, \tag{4.38}$$

then it is a ψ_θ-invariant compact set containing K_0. Because we have assumed that $h^N(0) \in K_0$ a.s. for all $N \in \mathbb{N}$, we must have that for all $t \geq 0$, $\psi_\theta(h^N(0), t) \in \hat{K}_0$ a.s.. But U_{eq} is open in $\mathbb{R}_+^q$ (see Section 2.2) and so there is a $\gamma > 0$ such that

$$K = \left\{x \in U_{\mathrm{eq}} : \inf_{y \in \hat{K}_0} \|y - x\| \leq \gamma\right\}$$

is a compact subset of U_{eq}. Observe that if for some $t \geq 0$, $h^N(t) \notin K$ then we must have that $\|h^N(t) - \psi_\theta(h^N(0), Nt)\| > \gamma$. Therefore

$$\lim_{N \to \infty} \mathbb{P}\left(h^N(t) \notin K \text{ for any } t \in [0, T]\right) \leq \lim_{N \to \infty} \mathbb{P}\left(\sup_{t \in [0, T]} \|h^N(t) - \psi_\theta(h^N(0), Nt)\| > \gamma\right).$$

The limit on the right is 0 due to (4.35) and this proves the lemma. $\qquad\square$

4.3 Solution to a system of partial differential equations

Recall the discussion at the end of Section 2.4. To prove Theorem 2.6 we require a function Λ that allows us to construct a $\mathcal{P}(E)$ valued process $\{\nu^N(t) : t \geq 0\}$ (see (2.44)) whose dynamics is *well-behaved* as N approaches ∞. The goal of this section is to guarantee that such a function Λ exists.

Specifically, we need to show that for some open set $\hat{U}_{\mathrm{eq}} \subset \mathbb{R}^q$ containing U_{eq} (given by (2.20)), we have a function $\Lambda \in C^2(\hat{U}_{\mathrm{eq}}, \mathbb{R}_*^q)$ which satisfies the following:

$$A^T(h)\Lambda(h) + [J\Lambda(h)]\theta(h) = \bar{0}_q \text{ for all } h \in \hat{U}_{\mathrm{eq}}, \tag{4.39}$$

$$\langle \Lambda(h), h \rangle = 1 \text{ for all } h \in \hat{U}_{\mathrm{eq}} \tag{4.40}$$

$$\text{and} \quad \Lambda(h_{\mathrm{eq}}) = v_{\mathrm{eq}}. \tag{4.41}$$

Here v_{eq} is defined in (2.34) and $[J\Lambda(h)]$ in equation (4.39) refers to the Jacobian matrix of Λ at h. The significance of the above relations will become clear in Section 4.4.

The major difficulty in solving (4.39) arises in the neighbourhood of h_{eq}. This is because $\theta(h_{\text{eq}}) = \bar{0}_q$ (part (A) of Assumption 2.1), which causes degeneracy in the system. However the next proposition shows that by employing power series expansions we can get around this problem and find an analytic solution to (4.39) in a neighbourhood of h_{eq}. We later construct an open set $\hat{U}_{\text{eq}}$ containing U_{eq} and extend the solution over the whole $\hat{U}_{\text{eq}}$. We also show that this solution has all the properties we desire.

Proposition 4.6. *There exists an open set V containing h_{eq} such that the equation (4.39) has an analytic solution Λ on V satisfying (4.41).*

Proof. We first transform the equation (4.39) into another equation that is easier to work with. Let $\lambda_1, \ldots, \lambda_q$ be the eigenvalues of the matrix $[J\theta(h_{\text{eq}})]$. We will prove this proposition under the assumption that all these eigenvalues are real. We later remark how the proof changes when they take complex values.

We know from part (B) of Assumption 2.1 that $\lambda_i < 0$ for each $i \in Q$. Pick an $\epsilon_0 \in (0, 1)$ such that

$$\lambda_i < -4\epsilon_0 \text{ for all } i \in Q. \tag{4.42}$$

Let $M_1 \in \mathbb{M}(q, q)$ be the matrix representing the Jordan canonical form of $[J\theta(h_{\text{eq}})]$. Its diagonal is occupied by $\lambda_1, \ldots, \lambda_q$, while its super-diagonal entries are either 0 or 1. All the other entries are 0. Let $P_1 \in \mathbb{M}(q, q)$ be the invertible matrix such that $P_1[J\theta(h_{\text{eq}})]P_1^{-1} = M_1$. Let $P_2 = \text{Diag}(1, \epsilon_0, \epsilon_0^2, \ldots, \epsilon_0^{q-1})$, $P = P_2^{-1}P_1$ and $M = P_2^{-1}M_1P_2$. Then

$$P[J\theta(h_{\text{eq}})]P^{-1} = M, \tag{4.43}$$

and M is just the matrix M_1 with each 1 on the super-diagonal replaced by ϵ_0.

In this proof, $\bar{0}$ will always denote the vector of zeroes in $\mathbb{R}^q$. Since $h_{\text{eq}} \in \mathbb{R}_*^q$ (see part (A) of Lemma A.1) and the map $x \mapsto h_{\text{eq}} + P^{-1}x$ is continuous, we can find a $r_0 > 0$ such that for any $x \in B_{r_0}^q(\bar{0})$ we have $h_{\text{eq}} + P^{-1}x \in \mathbb{R}_*^q$, where $B_{r_0}^q(\bar{0})$ is the open ball in $\mathbb{R}^q$ with radius r_0 centered at $\bar{0}$. For all $x \in B_{r_0}^q(\bar{0})$, let $\hat{A}(x) \in \mathbb{M}(q, q)$ and $\hat{\theta}(x) \in \mathbb{R}^q$ be given by

$$\hat{A}(x) = A^T(h_{\text{eq}} + P^{-1}x) \text{ and } \hat{\theta}(x) = P\theta(h_{\text{eq}} + P^{-1}x).$$

Suppose $\beta : U \to \mathbb{R}^q$ is a function which is analytic in an open set $U \subset B_{r_0}^q(\bar{0})$ and for all $x \in U$

$$\hat{A}(x)\beta(x) + [J\beta(x)]\hat{\theta}(x) = \bar{0} \tag{4.44}$$

along with

$$\beta(\bar{0}) = v_{\text{eq}}, \tag{4.45}$$

where v_{eq} is given by (2.34). If $V \subset \mathbb{R}_+^q$ is the image of U under the map $x \mapsto h_{\text{eq}} + P^{-1}x$, then V is an open set containing h_{eq} and the function $\Lambda : V \to \mathbb{R}^q$ defined by $\Lambda(h) = \beta(P(h - h_{\text{eq}}))$ is an analytic solution to (4.39) satisfying (4.41). Hence to prove the proposition it suffices to show that equation (4.44) has a solution β in some neighbourhood of $\bar{0}$ which satisfies (4.45).

We will be using the multi-index notation to write the power series in q variables. For any multi-index $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_q) \in \mathbb{N}_0^q$ let $|\alpha| = \alpha_1 + \alpha_2 + \cdots + \alpha_q$ and $\alpha! = \alpha_1!\alpha_2!\ldots\alpha_q!$. For two multi-indices $\nu = (\nu_1, \ldots, \nu_q) \in \mathbb{N}_0^q$ and $\alpha = (\alpha_1, \ldots, \alpha_q) \in \mathbb{N}_0^q$, we say that $\nu \leq \alpha$ if $\nu_i \leq \alpha_i$ for all $i = 1, \ldots, q$ and we say that $\nu < \alpha$ if $\nu \leq \alpha$ and $\nu \neq \alpha$. If $\nu \leq \alpha$ then

$$\binom{\alpha}{\nu} = \frac{\alpha!}{\nu!(\alpha - \nu)!}.$$

For any vector $x \in \mathbb{R}^q$ and multi-index $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_q) \in \mathbb{N}_0^q$ define

$$x^\alpha = x_1^{\alpha_1} x_2^{\alpha_2} \ldots x_q^{\alpha_q}$$

and the differential operator D_α as

$$D_\alpha = \frac{\partial^{\alpha_1}}{\partial x_1^{\alpha_1}} \cdots \frac{\partial^{\alpha_q}}{\partial x_q^{\alpha_q}}.$$

The operator D_α acts component-wise on matrix and vector valued functions.

Consider the function β given by the power series

$$\beta(x) = v_{\text{eq}} + \sum_{|\alpha|=1}^\infty \gamma_\alpha x^\alpha, \tag{4.46}$$

where $\gamma_\alpha \in \mathbb{R}^q$ is given by

$$\gamma_\alpha = \frac{D_\alpha \beta(0)}{\alpha!}. \tag{4.47}$$

This function β satisfies (4.45). To prove the proposition it suffices to show that the vectors γ_α can be suitably chosen such that β satisfies (4.44) and there exists a positive constant C such that

$$\|\gamma_\alpha\|_\infty \leq C^{|\alpha|} \text{ for all } \alpha \in \mathbb{N}_0^q. \tag{4.48}$$

The last condition ensures the absolute convergence of the power series (4.46) in a neighbourhood of $\bar{0}$.

Since $\hat{\theta}(\bar{0}) = \theta(h_{\text{eq}}) = \bar{0}$, if we plug $x = \bar{0}$ in (4.44) we obtain

$$\hat{A}(\bar{0})\beta(\bar{0}) = A^T(h_{\text{eq}})v_{\text{eq}} = 0.$$

This is satisfied because of the choice of v_{eq} (see (2.34)).

Applying the operator D_α to equation (4.44) and using the product rule for multi-derivatives we get

$$\bar{0} = D_\alpha \left(\hat{A}(x)\beta(x) \right) + D_\alpha \left([J\beta(x)]\hat{\theta}(x) \right)$$

$$= \sum_{\nu \leq \alpha} \binom{\alpha}{\nu} \left(D_{(\alpha-\nu)}\hat{A}(x) \right) (D_\nu \beta(x)) + \sum_{\nu \leq \alpha} \binom{\alpha}{\nu} (D_\nu[J\beta(x)]) \left(D_{(\alpha-\nu)}\hat{\theta}(x) \right)$$

$$= \hat{A}(x)D_\alpha\beta(x) + (D_\alpha[J\beta(x)])\hat{\theta}(x) + \sum_{\{\nu < \alpha, |\alpha-\nu|=1\}} \binom{\alpha}{\nu} (D_\nu[J\beta(x)]) \left(D_{(\alpha-\nu)}\hat{\theta}(x) \right)$$

$$+ \sum_{\nu < \alpha} \binom{\alpha}{\nu} \left(D_{(\alpha-\nu)}\hat{A}(x) \right) (D_\nu \beta(x)) + \sum_{\{\nu < \alpha, |\alpha-\nu|>1\}} \binom{\alpha}{\nu} (D_\nu[J\beta(x)]) \left(D_{(\alpha-\nu)}\hat{\theta}(x) \right).$$

On rearranging we obtain

$$\hat{A}(x)D_\alpha\beta(x) + (D_\alpha[J\beta(x)])\hat{\theta}(x) + \sum_{\{\nu < \alpha, |\alpha-\nu|=1\}} \binom{\alpha}{\nu} (D_\nu[J\beta(x)]) \left(D_{(\alpha-\nu)}\hat{\theta}(x) \right)$$

$$= -\sum_{\nu < \alpha} \binom{\alpha}{\nu} \left(D_{(\alpha-\nu)}\hat{A}(x) \right) (D_\nu \beta(x)) - \sum_{\{\nu < \alpha, |\alpha-\nu|>1\}} \binom{\alpha}{\nu} (D_\nu[J\beta(x)]) \left(D_{(\alpha-\nu)}\hat{\theta}(x) \right).$$

$$\tag{4.49}$$

For any $j \in Q$, let $e_j \in \mathbb{N}_0^q$ be the multi-index $(0, \ldots, 0, 1, 0, \ldots, 0)$, with the 1 at the j-th position. Observe that if $|\alpha - \nu| = 1$ then $\nu = \alpha - e_j$ for some $j \in Q$. Therefore

$$
\sum_{\{\nu < \alpha, |\alpha - \nu| = 1\}} \binom{\alpha}{\nu} (D_\nu[J\beta(x)]) \left(D_{(\alpha - \nu)} \hat{\theta}(x) \right) = \sum_{j \in Q} \alpha_j \left(D_{(\alpha - e_j)}[J\beta(x)] \right) \left(\partial_j \hat{\theta}(x) \right)
$$

$$
= \sum_{j,k \in Q} \alpha_j \left(\partial_j \hat{\theta}_k(x) \right) \left(D_{(\alpha - e_j + e_k)} \beta(x) \right)
$$

$$
= \sum_{j,k \in Q} \alpha_j [J\hat{\theta}(x)]_{kj} \left(D_{(\alpha - e_j + e_k)} \beta(x) \right).
$$

Note that $[J\hat{\theta}(\bar{0})] = M$ (see (4.43)). This matrix has the eigenvalues $\lambda_1, \ldots, \lambda_q$ on the diagonal and either 0 or ϵ_0 on the super-diagonal. For each $j = 2, \ldots, q$ let $\epsilon_j = \epsilon_0$ if $M_{(j-1)j} = \epsilon_0$ and $\epsilon_j = 0$ otherwise. Then for $x = \bar{0}$ we obtain

$$
\sum_{\{\nu < \alpha, |\alpha - \nu| = 1\}} \binom{\alpha}{\nu} (D_\nu[J\beta(\bar{0})]) \left(D_{(\alpha - \nu)} \hat{\theta}(\bar{0}) \right)
$$

$$
= \sum_{j \in Q} \alpha_j \lambda_j D_\alpha \beta(\bar{0}) + \sum_{j=2}^{q} \alpha_j \epsilon_j D_{(\alpha - e_j + e_{j-1})} \beta(\bar{0}).
$$

Note that $\hat{\theta}(\bar{0}) = P\theta(h_{\text{eq}}) = \bar{0}$ and for each $\alpha \in \mathbb{N}_0^q$, γ_α is given by (4.47). We plug $x = \bar{0}$ in (4.49) and divide by $\alpha!$ to get

$$
\hat{A}(\bar{0})\gamma_\alpha + \sum_{j \in Q} \alpha_j \lambda_j \gamma_\alpha + \sum_{j=2, \alpha_j > 0}^{q} (\alpha_{j-1} + 1)\epsilon_j \gamma_{(\alpha - e_j + e_{j-1})} = Y_\alpha \tag{4.50}
$$

where $Y_\alpha = -\sum_{\nu < \alpha} \frac{\left(D_{(\alpha - \nu)} \hat{A}(\bar{0}) \right)}{(\alpha - \nu)!} \gamma_\nu - \sum_{\{\nu < \alpha, |\alpha - \nu| > 1\}} \frac{(D_\nu[J\beta(\bar{0})])}{\nu!} \frac{\left(D_{(\alpha - \nu)} \hat{\theta}(\bar{0}) \right)}{(\alpha - \nu)!}.$

The second term can be simplified as

$$
\sum_{\{\nu < \alpha, |\alpha - \nu| > 1\}} \frac{(D_\nu[J\beta(\bar{0})])}{\nu!} \frac{\left(D_{(\alpha - \nu)} \hat{\theta}(\bar{0}) \right)}{(\alpha - \nu)!} = \sum_{\{\nu < \alpha, |\alpha - \nu| > 1\}} \sum_{j \in Q} \frac{\left(D_{(\alpha - \nu)} \hat{\theta}_j(\bar{0}) \right)}{(\alpha - \nu)!} \frac{\left(D_{(\nu + e_j)} \beta(\bar{0}) \right)}{\nu!}
$$

$$
= \sum_{\{\nu < \alpha, |\alpha - \nu| > 1\}} \sum_{j \in Q} (\nu_j + 1) \frac{\left(D_{(\alpha - \nu)} \hat{\theta}_j(\bar{0}) \right)}{(\alpha - \nu)!} \gamma_{(\nu + e_j)}.
$$

Therefore we can write Y_α as

$$
Y_\alpha = -\sum_{\nu < \alpha} \frac{\left(D_{(\alpha - \nu)} \hat{A}(\bar{0}) \right)}{(\alpha - \nu)!} \gamma_\nu - \sum_{\{\nu < \alpha, |\alpha - \nu| > 1\}} \sum_{j \in Q} (\nu_j + 1) \frac{\left(D_{(\alpha - \nu)} \hat{\theta}_j(\bar{0}) \right)}{(\alpha - \nu)!} \gamma_{(\nu + e_j)}. \tag{4.51}
$$

For each $k \in \mathbb{N}$, let S_k be the set of multi-indices given by $S_k = \{\alpha \in \mathbb{N}_0^q : |\alpha| = k\}$. The number of elements in S_k is

$$
s_k = \binom{k + q - 1}{q - 1}.
$$

We order the multi-indices in S_k as follows. We say that $\nu \preceq \alpha$ if and only if $\sum_{i \in Q} i\nu_i \leq \sum_{i \in Q} i\alpha_i$. Let $\alpha^k(1), \ldots, \alpha^k(s_k)$ be all the elements of S_k listed in the order given by $\preceq$.

Let the matrix $\Xi^{(k)} \in \mathbb{M}(qs_k, qs_k)$ be a block matrix composed of s_k^2 blocks of size $q \times q$. For each $i, j \in \{1, 2, \ldots, s_k\}$ the block starting at row $q(i-1) + 1$ and column $q(j-1) + 1$ of matrix $\Xi^{(k)}$ is occupied by the matrix $L_{ij} \in \mathbb{M}(q, q)$ defined as follows. If $i = j$ then $L_{ii} = \hat{A}(\bar{0}) + \left(\sum_{l \in Q} \alpha_l^k(i) \lambda_l \right) I_q$. If i and j are such that $\alpha^k(j) = \alpha^k(i) - e_l + e_{l-1}$ for some $l \in \{2, \ldots, q\}$ then $L_{ij} = \epsilon_l(\alpha_{l-1}^k(i) + 1) I_q$. For every other i and j, L_{ij} is just a matrix of zeroes. The matrix $\Xi^{(k)}$ is lower block-triangular and its determinant is given by

$$\mathrm{Det}\left(\Xi^{(k)} \right) = \prod_{i=1}^{s_k} \mathrm{Det}\left(\hat{A}(\bar{0}) + \left(\sum_{l \in Q} \alpha_l^k(i) \lambda_l \right) I_q \right).$$

The eigenvalues $\lambda_1, \ldots, \lambda_q$ satisfy (4.42). Since all the eigenvalues of the matrix $\hat{A}(\bar{0}) = A^T(h_{\mathrm{eq}})$ have non-positive real parts (see part (B) of Lemma A.1), the above determinant is non-zero. Hence the matrix $\Xi^{(k)}$ is invertible.

Let $X^{(k)}$ and $Y^{(k)}$ be the vectors in $\mathbb{R}^{qs_k}$ given by

$$X^{(k)} = \left(\gamma_{\alpha^k(1)}, \gamma_{\alpha^k(2)}, \ldots, \gamma_{\alpha^k(s_k)} \right) \text{ and } Y^{(k)} = \left(Y_{\alpha^k(1)}, Y_{\alpha^k(2)}, \ldots, Y_{\alpha^k(s_k)} \right).$$

Using (4.50) we obtain the following linear system

$$\Xi^{(k)} X^{(k)} = Y^{(k)}$$

and since the matrix $\Xi^{(k)}$ is invertible

$$X^{(k)} = [\Xi^{(k)}]^{-1} Y^{(k)}. \tag{4.52}$$

Note that $Y^{(k)}$ only depends on $\{\gamma_\alpha : \alpha \in S_l \text{ for } l \in \{0, 1, \ldots, k-1\}\}$. Hence for each $k \in \mathbb{N}$ we can solve for the whole set $\{\gamma_\alpha : \alpha \in S_k\}$ using (4.52). Doing this iteratively for each k we can solve for γ_α for all $\alpha \in \mathbb{N}_0^q$. The function β given by (4.46) with this choice of γ_α's will solve (4.44) in a neighbourhood of $\bar{0}$ if we can show that (4.48) holds for some $C > 0$. Showing this will be our next task.

Any entry on the diagonal of $\Xi^{(k)}$ has the form $\hat{A}_{ii}(\bar{0}) + \sum_{j \in Q} \lambda_j \alpha_j$ for some $\alpha \in S_k$ and $i \in Q$. Observe that $\hat{A}(\bar{0}) = A^T(h_{\mathrm{eq}})$ and this matrix only has non-positive entries on its diagonal (see (2.13)). From (4.42), for $\alpha \in S_k$ we obtain the estimate

$$\left| \hat{A}_{ii}(\bar{0}) + \sum_{j \in Q} \lambda_j \alpha_j \right| \geq 4\epsilon_0 k. \tag{4.53}$$

For each row of $\Xi^{(k)}$, the sum of the absolute values of the non-diagonal entries is bounded above by

$$\max_{i \in Q} \left| \sum_{j \in Q, j \neq i} \hat{A}_{ij}(\bar{0}) + \sum_{l=2}^q \epsilon_l(\alpha_{l-1} + 1) \right| \leq \max_{i \in Q} \left(\sum_{j \in Q, j \neq i} |\hat{A}_{ij}(\bar{0})| \right) + \epsilon_0(k + q). \tag{4.54}$$

Hence from (4.53) and (4.54) we can conclude that there exists a $K_0 \in \mathbb{N}$ such that for all $k \geq K_0$ the matrix $\Xi^{(k)}$ is strictly diagonally dominant and we have

$$\min_{1 \leq l \leq qs_k} \left| \left| [\Xi^{(k)}]_{ll} \right| - \sum_{r=1, r \neq l}^{qs_k} \left| [\Xi^{(k)}]_{lr} \right| \right| \geq k\epsilon_0.$$

Theorem 1 in Varah [36] shows that for all $k \geq K_0$

$$\left\| [\Xi^{(k)}]^{-1} \right\|_\infty \leq \frac{1}{k\epsilon_0}. \tag{4.55}$$

Part (D) of Assumption 2.1 says that for each $i, j \in Q$, the functions ρ_i and β_{ij} are analytic in a neighbourhood of h_{eq}. This implies that there is a neighbourhood U of $\bar{0}$ such that the $\mathbb{M}(q,q)$-valued function $\hat{A}$ and the $\mathbb{R}^q$-valued function $\hat{\theta}$ are analytic component-wise on U. Therefore there is a constant C_0 such that

$$\|D_\alpha \hat{A}(\bar{0})\|_\infty \leq C_0^{|\alpha|}\alpha! \text{ and } \|D_\alpha \hat{\theta}(\bar{0})\|_\infty \leq C_0^{|\alpha|}\alpha! \text{ for all } \alpha \in \mathbb{N}_0^q. \tag{4.56}$$

We can assume that $C_0 > 1$. Choose a $\delta > 0$ satisfying

$$\delta < \left(\frac{\epsilon_0}{C_0 q(q+1)2^{q+3}} \right) \tag{4.57}$$

and define $C = C_0/\delta$. We will prove (4.48) by induction. Let $k > K_0$ and suppose that C is large enough to satisfy

$$\|\gamma_\nu\|_\infty \leq C^{|\nu|} \tag{4.58}$$

for all $l \in \{1, 2, \ldots, k-1\}$ and $\nu \in S_l$. To prove (4.48) we need to show that $\|\gamma_\alpha\|_\infty \leq C^k$ for all $\alpha \in S_k$. This is equivalent to showing that $\left\| X^{(k)} \right\|_\infty \leq C^k$. From (4.52) and (4.55) we have

$$\left\| X^{(k)} \right\|_\infty \leq \left\| [\Xi^{(k)}]^{-1} \right\|_\infty \left\| Y^{(k)} \right\|_\infty \leq \frac{1}{k\epsilon_0} \left\| Y^{(k)} \right\|_\infty.$$

Hence to prove (4.48) it suffices to show that

$$\left\| Y^{(k)} \right\|_\infty = \max_{\alpha \in S_k} \|Y_\alpha\|_\infty \leq k\epsilon_0 C^k. \tag{4.59}$$

From (4.51), (4.56) and (4.58), for any $\alpha \in S_k$ we get

$$\|Y_\alpha\|_\infty \leq \sum_{\nu < \alpha} C_0^{|\alpha-\nu|}C^{|\nu|} + \sum_{\{\nu < \alpha, |\alpha-\nu| > 1\}} \sum_{j \in Q} (\nu_j + 1)C_0^{|\alpha-\nu|}C^{|\nu|+1}$$
$$= \sum_{\nu < \alpha} C_0^{|\alpha-\nu|}C^{|\nu|} + \sum_{\{\nu < \alpha, |\alpha-\nu| > 1\}} (|\nu| + q)C_0^{|\alpha-\nu|}C^{|\nu|+1}.$$

But $C = C_0/\delta$ and $|\alpha| = k$. Hence

$$\|Y_\alpha\|_\infty \leq C^k \left(\sum_{\nu < \alpha} \delta^{|\alpha-\nu|} + C \sum_{\{\nu < \alpha, |\alpha-\nu| > 1\}} (|\nu| + q)\delta^{|\alpha-\nu|} \right)$$
$$\leq C^k \left(\sum_{\nu < \alpha} \delta^{|\alpha-\nu|} + 2kC \sum_{\{\nu < \alpha, |\alpha-\nu| > 1\}} \delta^{|\alpha-\nu|} \right). \tag{4.60}$$

Note that

$$1 + \sum_{\nu < \alpha} \delta^{|\alpha-\nu|} = \sum_{\nu \leq \alpha} \delta^{|\alpha-\nu|} = \sum_{\nu_1=0}^{\alpha_1} \sum_{\nu_2=0}^{\alpha_2} \cdots \sum_{\nu_q=0}^{\alpha_q} \prod_{i=1}^{q} \delta^{(\alpha_i-\nu_i)}$$
$$= \prod_{i=1}^{q} \left(\frac{1-\delta^{\alpha_i+1}}{1-\delta} \right)$$
$$= \prod_{i=1, \alpha_i>0}^{q} \left(\frac{1-\delta^{\alpha_i+1}}{1-\delta} \right)$$

and this shows that

$$\sum_{\nu < \alpha} \delta^{|\alpha - \nu|} \le (1 - \delta)^{-n(\alpha)} - 1, \tag{4.61}$$

where $n(\alpha)$ be the number of non-zero coordinates of α. Similarly

$$\sum_{\{\nu < \alpha, |\alpha - \nu| > 1\}} \delta^{|\alpha - \nu|} = \sum_{\nu \le \alpha} \delta^{|\alpha - \nu|} - 1 - n(\alpha)\delta \le (1 - \delta)^{-n(\alpha)} - 1 - n(\alpha)\delta. \tag{4.62}$$

Since $\delta \in (0, 1/2)$ and $n(\alpha) \le q$, by Taylor's theorem we see that

$$(1 - \delta)^{-n(\alpha)} - 1 \le q2^{q+1}\delta \quad \text{and} \quad \sum_{\nu \le \alpha} \delta^{|\alpha - \nu|} - 1 - n(\alpha)\delta \le q(q+1)2^{q+1}\delta^2.$$

Using these estimates, (4.61), (4.62) and (4.60) we get

$$\|Y_\alpha\|_\infty \le C^k \left(q2^{q+1}\delta + 2kCq(q+1)2^{q+1}\delta^2 \right) = C^k \delta k \left(C_0 q(q+1)2^{q+3} \right).$$

But δ satisfies (4.57) which shows (4.59) and completes the proof of the proposition.

At the beginning of the proof, we had assumed that the eigenvalues $\lambda_1, \ldots, \lambda_q$ of the matrix $[J\theta(h_{\text{eq}})]$ are all real-valued. If that is not true then the invertible matrix P that appears in (4.43) has complex entries. Let $\mathbb{C}$ be the field of complex numbers. Define a map $\phi : \mathbb{R}^q \to \mathbb{C}^q$ by $\phi(h) = P(h - h_{\text{eq}})$. The image of this map, denoted by $\phi(\mathbb{R}^q)$, sits as a q-dimensional real vector space in $\mathbb{C}^q$. The map ϕ is an infinitely differentiable isomorphism between $\mathbb{R}^q$ and $\phi(\mathbb{R}^q)$ and using this we can define derivatives of real-valued functions over $\phi(\mathbb{R}^q)$. As above, we can obtain an analytic solution β of (4.44) satisfying (4.45), defined on some open set U in $\phi(\mathbb{R}^q)$ containing $\bar{0}$. On $V = \phi^{-1}(U)$, the function Λ defined by $\Lambda(h) = \beta(\phi(h))$ will then be an analytic solution to (4.39) satisfying (4.41). $\qquad \square$

The above proposition provides us with an analytic solution to (4.39) in a neighbourhood of h_{eq}. Our next task is to extend it to a solution in $C^2(\hat{U}_{\text{eq}}, \mathbb{R}^q_*)$ where $\hat{U}_{\text{eq}}$ is an open set in $\mathbb{R}^q$ containing U_{eq}.

Recall from Section 2.2 that for all $i, j \in Q$, β_{ij}, ρ_i are functions in $C^2(\mathbb{R}^q_+, \mathbb{R}_+)$. Let $O \subset \mathbb{R}^q$ be the open set containing $\mathbb{R}^q_+$ defined by

$$O = \{h \in \mathbb{R}^q : h_i > -1 \text{ for all } i = 1, \ldots, q\}.$$

Then we can extend the functions β_{ij}, ρ_i to functions $\hat{\beta}_{ij}, \hat{\rho}_i \in C^2(\mathbb{R}^q, \mathbb{R}_+)$ such that $\hat{\beta}_{ij}(h) = 0$ and $\hat{\rho}_i(h) = 0$ for all $h \notin O$. Moreover since each β_{ij} is bounded, we can make sure that its extension $\hat{\beta}_{ij}$ is also bounded. For each $h \in \mathbb{R}^q$ let $\hat{A}(h) \in \mathbb{M}(q, q)$ be the matrix defined by (2.13) with β_{ij}, ρ_i replaced by $\hat{\beta}_{ij}, \hat{\rho}_i$. Also let $\hat{\theta} \in C^2(\mathbb{R}^q, \mathbb{R}^q)$ be the function given by

$$\hat{\theta}(h) = \hat{A}(h)h \text{ for } h \in \mathbb{R}^q. \tag{4.63}$$

Corresponding to $\hat{\theta}$ we can define the *flow* map $\hat{\psi} \in C^2(\mathbb{R}^q \times \mathbb{R}_+, O)$ as the unique solution to the equation analogous to (2.18), with θ replaced by $\hat{\theta}$. Define the region of attraction of the fixed point h_{eq} as

$$\hat{U}_{\text{eq}} = \left\{h \in O : \lim_{t \to \infty} \hat{\psi}(h, t) = h_{\text{eq}}\right\}.$$

Then $\hat{U}_{\text{eq}}$ is an open set in $\mathbb{R}^q$ (see Lemma 3.2 in [24]) containing U_{eq}.

Proposition 4.7. *There exists a solution* $\Lambda \in C^2(\hat{U}_{\mathrm{eq}}, \mathbb{R}_+^q)$ *of* (4.39) *satisfying* (4.40) *and* (4.41).

Proof. Suppose that $U \subset \hat{U}_{\mathrm{eq}}$ is any $\hat{\psi}$-invariant open set and the function $\Lambda \in C^2(U, \mathbb{R}^q)$ satisfies (4.39) and (4.41). We first show that this function automatically satisfies (4.40) on U. Using (4.39) and the $\hat{\psi}$-invariance of U we get

$$\frac{d}{dt}\Lambda(\hat{\psi}(h,t)) = [J\Lambda(\hat{\psi}(h,t))]\hat{\theta}(\hat{\psi}(h,t)) = -\hat{A}^T(\hat{\psi}(h,t))\Lambda(\hat{\psi}(h,t)). \tag{4.64}$$

Observe that

$$\begin{aligned}
\frac{d}{dt}\langle\hat{\psi}(h,t),\Lambda(\hat{\psi}(h,t))\rangle &= \left\langle\frac{d}{dt}\hat{\psi}(h,t),\Lambda(\hat{\psi}(h,t))\right\rangle + \left\langle\hat{\psi}(h,t),\frac{d}{dt}\Lambda(\hat{\psi}(h,t))\right\rangle \\
&= \left\langle\hat{\theta}\left(\hat{\psi}(h,t)\right),\Lambda(\hat{\psi}(h,t))\right\rangle - \left\langle\hat{\psi}(h,t),\hat{A}^T(\hat{\psi}(h,t))\Lambda(\hat{\psi}(h,t))\right\rangle \\
&= \left\langle\hat{\theta}\left(\hat{\psi}(h,t)\right),\Lambda(\hat{\psi}(h,t))\right\rangle - \left\langle\hat{A}(\hat{\psi}(h,t))\hat{\psi}(h,t),\Lambda(\hat{\psi}(h,t))\right\rangle \\
&= 0,
\end{aligned}$$

where the last equality holds due to (4.63). This shows that for any fixed $h \in U$ the function $\langle\hat{\psi}(h,t),\Lambda(\hat{\psi}(h,t))\rangle$ is a constant function of time. Therefore (4.41) implies that for any $h \in U$

$$\langle h,\Lambda(h)\rangle = \lim_{t\to\infty}\langle\hat{\psi}(h,t),\Lambda(\hat{\psi}(h,t))\rangle = \langle h_{\mathrm{eq}},\Lambda(h_{\mathrm{eq}})\rangle = 1.$$

This proves that Λ satisfies (4.40) on U. For any $h \in U$ and $0 \le t \le t_0$, let $\Phi(h,t,t_0)$ be the matrix defined in Lemma A.2. Since Λ satisfies (4.64) we must have

$$\Lambda(\hat{\psi}(h,t)) = \Phi(h,t,t_0)\Lambda(\hat{\psi}(h,t_0)). \tag{4.65}$$

From Proposition 4.6 we know that on some open set $V \subset \mathbb{R}^q$ containing h_{eq} we can find a solution $\bar{\Lambda} \in C^2(V, \mathbb{R}^q)$ that satisfies (4.39) along with (4.41). Since $v_{\mathrm{eq}} \in \mathbb{R}_*^q$ (that is, it is positive component-wise) and $\bar{\Lambda}$ is a continuous function, by shrinking V if necessary, we can ensure that the image of V under $\bar{\Lambda}$ lies in $\mathbb{R}_*^q$. Since V is open, there exists a $r \in (0,1)$ such that $B_r^q(h_{\mathrm{eq}}) \subset V$, where $B_r^q(h_{\mathrm{eq}})$ is the open ball in $\mathbb{R}^q$ centered at h_{eq} with radius r. As in the proof of Lemma 4.5, we can find a $\hat{\psi}$-invariant open set $W \subset B_r^q(h_{\mathrm{eq}})$ which contains h_{eq}.

For each $n \in \mathbb{N}$ define an open set

$$O_n = \{h \in \hat{U}_{\mathrm{eq}} : \hat{\psi}(h,n) \subset W\}.$$

Each O_n is $\hat{\psi}$-invariant. Furthermore $W \subset O_1 \subset O_2 \ldots$ and $\bigcup_{n=1}^{\infty} O_n = \hat{U}_{\mathrm{eq}}$. Define $\lambda_n(h,t)$ for each $h \in O_n$ and $t \in [0,n)$ by

$$\lambda_n(h,t) = \Phi(h,t,n)\bar{\Lambda}(\hat{\psi}(h,n)). \tag{4.66}$$

Observe that $\hat{\psi}(h,n) \in W \subset V$ and so $\bar{\Lambda}(\hat{\psi}(h,n))$ is well-defined and also $\bar{\Lambda}(\hat{\psi}(h,n)) \in \mathbb{R}_*^q$. Part (C) of Lemma A.2 shows that $\lambda_n(h,t) \in \mathbb{R}_*^q$. Since $\bar{\Lambda} \in C^2(V, \mathbb{R}_*^q)$, $\Phi(\cdot,\cdot,n) \in C^2(\hat{U}_{\mathrm{eq}} \times [0,n], \mathbb{M}_{\mathbb{R}}(q,q))$ (see Lemma A.2) and $\hat{\psi} \in C^2(\mathbb{R}^q \times \mathbb{R}_+, \mathbb{R}^q)$ we must have that $\lambda_n \in C^2(O_n \times [0,n), \mathbb{R}_*^q)$. Note that $\bar{\Lambda}$ satisfies (4.65) for all $h \in W$ and so for $0 \le t \le t_0$

$$\bar{\Lambda}(\hat{\psi}(h,t)) = \Phi(h,t,t_0)\bar{\Lambda}(\hat{\psi}(h,t_0)). \tag{4.67}$$

Therefore if $h \in W$, then for any $n \in \mathbb{N}$ and $t \ge 0$ we have

$$\lambda_n(h,t) = \bar{\Lambda}(\hat{\psi}(h,t)). \tag{4.68}$$

Using parts (A) and (B) of Lemma A.2, (4.67) and the semigroup property of $\hat{\psi}$ (similar to (2.19)) we can also see that for any $h \in O_n$

$$
\begin{aligned}
\lambda_n(\hat{\psi}(h,t),0) &= \Phi(\hat{\psi}(h,t),0,n)\bar{\Lambda}\left(\hat{\psi}(\hat{\psi}(h,t),n)\right) \\
&= \Phi(h,t,n+t)\bar{\Lambda}\left(\hat{\psi}(h,n+t)\right) \\
&= \Phi(h,t,n)\Phi(h,n,n+t)\bar{\Lambda}(\hat{\psi}(h,n+t)) \\
&= \Phi(h,t,n)\bar{\Lambda}(\hat{\psi}(h,n)) \\
&= \lambda_n(h,t).
\end{aligned}
$$

Let $h \in O_n$ and $m \geq n$. Then $\hat{\psi}(h,n) \in W$. From part (A) of Lemma A.2 and (4.67) we can deduce that for any $t \in [0,n)$

$$
\begin{aligned}
\lambda_m(h,t) = \Phi(h,t,m)\bar{\Lambda}(\hat{\psi}(h,m)) &= \Phi(h,t,n)\Phi(h,n,m)\bar{\Lambda}(\hat{\psi}(h,m)) \\
&= \Phi(h,t,n)\bar{\Lambda}(\hat{\psi}(h,n)) \\
&= \lambda_n(h,t).
\end{aligned}
$$

Hence if we define the map $\lambda : \hat{U}_{\mathrm{eq}} \times \mathbb{R}_+ \to \mathbb{R}_*^q$ by

$$
\lambda(h,t) = \lambda_n(h,t) \text{ if } (h,t) \in O_n \times [0,n),
$$

then λ is a well-defined function in $C^2(\hat{U}_{\mathrm{eq}} \times \mathbb{R}_+, \mathbb{R}_*^q)$ which satisfies

$$
\lambda(h,t) = \lambda(\hat{\psi}(h,t),0) \text{ for all } (h,t) \in \hat{U}_{\mathrm{eq}} \times \mathbb{R}_+. \tag{4.69}
$$

From (4.66) and the definition of the matrix Φ we can see that

$$
\frac{d\lambda(h,t)}{dt} = -\hat{A}^T(\hat{\psi}(h,t))\lambda(h,t). \tag{4.70}
$$

Define $\Lambda : \hat{U}_{\mathrm{eq}} \to \mathbb{R}_*^q$ by

$$
\Lambda(h) = \lambda(h,0).
$$

Then this map is in $C^2(\hat{U}_{\mathrm{eq}}, \mathbb{R}_*^q)$ and (4.69) implies that for any $(h,t) \in \hat{U}_{\mathrm{eq}} \times \mathbb{R}_+$

$$
\frac{d\lambda(h,t)}{dt} = \frac{d\lambda(\hat{\psi}(h,t),0)}{dt} = \frac{d}{dt}\Lambda(\hat{\psi}(h,t)) = [J\Lambda(\hat{\psi}(h,t))]\frac{d\hat{\psi}(h,t)}{dt} = [J\Lambda(\hat{\psi}(h,t))]\hat{\theta}(\hat{\psi}(h,t)).
$$

Using (4.70) we obtain

$$
[J\Lambda(\hat{\psi}(h,t))]\hat{\theta}(\hat{\psi}(h,t)) = -\hat{A}^T(\hat{\psi}(h,t))\lambda(h,t) = -\hat{A}^T(\hat{\psi}(h,t))\Lambda(\hat{\psi}(h,t)).
$$

If we set $t = 0$ then we see that Λ is a solution to (4.39). Since $\bar{\Lambda}$ satisfies (4.41), equation (4.68) implies that Λ will also satisfy it. We have already shown that such a solution of (4.39) will automatically satisfy (4.40) for all $h \in \hat{U}_{\mathrm{eq}}$. This completes the proof of the proposition. $\qquad\square$

4.4 Fleming-Viot convergence

In this section we will finally prove the main result of our paper, which is Theorem 2.6. Let $\{\mu^N(t) : t \geq 0\}$ be a $\mathcal{M}_{N,a}^q(E)$-valued process with generator $\mathbf{A}_l^N$ for some $l \in \{0,1,2,3\}$. As outlined at the end of Section 2.4, we first *extract* a $\mathcal{P}(E)$-valued process $\{\nu^N(t) : t \geq 0\}$ from the process $\{\mu^N(t) : t \geq 0\}$. This step requires a solution Λ of (4.39) whose existence was shown in Section 4.3. We then show that as $N \to \infty$,

we have $\nu^N \Rightarrow \nu$ where $\{\nu(t) : t \geq 0\}$ is an appropriately defined Fleming-Viot process. This convergence and Proposition 2.5 prove Theorem 2.6. Before we proceed we need some preliminary results.

For any set $A \subset \mathbb{R}_+^q$ define

$$\mathcal{M}_F^q(E : A) = \left\{ \mu \in \mathcal{M}_F^q(E) : H(\mu) \in A \right\}.$$

Note that if A is a compact set then the set $\mathcal{M}_F^q(E : A)$ is also compact.

Recall the definition of the set U_{eq} from (2.20). Let $\{F_N : N \in \mathbb{N}\}$ be a sequence of real-valued functions on $\mathcal{M}_F^q(E : U_{\text{eq}})$. We will say that this sequence belongs to class $o(N^{-m})$ for some $m \in \mathbb{N}_0$, if and only if for each compact $K \subset U_{\text{eq}}$ we have

$$\limsup_{N \to \infty} \sup_{\mu \in \mathcal{M}_F^q(E:K)} N^m |F_N(\mu)| = 0.$$

For two such sequences $\{F_N : N \in \mathbb{N}\}$ and $\{G_N : N \in \mathbb{N}\}$, we say that $F_N(\mu) = G_N(\mu) + o(N^{-m})$ for all $\mu \in \mathcal{M}_F^q(E : U_{\text{eq}})$ if and only if the sequence of functions $\{(F_N - G_N) : N \in \mathbb{N}\}$ is in the class $o(N^{-m})$.

From now on let $\Lambda \in C^2(\hat{U}_{\text{eq}}, \mathbb{R}_*^q)$ be a function that satisfies (4.39), (4.40) and (4.41) on some open set $\hat{U}_{\text{eq}} \subset \mathbb{R}^q$ containing U_{eq}. Such a function exists by Proposition 4.7. Define a continuous map $\Gamma : \mathcal{M}_F^q(E : U_{\text{eq}}) \to \mathcal{P}(E)$ by

$$\Gamma(\mu) = \nu, \tag{4.71}$$

where the measure ν is given by

$$\nu(S) = \sum_{i \in Q} \Lambda_i(h) \mu_i(S) \text{ for any } S \in \mathcal{B}(E), \tag{4.72}$$

with $h = H(\mu)$ being the density vector corresponding to μ. Note that for each $h \in U_{\text{eq}}$, $\Lambda(h)$ is a vector which is positive in each component and hence $\nu(S) \geq 0$ for all $S \in \mathcal{B}(E)$. Since the function Λ satisfies (4.40) we have

$$\nu(E) = \sum_{i \in Q} \Lambda_i(h) \mu_i(E) = \sum_{i \in Q} \Lambda_i(h) h_i = 1.$$

This shows that ν is a probability measure on E.

Let Υ' be the class of functions in $C\left(\mathcal{M}_F^q(E : U_{\text{eq}})\right)$ given by

$$\Upsilon' = \{F(\mu) = \left(h_j \langle f, \mu_i \rangle - h_i \langle f, \mu_j \rangle\right) L(\mu) : (h_1, \ldots, h_q) = H(\mu), \tag{4.73}$$
$$f \in C(E), L \in C\left(\mathcal{M}_F^q(E : U_{\text{eq}})\right) \text{ and } i, j \in Q\}.$$

Let Υ be the smallest algebra of functions in $C\left(\mathcal{M}_F^q(E : U_{\text{eq}})\right)$ containing Υ'. Observe that if $G \in C\left(\mathcal{M}_F^q(E : U_{\text{eq}})\right)$ and $L \in \Upsilon$, then the product GL is in Υ. Given two functions $G_1, G_2 \in C\left(\mathcal{M}_F^q(E : U_{\text{eq}})\right)$ we say that $G_2(\mu) = G_1(\mu) + \Upsilon$ for all $\mu \in \mathcal{M}_F^q(E : U_{\text{eq}})$ if and only if the function $(G_2 - G_1)$ is in the class Υ.

Let $F \in C(\mathcal{P}(E))$ be a function in the class $\mathcal{C}_0$ defined by (2.4). Then F has the form

$$F(\nu) = \prod_{j=1}^m \langle f_j, \nu \rangle, \tag{4.74}$$

where $f_1, \ldots, f_m \in \mathcal{D}_0$. Corresponding to F, define the functions $F_l, F_{lk} \in \mathcal{C}_0$ for all distinct $l, k \in Q$ by

$$F_l(\nu) = \prod_{j=1, j \neq l}^m \langle f_j, \nu \rangle \text{ and } F_{lk}(\nu) = \prod_{j=1, j \neq l, k}^m \langle f_j, \nu \rangle. \tag{4.75}$$

Using any $F \in \mathcal{C}_0$ we construct a function $\hat{F} \in \mathcal{C}_0^q$ as follows. We first extend the definition of Λ to the whole of $\mathbb{R}^q$ by letting $\Lambda(h) = \bar{0}_q$ for all $h \notin \hat{U}_{\mathrm{eq}}$. If F has the form (4.74) then consider the function $\hat{F} : \mathcal{M}_F^q(E) \to \mathbb{R}$ given by

$$\hat{F}(\mu) = \prod_{j=1}^m \left(\sum_{i \in Q} \Lambda_i(h) \langle f_j, \mu_i \rangle \right), \qquad (4.76)$$

where $h = H(\mu)$. Due to (4.40), the function $\hat{F}$ is in the class $\mathcal{C}_0^q$ defined by (2.11). The next result demonstrates how the action of various operators on functions of the form (4.76) can be approximated.

Proposition 4.8. *Let $F \in \mathcal{C}_0$ have the form (4.74). Corresponding to F let $\hat{F} \in \mathcal{C}_0^q$ have the form (4.76) and for distinct $l, k \in Q$ let F_l, F_{lk} be given by (4.75). Then for all $\mu \in \mathcal{M}_F^q(E : U_{\mathrm{eq}})$ with $h = H(\mu)$ and $\nu = \Gamma(\mu)$ we have the following.*

(A) Let $\mathbf{R}^N$ be the operator given by (2.21). Then

$$N\mathbf{R}^N \hat{F}(\mu) = \sum_{1 \le l \ne k \le m} \gamma(h) \left(\langle f_l f_k, \nu \rangle - \langle f_l, \nu \rangle \langle f_k, \nu \rangle \right) F_{lk}(\nu) + \Upsilon + o(1), \qquad (4.77)$$

where

$$\gamma(h) = \frac{1}{2} \left[\sum_{i,j \in Q} \beta_{ij}(h) h_i (\Lambda_j(h))^2 + \sum_{i \in Q} \rho_i(h) h_i (\Lambda_i(h))^2 \right]. \qquad (4.78)$$

(B) Let $\mathbf{B}^N$ be the operator given by (2.12). Then

$$\mathbf{B}^N \hat{F}(\mu) = \sum_{l=1}^m \left(\sum_{i \in Q} \Lambda_i(h) h_i \langle B_i f_l, \nu \rangle \right) F_l(\nu) + \Upsilon + o(1). \qquad (4.79)$$

(C) Let $\mathbf{G}_1^N$ be the operator given by (2.28). Then

$$\mathbf{G}_1^N \hat{F}(\mu) = \sum_{l=1}^m \left[\left(\langle b^s(\cdot, h) f_l(\cdot), \nu \rangle - \langle b^s(\cdot, h), \nu \rangle \langle f_l, \nu \rangle \right) \right. \qquad (4.80)$$
$$\left. + \left(\langle d^s(\cdot, h), \nu \rangle \langle f_l, \nu \rangle - \langle d^s(\cdot, h) f_l(\cdot), \nu \rangle \right) \right] F_l(\nu) + \Upsilon + o(1),$$

where for any $x \in E$ and $h \in \mathbb{R}_+^q$

$$b^s(x, h) = \sum_{i,j \in Q} b_{ij}^s(x, h) \Lambda_j(h) h_i \quad \text{and} \quad d^s(x, h) = \sum_{i \in Q} d_i^s(x, h) \Lambda_i(h) h_i. \qquad (4.81)$$

(D) Let $\mathbf{G}_2^N$ be the operator given by (2.29). Then

$$\mathbf{G}_2^N \hat{F}(\mu) = \sum_{l=1}^m \left(\sum_{i,j \in Q} \beta_{ij}(h) \Lambda_j(h) h_i \langle C_{ij} f_l, \nu \rangle \right) F_l(\nu) + \Upsilon + o(1), \qquad (4.82)$$

where the operators C_{ij} are as in Assumption 2.2.

(E) Let $\mathbf{G}_3^N$ be the operator given by (2.30). Then

$$\mathbf{G}_3^N \hat{F}(\mu) = \sum_{l=1}^m \left(\sum_{i \in Q} \kappa_i(h) \Lambda_i(h) \int_E (f_l(x) - \langle f_l, \nu \rangle) \Theta_i(dx) \right) F_l(\nu) + \Upsilon + o(1). \qquad (4.83)$$

Proof. For any $j \in Q$, let e_j be the vector in $\mathbb{R}^q$ of the form $e_j = (0, \ldots, 0, 1, 0, \ldots, 0)$ with the 1 at the j-th position. Since $\Lambda \in C^2(\hat{U}_{\text{eq}}, \mathbb{R}^q_+)$ and $\hat{U}_{\text{eq}}$ is an open set containing U_{eq}, if $h \in U_{\text{eq}}$, then using Taylor's theorem we can write

$$\Lambda_i \left(h \pm \frac{1}{N} e_j \right) = \Lambda_i(h) \pm \frac{1}{N} \frac{\partial \Lambda_i(h)}{\partial h_j} + \frac{1}{2N^2} \frac{\partial^2 \Lambda_i(h)}{\partial h_j^2} + o(N^{-2})$$

for any $i, j \in Q$. But then for any $\mu \in \mathcal{M}_F(E : U_{\text{eq}})$ and $x \in E$

$$
\begin{aligned}
\hat{F} \left(\mu \pm \frac{1}{N} \delta_x^j \right) &= \prod_{l=1}^{m} \left(\sum_{i \in Q} \Lambda_i \left(h \pm \frac{1}{N} e_j \right) \langle f_l, \mu_i \rangle \pm \Lambda_j \left(h \pm \frac{1}{N} e_j \right) \frac{f_l(x)}{N} \right) \\
&= \prod_{l=1}^{m} \Bigg[\sum_{i \in Q} \Lambda_i(h) \langle f_l, \mu_i \rangle \pm \frac{1}{N} \left(\sum_{i \in Q} \frac{\partial \Lambda_i(h)}{\partial h_j} \langle f_l, \mu_i \rangle + \Lambda_j(h) f_l(x) \right) \\
&\quad + \frac{1}{N^2} \left(\frac{1}{2} \sum_{i \in Q} \frac{\partial \Lambda_i^2(h)}{\partial h_j^2} \langle f_l, \mu_i \rangle + \frac{\partial \Lambda_j(h)}{\partial h_j} f_l(x) \right) + o(N^{-2}) \Bigg] \\
&= \hat{F}(\mu) \pm \frac{1}{N} \sum_{l=1}^{m} \chi_l^j(\mu, x) F_l(\nu) + \frac{1}{N^2} \sum_{l=1}^{m} \phi_l^j(\mu, x) F_l(\nu) \\
&\quad + \frac{1}{2N^2} \sum_{1 \leq l \neq k \leq m} \chi_l^j(\mu, x) \chi_k^j(\mu, x) F_{lk}(\nu) + o(N^{-2}),
\end{aligned}
$$

where F_l, F_{lk} are as in (4.75) and $\chi_l^j(\mu, x), \phi_l^j(\mu, x)$ are given by

$$\chi_l^j(\mu, x) = \sum_{i \in Q} \frac{\partial \Lambda_i(h)}{\partial h_j} \langle f_l, \mu_i \rangle + \Lambda_j(h) f_l(x)$$

$$\text{and} \quad \phi_l^j(\mu, x) = \frac{1}{2} \sum_{i \in Q} \frac{\partial \Lambda_i^2(h)}{\partial h_j^2} \langle f_l, \mu_i \rangle + \frac{\partial \Lambda_j(h)}{\partial h_j} f_l(x). \tag{4.84}$$

On rearranging we obtain

$$
\begin{aligned}
\hat{F} \left(\mu \pm \frac{1}{N} \delta_x^j \right) - \hat{F}(\mu) &= \pm \frac{1}{N} \sum_{l=1}^{m} \chi_l^j(\mu, x) F_l(\nu) + \frac{1}{N^2} \sum_{l=1}^{m} \phi_l^j(\mu, x) F_l(\nu) \\
&\quad + \frac{1}{2N^2} \sum_{1 \leq l \neq k \leq m} \chi_l^j(\mu, x) \chi_k^j(\mu, x) F_{lk}(\nu) + o(N^{-2}).
\end{aligned}
\tag{4.85}
$$

Therefore for any $\mu \in \mathcal{M}_F(E : U_{\text{eq}})$

$$N\mathbf{R}^N\hat{F}(\mu) = N^2 \sum_{i,j \in Q} \int_E \beta_{ij}(h) \left(\hat{F}\left(\mu + \frac{1}{N}\delta_x^j\right) - \hat{F}(\mu) \right) \mu_i(dx)$$

$$+ N^2 \sum_{i \in Q} \int_E \rho_i(h) \left(\hat{F}\left(\mu - \frac{1}{N}\delta_x^i\right) - \hat{F}(\mu) \right) \mu_i(dx)$$

$$= N \sum_{l=1}^m \left(\sum_{i,j \in Q} \beta_{ij}(h) \langle \chi_l^j(\mu,\cdot), \mu_i \rangle - \sum_{i \in Q} \rho_i(h) \langle \chi_l^i(\mu,\cdot), \mu_i \rangle \right) F_l(\nu)$$

$$+ \sum_{l=1}^m \left(\sum_{i,j \in Q} \beta_{ij}(h) \langle \phi_l^j(\mu,\cdot), \mu_i \rangle + \sum_{i \in Q} \rho_i(h) \langle \phi_l^i(\mu,\cdot), \mu_i \rangle \right) F_l(\nu)$$

$$+ \frac{1}{2} \sum_{1 \le l \ne k \le m} \left(\sum_{i,j \in Q} \beta_{ij}(h) \langle \chi_l^j(\mu,\cdot) \chi_k^j(\mu,\cdot), \mu_i \rangle \right.$$

$$\left. + \sum_{i \in Q} \rho_i(h) \langle \chi_l^i(\mu,\cdot) \chi_l^i(\mu,\cdot), \mu_i \rangle \right) F_{lk}(\nu) + o(1).$$

But note that

$$\sum_{i,j \in Q} \beta_{ij}(h) \langle \chi_l^j(\mu,\cdot), \mu_i \rangle + \sum_{i \in Q} \rho_i(h) \langle \chi_l^i(\mu,\cdot), \mu_i \rangle$$

$$= \sum_{r \in Q} \langle f_l, \mu_r \rangle \left(\sum_{j \in Q} \frac{\partial \Lambda_r(h)}{\partial h_j} \left(\sum_{i \in Q} \beta_{ij}(h)h_i - \rho_j(h)h_j \right) + \sum_{j \in Q} \beta_{rj}(h)\Lambda_j(h) - \rho_r(h)\Lambda_r(h) \right)$$

$$= \sum_{r \in Q} \langle f_l, \mu_r \rangle \left(\sum_{j \in Q} \frac{\partial \Lambda_r(h)}{\partial h_j} \theta_j(h) + \sum_{j \in Q} A_{rj}(h)\Lambda_j(h) \right),$$

where the matrix $A(h)$ and the vector $\theta(h)$ are defined by (2.13) and (2.14). Since the function Λ satisfies (4.39), the expression on the right is just 0. Hence the formula for $N\mathbf{R}^N\hat{F}(\mu)$ simplifies to

$$N\mathbf{R}^N\hat{F}(\mu)$$

$$= \sum_{l=1}^m \left(\sum_{i,j \in Q} \beta_{ij}(h) \langle \phi_l^j(\mu,\cdot), \mu_i \rangle + \sum_{i \in Q} \rho_i(h) \langle \phi_l^i(\mu,\cdot), \mu_i \rangle \right) F_l(\nu) \qquad (4.86)$$

$$+ \frac{1}{2} \sum_{1 \le l \ne k \le m} \left(\sum_{i,j \in Q} \beta_{ij}(h) \langle \chi_l^j(\mu,\cdot) \chi_k^j(\mu,\cdot), \mu_i \rangle + \sum_{i \in Q} \rho_i(h) \langle \chi_l^i(\mu,\cdot) \chi_l^i(\mu,\cdot), \mu_i \rangle \right) F_{lk}(\nu) + o(1).$$

Equation (4.40) says that for all $h \in \hat{U}_{\text{eq}}$

$$\sum_{i \in Q} h_i \Lambda_i(h) = 1.$$

Pick a $j \in Q$. Differentiating the above equation with respect to h_j we get

$$\sum_{i \in Q} h_i \frac{\partial \Lambda_i(h)}{\partial h_j} + \Lambda_j(h) = 0 \qquad (4.87)$$

and differentiating again with respect to h_j we obtain

$$\sum_{i \in Q} h_i \frac{\partial^2 \Lambda_i(h)}{\partial h_j^2} + 2\frac{\partial \Lambda_j(h)}{\partial h_j} = 0. \tag{4.88}$$

Recall that for any $\mu \in \mathcal{M}_F(E : U_{\text{eq}})$, $\nu = \Gamma(\mu)$ is given by (4.72). Using (4.40) one can verify that for any $f \in C(E)$ and $i \in Q$

$$\langle f, \mu_i \rangle = h_i \langle f, \nu \rangle + \sum_{j \in Q} \left(h_j \langle f, \mu_i \rangle - h_i \langle f, \mu_j \rangle \right) \Lambda_j(h). \tag{4.89}$$

But the second term on the right is a function in the class Υ. Hence for all $\mu \in \mathcal{M}_F(E : U_{\text{eq}})$

$$\langle f, \mu_i \rangle = h_i \langle f, \nu \rangle + \Upsilon. \tag{4.90}$$

From the definitions of χ_l^j and ϕ_l^j (see (4.84)) it is immediate that for any $j \in Q$, $l \in \{1, \dots, m\}$ and $x \in E$ we have the following relations. For all $\mu \in \mathcal{M}_F(E : U_{\text{eq}})$

$$\chi_l^j(\mu, x) = \Lambda_j(h) f_l(x) + \left(\sum_{i \in Q} h_i \frac{\partial \Lambda_i(h)}{\partial h_j} \right) \langle f_l, \nu \rangle + \Upsilon$$

$$\text{and} \ \ \phi_l^j(\mu, x) = \frac{\partial \Lambda_j(h)}{\partial h_j} f_l(x) + \left(\frac{1}{2} \sum_{i \in Q} h_i \frac{\partial \Lambda_i^2(h)}{\partial h_j^2} \right) \langle f_l, \nu \rangle + \Upsilon.$$

Using (4.87) and (4.88) we obtain

$$\chi_l^j(\mu, x) = \Lambda_j(h) \left(f_l(x) - \langle f_l, \nu \rangle \right) + \Upsilon \tag{4.91}$$

$$\text{and} \ \ \phi_l^j(\mu, x) = \frac{\partial \Lambda_j(h)}{\partial h_j} \left(f_l(x) - \langle f_l, \nu \rangle \right) + \Upsilon. \tag{4.92}$$

Recall that the class Υ is invariant under multiplication by functions in $C(\mathcal{M}_F(E : U_{\text{eq}}))$. It can be checked that for any $i, j \in Q$ and $l, k \in \{1, \dots, m\}$

$$\langle \chi_l^j(\mu, \cdot) \chi_k^j(\mu, \cdot), \mu_i \rangle = h_i (\Lambda_j(h))^2 \left(\langle f_l f_k, \nu \rangle - \langle f_l, \nu \rangle \langle f_k, \nu \rangle \right) + \Upsilon$$

and the function $\mu \mapsto \langle \phi_l^j(\mu, \cdot), \mu_i \rangle$ belongs to class Υ. Substituting these two relations in (4.86) proves part (A) of this proposition.

Recall the definition of the operator $\mathbf{B}_i^n$ from Section 2.1. If $G(\nu) = \prod_{j=1}^l \langle g_j, \nu \rangle \in \mathcal{C}_0$ then one can verify (see Section 2.2 in [7]) that there is a constant c (depending on l and $g_1, \dots, g_l$) such that

$$\sup_{n \in \mathbb{N}} \sup_{\nu \in \mathcal{P}_{n,a}} \left(n \left| \mathbf{B}_i^n G(\nu) - \sum_{j=1}^l \langle B_i g_j, \nu \rangle \prod_{k=1, k \neq j}^l \langle g_k, \nu \rangle \right| \right) \leq c.$$

From the definition of the operator $\mathbf{B}^N$ and (4.90) it is immediate that

$$\mathbf{B}^N \hat{F}(\mu) = \sum_{l=1}^m \left(\sum_{i \in Q} \Lambda_i(h) \langle B_i f_l, \mu_i \rangle \right) F_l(\nu) + o(1)$$

$$= \sum_{l=1}^m \left(\sum_{i \in Q} \Lambda_i(h) h_i \langle B_i f_l, \nu \rangle \right) F_l(\nu) + \Upsilon + o(1).$$

This proves part (B) of the proposition.

Using (4.85), (4.91) and (4.90) we get

$$
\begin{aligned}
\mathbf{G}_1^N \hat{F}(\mu) = N & \left(\sum_{i,j \in Q} \int_E b_{ij}^s(x,h) \left(\hat{F}\left(\mu + \frac{1}{N} \delta_x^j \right) - \hat{F}(\mu) \right) \mu_i(dx) \right. \\
& \left. + \sum_{i \in Q} \int_E d_i^s(x,h) \left(\hat{F}\left(\mu - \frac{1}{N} \delta_x^i \right) - \hat{F}(\mu) \right) \mu_i(dx) \right) \\
= \sum_{l=1}^m & \left(\sum_{i,j \in Q} \left\langle b_{ij}^s(\cdot,h) \chi_l^j(\mu,\cdot), \mu_i \right\rangle \right. \\
& \left. - \sum_{i \in Q} \left\langle d_i^s(\cdot,h) \chi_l^i(\mu,\cdot), \mu_i \right\rangle \right) F_l(\nu) + \Upsilon + o(1) \\
= \sum_{l=1}^m & \left(\sum_{i,j \in Q} \Lambda_j(h) \left\langle b_{ij}^s(\cdot,h) \left(f_l(\cdot) - \langle f_l, \nu \rangle \right), \mu_i \right\rangle \right. \\
& \left. - \sum_{i \in Q} \Lambda_i(h) \left\langle d_i^s(\cdot,h) \left(f_l(\cdot) - \langle f_l, \nu \rangle \right), \mu_i \right\rangle \right) F_l(\nu) + \Upsilon + o(1) \\
= \sum_{l=1}^m & \left(\sum_{i,j \in Q} \Lambda_j(h) h_i \left\langle b_{ij}^s(\cdot,h) \left(f_l(\cdot) - \langle f_l, \nu \rangle \right), \nu \right\rangle \right. \\
& \left. - \sum_{i \in Q} \Lambda_i(h) h_i \left\langle d_i^s(\cdot,h) \left(f_l(\cdot) - \langle f_l, \nu \rangle \right), \nu \right\rangle \right) F_l(\nu) + \Upsilon + o(1) \\
= \sum_{l=1}^m & \left(\left(\langle b^s(\cdot,h) f_l(\cdot), \nu \rangle - \langle b^s(\cdot,h), \nu \rangle \langle f_l, \nu \rangle \right) \right. \\
& \left. + \left(\langle d^s(\cdot,h), \nu \rangle \langle f_l, \nu \rangle - \langle d^s(\cdot,h) f_l(\cdot), \nu \rangle \right) \right) F_l(\nu) + \Upsilon + o(1),
\end{aligned}
$$

where the functions b^s and d^s are defined in (4.81). This proves part (C).

Observe that Assumption 2.2 implies that for any $f \in \mathcal{D}_0$ and $i, j \in Q$

$$
N \int_E p_{ij}^N(x) \int_E (f(y) - f(x)) \vartheta_{ij}^N(x,dy) \mu_i(dx) = \langle C_{ij} f, \mu_i \rangle + o(1)
$$

and

$$
\int_E p_{ij}^N(x) \int_E (f(y) - f(x)) \vartheta_{ij}^N(x,dy) \mu_i(dx) = o(1).
$$

Using (4.85) , (4.84) and (4.90) we obtain

$$\mathbf{G}_2^N \hat{F}(\mu)$$

$$= N^2 \sum_{i,j \in Q} \beta_{ij}(h) \int_E p_{ij}^N(x) \left(\int_E \left(\hat{F}\left(\mu + \frac{1}{N}\delta_y^j\right) - \hat{F}\left(\mu + \frac{1}{N}\delta_x^j\right) \right) \vartheta_{ij}^N(x, dy) \right) \mu_i(dx)$$

$$= \sum_{l=1}^m \left(N \sum_{i,j \in Q} \beta_{ij}(h) \int_E p_{ij}^N(x) \int_E \left(\chi_l^j(\mu, y) - \chi_l^j(\mu, x) \right) \vartheta_{ij}^N(x, dy) \mu_i(dx) \right) F_l(\nu) + o(1)$$

$$= \sum_{l=1}^m \left(N \sum_{i,j \in Q} \beta_{ij}(h)\Lambda_j(h) \int_E p_{ij}^N(x) \left[\int_E \left(f_l(y) - f_l(x) \right) \vartheta_{ij}^N(x, dy) \right] \mu_i(dx) \right) F_l(\nu) + o(1)$$

$$= \sum_{l=1}^m \left(\sum_{i,j \in Q} \beta_{ij}(h)\Lambda_j(h) \langle C_{ij} f_l, \mu_i \rangle \right) F_l(\nu) + o(1)$$

$$= \sum_{l=1}^m \left(\sum_{i,j \in Q} \beta_{ij}(h)\Lambda_j(h) h_i \langle C_{ij} f_l, \nu \rangle \right) F_l(\nu) + \Upsilon + o(1).$$

This proves part (D).

Again using (4.85), (4.91) and (4.90) we see that

$$\mathbf{G}_3^N \hat{F}(\mu) = N \sum_{i=1}^q \kappa_i(h) \int_E \left(\hat{F}\left(\mu + \frac{1}{N}\delta_x^i\right) - \hat{F}(\mu) \right) \Theta_i(dx)$$

$$= \sum_{l=1}^m \left(\sum_{i \in Q} \kappa_i(h) \int_E \chi_l^i(\mu, x) \Theta_i(dx) \right) F_l(\nu) + o(1)$$

$$= \sum_{l=1}^m \left(\sum_{i \in Q} \kappa_i(h)\Lambda_i(h) \int_E \left(f_l(x) - \langle f_l, \nu \rangle \right) \Theta_i(dx) \right) F_l(\nu) + \Upsilon + o(1).$$

This proves part (E) and completes the proof of this proposition. $\qquad\square$

Remark 4.9. *For any $(h, \nu) \in U_{\text{eq}} \times \mathcal{P}(E)$ let $\zeta_F^R(h, \nu), \zeta_F^B(h, \nu), \zeta_F^{G,1}(h, \nu), \zeta_F^{G,2}(h, \nu)$ and $\zeta_F^{G,3}(h, \nu)$ be the first terms that appear on the right hand side of equations (4.77), (4.79), (4.80), (4.82) and (4.83) respectively. Let $\zeta_F^{G,0}(h, \nu) = 0$ for all $(h, \nu) \in U_{\text{eq}} \times \mathcal{P}(E)$. Recall the definitions of the operators $\mathbf{A}_l$ for $l \in \{0, 1, 2, 3\}$ from (2.39), (2.40), (2.41) and (2.42). One can verify that for each $l \in \{0, 1, 2, 3\}$, $F \in \mathcal{C}_0$ and $\nu \in \mathcal{P}(E)$*

$$\mathbf{A}_l F(\nu) = \zeta_F^B(h_{\text{eq}}, \nu) + \zeta_F^R(h_{\text{eq}}, \nu) + \zeta_F^{G,l}(h_{\text{eq}}, \nu). \tag{4.93}$$

To check this relation observe that Λ satisfies (4.41). Furthermore for each $i \in Q$, $\theta_i(h_{\text{eq}}) = \sum_{j \in Q} \beta_{ji}(h_{\text{eq}}) h_{\text{eq},j} - \rho_i(h_{\text{eq}}) h_{\text{eq},i} = 0$, which shows that the value of the function γ (given by (4.78)) at h_{eq} is equal to the constant γ_{smpl} (given by (2.36)).

We now prove the main theorem of the paper.

Proof of Theorem 2.6. Fix a $l \in \{0, 1, 2, 3\}$ and let $\{\mu^N(t) : t \geq 0\}$ be a solution to the martingale problem for $\mathbf{A}_l^N$. Let $\{h^N(t) = H\left(\mu^N(t)\right) : t \geq 0\}$ be the corresponding density process. Since $\mu^N(0) \Rightarrow \mu(0)$ as $N \to \infty$ and $H(\mu(0)) \in U_{\text{eq}}$ a.s. we must have that $h^N(0) \Rightarrow h(0)$ and $h(0) \in U_{\text{eq}}$ a.s. It suffices to prove the theorem under the assumption that for all $N \in \mathbb{N}$, $h^N(0) \in K_0$ a.s. for some compact $K_0 \subset U_{\text{eq}}$.

By Lemma 4.5, we can find a bigger compact set $K \subset U_{\text{eq}}$ containing K_0 such that if we define the stopping time σ^N by

$$\sigma^N = \inf \left\{ t \geq 0 : h^N(t) \notin K \right\}$$

then $\sigma^N \Rightarrow \infty$ as $N \to \infty$.

Let $F \in \mathcal{C}_0$ be a function of the form (4.74) and let $\hat{F} \in \mathcal{C}_0^q$ have the form (4.76). From (2.31) and Proposition 4.8 we can conclude that for all $\mu \in \mathcal{M}_F(E : U_{\text{eq}})$

$$\mathbf{A}_l^N \hat{F}(\mu) = \zeta_F^B(h, \nu) + \zeta_F^R(h, \nu) + \zeta_F^{G,l}(h, \nu) + \Upsilon + o(1), \tag{4.94}$$

where $h = H(\mu)$, $\nu \in \Gamma(\mu)$ and the continuous functions $\zeta_F^B, \zeta_F^R, \zeta_F^{G,0}, \zeta_F^{G,1}, \zeta_F^{G,2}, \zeta_F^{G,3}$ from $U_{\text{eq}} \times \mathcal{P}(E)$ to $\mathbb{R}$ are defined in Remark 4.9.

The relation (4.94) implies that

$$\sup_{N \in \mathbb{N}} \sup_{\mu \in \mathcal{M}_F(E:K)} |\mathbf{A}_l^N \hat{F}(\mu)| < \infty. \tag{4.95}$$

The function $\hat{F}$ belongs to the domain of $\mathbf{A}_l^N$ and the process $\{\mu^N(t) : t \geq 0\}$ is a solution to the martingale problem for $\mathbf{A}_l^N$. Hence

$$\hat{F}(\mu^N(t)) - \hat{F}^N(\mu^N(0)) - \int_0^t \mathbf{A}_l^N \hat{F}(\mu^N(s)) ds$$

is a martingale and using the optional sampling theorem we see that

$$m_F^N(t) = \hat{F}(\mu^N(t \wedge \sigma^N)) - \hat{F}^N(\mu^N(0)) - \int_0^{t \wedge \sigma^N} \mathbf{A}_l^N \hat{F}(\mu^N(s)) ds$$

is also a martingale. Note that for all $t \geq 0$, $\mu^N(t \wedge \sigma^N)$ is in the set $\mathcal{M}_F(E : U_{\text{eq}})$. Define a $\mathcal{P}(E)$-valued process by

$$\nu^N(t) = \Gamma\left(\mu^N(t \wedge \sigma^N) \right) \text{ for } t \geq 0. \tag{4.96}$$

The form of the functions F and $\hat{F}$ shows that $\hat{F}(\mu^N(t \wedge \sigma^N)) = F(\nu^N(t))$ for any $t \geq 0$. Hence the martingale m_F^N can be rewritten as

$$m_F^N(t) = F(\nu^N(t)) - F(\nu^N(0)) - \int_0^{t \wedge \sigma^N} \mathbf{A}_l^N \hat{F}(\mu^N(s)) ds. \tag{4.97}$$

The linear span of functions in the class $\mathcal{C}_0$ is a dense sub-algebra of $C(\mathcal{P}(E))$ and for every $F \in \mathcal{C}_0$ we have the martingale relation (4.97). Theorems 3.9.1 and 3.9.4 in Ethier and Kurtz [12] along with the estimate (4.95) imply that the sequence of processes $\{\nu^N : N \in \mathbb{N}\}$ is tight in the space $D_{\mathcal{P}(E)}[0, \infty)$.

Let t_N be a sequence satisfying the conditions of Theorem 2.6. Pick a $T > 0$. It is easy to see that

$$\sup_{t \in [0,T]} \left| \int_{t_N}^{t \wedge \sigma^N + t_N} \mathbf{A}_l^N \hat{F}(\mu^N(s)) ds - \int_0^{t \wedge \sigma^N} \mathbf{A}_l^N \hat{F}(\mu^N(s)) ds \right| \Rightarrow 0 \text{ as } N \to \infty. \tag{4.98}$$

From parts (A) and (B) of Proposition 2.5 we know that as $N \to \infty$

$$\sup_{t \in [0,T]} \left\| h^N(t + t_N) - h_{\text{eq}} \right\|_1 \Rightarrow 0 \tag{4.99}$$

and for any $f \in C(E)$ and $i, j \in Q$

$$\sup_{t \in [0,T]} \left| h_j^N(t + t_N)\langle f, \mu_i^N(t + t_N)\rangle - h_i^N(t + t_N)\langle f, \mu_j^N(t + t_N)\rangle \right| \Rightarrow 0. \qquad (4.100)$$

Note that this also implies that if L is a function in the class Υ then

$$\sup_{t \in [0,T]} \left| L(\mu^N(t + t_N)) \right| \Rightarrow 0 \text{ as } N \to \infty. \qquad (4.101)$$

We argued before that the sequence of processes $\{\nu^N : N \in \mathbb{N}\}$ is tight. Let $\{\nu(t) : t \geq 0\}$ be a limit point. Then along some sequence k_N, $\nu^N \Rightarrow \nu$ as $N \to \infty$. Since $\sigma^N \Rightarrow \infty$, using the continuous mapping theorem and (4.99) we obtain that along the subsequence k_N

$$\sup_{t \in [0,T]} \left| \int_{t_N}^{t \wedge \sigma^N + t_N} \zeta(h^N(s), \nu^N(s))ds - \int_0^t \zeta(h_{\mathrm{eq}}, \nu(s))ds \right| \Rightarrow 0 \text{ as } N \to \infty, \qquad (4.102)$$

where ζ is any of the continuous functions $\zeta_F^B, \zeta_F^R, \zeta_F^{G,0}, \zeta_F^{G,1}, \zeta_F^{G,2}, \zeta_F^{G,3}$ defined in Remark 4.9. From (4.94), (4.98), (4.101), (4.102) and (4.93) we get that along the subsequence k_N

$$\sup_{t \in [0,T]} \left| \int_0^{t \wedge \sigma^N} \mathbf{A}_l^N \hat{F}(\mu^N(s))ds - \int_0^t \mathbf{A}_l F(\nu(s))ds \right| \Rightarrow 0 \text{ as } N \to \infty. \qquad (4.103)$$

Using (4.103) and the continuous mapping theorem we can conclude that for any $F \in \mathcal{C}_0$, as $N \to \infty$, the sequence of martingales m_F^N (given by (4.97)) converges in distribution along the subsequence k_N to the martingale given by

$$F(\nu(t)) - F(\nu(0)) - \int_0^t \mathbf{A}_l F(\nu(s))ds. \qquad (4.104)$$

This shows that $\{\nu(t) : t \geq 0\}$ is a solution to the martingale problem for $\mathbf{A}_l$. Let $\pi \in \mathcal{P}(\mathcal{P}(E))$ be the distribution of $\Gamma(\mu(0))$. Since $\mu^N(0) \Rightarrow \mu(0)$ as $N \to \infty$ and Γ is a continuous map we must also have that $\nu^N(0) \Rightarrow \nu(0)$, where $\nu(0)$ has distribution π. We argued in Section 2.4 that the martingale problem for each $\mathbf{A}_l$ is well-posed. Hence $\{\nu(t) : t \geq 0\}$ is the unique solution to the martingale problem for $(\mathbf{A}_l, \pi)$ and thus $\nu^N \Rightarrow \nu$ as $N \to \infty$, along the entire sequence. Moreover the limiting process has sample paths in $C_{\mathcal{P}(E)}[0, \infty)$ almost surely.

Let $\{\hat{\mu}^N(t) : t \geq 0\}$ be the process defined by (2.43). Pick a $i \in Q$ and $f \in C(E)$. From (4.89) for any $0 \leq t < \sigma^N - t_N$ we can write

$$
\begin{aligned}
\langle f, \hat{\mu}_i^N(t)\rangle &= \langle f, \mu_i^N(t + t_N)\rangle \\
&= h_i^N(t + t_N)\langle f, \nu^N(t + t_N)\rangle \\
&\quad + \sum_{j \in Q} \left(h_j^N(t + t_N)\langle f, \mu_i^N(t + t_N)\rangle - h_i^N(t + t_N)\langle f, \mu_j^N(t + t_N)\rangle \right) \Lambda_j(h^N(t + t_N)).
\end{aligned}
$$

Since $\nu^N \Rightarrow \nu$, $\sigma^N \Rightarrow \infty$ and $t_N \to 0$, (4.99), (4.100) and the continuity of the sample paths of $\{\nu(t) : t \geq 0\}$ imply that for any $T > 0$

$$\sup_{t \in [0,T]} \left| \langle f, \hat{\mu}_i^N(t)\rangle - h_{\mathrm{eq},i}\langle f, \nu(t)\rangle \right| \Rightarrow 0 \text{ as } N \to \infty.$$

This holds for any $f \in C(E)$ and $i \in Q$. Hence by Theorem 3.7.1 in Dawson[7], $\hat{\mu}^N \Rightarrow h_{\mathrm{eq}}\nu$ as $N \to \infty$ in the Skorohod topology on $D_{\mathcal{M}_F^q(E)}[0, \infty)$. This completes the proof of Theorem 2.6. $\qquad \square$

Remark 4.10. *In the statement of Theorem 2.6 we did not specify how the initial distribution π of the limiting Fleming-Viot process $\{\nu(t) : t \geq 0\}$ is related to the distribution of $\mu(0)$. However the above proof makes it clear that π is the distribution of $\Gamma(\mu(0))$ where Γ is the map defined by (4.71).*

A Appendix.

Lemma A.1. *For each $h \in \mathbb{R}_+^q$ let $A(h) \in \mathbb{M}(q, q)$ be the matrix given by (2.13). Suppose that Assumption 2.1 is satisfied and let $h_{\text{eq}} \in \mathbb{R}_+^q$ be the nonzero vector such that*

$$A(h_{\text{eq}})h_{\text{eq}} = \bar{0}_q. \tag{A.1}$$

Then we have the following.

(A) *The vector h_{eq} is in $\mathbb{R}_*^q$, that is, all its components are strictly positive.*

(B) *The matrix $A(h_{\text{eq}})$ has 0 as an eigenvalue with algebraic multiplicity 1. All the other eigenvalues of $A(h_{\text{eq}})$ have strictly negative real parts.*

(C) *There exists a unique vector $v_{\text{eq}} \in \mathbb{R}_*^q$ such that $v_{\text{eq}}A(h_{\text{eq}}) = \bar{0}_q$ and $\langle v_{\text{eq}}, h_{\text{eq}} \rangle = 1$.*

(D) *Let $G(h_{\text{eq}}) \in \mathbb{M}(q, q)$ be the matrix given by*

$$G(h_{\text{eq}}) = \left(I_q - \frac{h_{\text{eq}}}{\langle \bar{1}_q, h_{\text{eq}} \rangle} \bar{1}_q^T \right) A(h_{\text{eq}}).$$

Then the matrix $G(h_{\text{eq}})$ has the same eigenvalues as the matrix $A(h_{\text{eq}})$.

(E) *Let $\bar{G}(h_{\text{eq}}) \in \mathbb{M}(q-1, q-1)$ be the matrix defined by*

$$\bar{G}_{ij}(h_{\text{eq}}) = G_{ij}(h_{\text{eq}}) - G_{iq}(h_{\text{eq}}) \text{ for all } i, j \in \{1, \ldots, q-1\}.$$

Then the matrix $\bar{G}(h_{\text{eq}})$ is stable, that is, all its eigenvalues have strictly negative real parts.

Proof. Observe that all the non-diagonal entries of the matrix $A(h_{\text{eq}})$ are nonnegative. Such matrices are sometimes referred to as Metzler-Leontief matrices in mathematical economics (see Section 2.3 in Seneta [32]). Their important property is that they can be transformed to a nonnegative matrix by adding a constant multiple of the identity matrix. This allows extensions of the Perron-Frobenius type results to such matrices.

Note that the matrix $A(h_{\text{eq}})$ is irreducible (part (C) of Assumption 2.1) and has 0 as an eigenvalue with h_{eq} as a right eigenvector (see (A.1)). Theorem 2.6 in [32] proves parts (A),(B) and (C) of this lemma.

We now prove parts (D) and (E). Let P and its inverse P^{-1} be the matrices in $\mathbb{M}(q, q)$ given by

$$P = \begin{bmatrix} I_{q-1} & \bar{1}_{q-1} \\ \bar{0}_{q-1}^T & 1 \end{bmatrix} \quad \text{and} \quad P^{-1} = \begin{bmatrix} I_{q-1} & -\bar{1}_{q-1} \\ \bar{0}_{q-1}^T & 1 \end{bmatrix}.$$

Observe that $\bar{1}_q^T G(h_{\text{eq}}) = \bar{0}_q$ and

$$P^T G(h_{\text{eq}})(P^T)^{-1} = \begin{bmatrix} \bar{G}(h_{\text{eq}}) & v \\ \bar{0}_{q-1}^T & 0 \end{bmatrix}, \tag{A.2}$$

where v is some vector in $\mathbb{R}^{q-1}$. Let $L = \text{Diag}(h_{\text{eq}})$ and $Q = P^{-1}L^{-1}$. Note that $Qh_{\text{eq}} = P^{-1}L^{-1}h_{\text{eq}} = P^{-1}\bar{1}_q = \bar{0}_q$ and hence

$$Q\left(I_q - \frac{h_{\text{eq}}}{\langle \bar{1}_q, h_{\text{eq}} \rangle} \bar{1}_q^T \right) = Q - \frac{1}{\langle \bar{1}_q, h_{\text{eq}} \rangle}[Qh_{\text{eq}}]\bar{1}_q^T = Q.$$

This shows that $QG(h_{\text{eq}})Q^{-1} = QA(h_{\text{eq}})Q^{-1}$. Hence the matrices $G(h_{\text{eq}})$ and $A(h_{\text{eq}})$ are similar and have the same eigenvalues. Part (B) of this lemma and (A.2) imply that the matrix $\bar{G}(h_{\text{eq}})$ is stable. This completes the proof of this lemma. □

Let $\hat{A}$, $\hat{\theta}$, $\hat{\psi}$ and $\hat{U}_{\text{eq}}$ be as defined in Section 4.3 (just prior to Proposition 4.7).

Lemma A.2. *Fix a $t_0 > 0$. For $(h, t) \in \hat{U}_{\text{eq}} \times [0, t_0]$ consider the following matrix equation*

$$\Phi(h, t, t_0) = I_q + \int_t^{t_0} \hat{A}^T(\hat{\psi}(h, u))\Phi(h, u, t_0) du, \tag{A.3}$$

where I_q is the $q \times q$ identity matrix. This equation has a unique solution in $C^2(\hat{U}_{\text{eq}} \times [0, t_0], \mathbb{M}_{\mathbb{R}}(q, q))$ that satisfies the following for any $h \in \hat{U}_{\text{eq}}$, $s \geq 0$ and $0 \leq t \leq t_0$.

(A) *If $s \in [t, t_0]$ then $\Phi(h, t, t_0) = \Phi(h, t, s)\Phi(h, s, t_0)$.*
(B) *For any $s \geq 0$, $\Phi(\hat{\psi}(h, s), t, t_0) = \Phi(h, t + s, t_0 + s)$.*
(C) *If $v_0 \in \mathbb{R}_*^q$ then $\Phi(h, t, t_0)v_0 \in \mathbb{R}_*^q$.*

Proof. The function $\hat{\psi}$ is in $C^2(\mathbb{R}^q \times \mathbb{R}_+, \mathbb{R}^q)$ and the matrix-valued function $(h, t) \mapsto \hat{A}^T(\hat{\psi}(h, t))$ is in $C^2(\mathbb{R}^q \times \mathbb{R}_+, \mathbb{M}(q, q))$. Standard existence and uniqueness results for ordinary differential equations guarantee that there is a unique solution for (A.3) in the class $C^2(\hat{U}_{\text{eq}} \times [0, t_0], \mathbb{M}_{\mathbb{R}}(q, q))$.

Part (A) of the lemma is just the Chapman-Kolmogorov property (see Proposition 2.12 in Chicone [6]). Note that due to the semigroup property for $\hat{\psi}$ (similar to (2.19)) both $\Phi(\hat{\psi}(h, s), t, t_0)$ and $\Phi(h, t + s, t_0 + s)$ satisfy the same equation for $t \in [0, t_0]$. Hence by uniqueness of solutions, part (B) is immediate.

We now prove part (C). Note that only the diagonal elements of the matrix $\hat{A}^T(\hat{\psi}(h, t))$ can be negative. For $t \in [0, t_0]$ let

$$c(t) = -\min_{1 \leq i \leq q} \hat{A}_{ii}(\hat{\psi}(h, t)).$$

Fix a $v_0 \in \mathbb{R}_*^q$ and define

$$L(t) = \exp\left(\int_{t_0 - t}^{t_0} c(s)ds\right) \Phi(h, t_0 - t, t_0)v_0.$$

Then

$$\frac{dL(t)}{dt} = \left(c(t_0 - t)I_q + \hat{A}^T(\hat{\psi}(h, t_0 - t))\right) L(t).$$

But the matrix $\left(c(t_0 - t)I_q + \hat{A}^T(\hat{\psi}(h, t_0 - t))\right)$ has all entries positive for any $t \in [0, t_0]$. Since $L(0) = v_0 \in \mathbb{R}_*^q$, for any $t \in [0, t_0]$ we have $dL(t)/dt \in \mathbb{R}_+^q$. Therefore $L(t) \in \mathbb{R}_*^q$ and this proves part (C). $\square$

References

[1] W. Allee. *Animal aggregations : A study in general sociology.* University of Chicago Press, Chicago, USA, 1931.

[2] S. J. Altschuler, S. B. Angenent, Y. Wang, and L. F. Wu. On the spontaneous emergence of cell polarity. *Nature*, 454:886–889, 2008.

[3] K. Ball, T. G. Kurtz, L. Popovic, and G. Rempala. Asymptotic analysis of multiscale approximations to reaction networks. *The Annals of Applied Probability*, 16(4):1925–1961, 2006. MR-2288709

[4] A. Butty, N. Perrinjaquet, A. Petit, M. Jaquenoud, J. Segall, K. Hofmann, C. Zwahlen, and M. Peter. A positive feedback loop stabilizes the guanine-nucleotide exchange factor cdc24 at sites of polarization. *EMBO Journal*, 21:1565–1576, 2002.

[5] Y. Cao, D. T. Gillespie, and L. R. Petzold. The slow-scale stochastic simulation algorithm. *The Journal of Chemical Physics*, 122(1), Jan. 2005.

[6] C. Chicone. *Ordinary differential equations with applications*, volume 34 of *Texts in Applied Mathematics*. Springer-Verlag, New York, 1999. MR-1707333

[7] D. A. Dawson. *École d'Été de Probabilités de Saint-Flour XXI—1991*, volume 1541 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1993. Papers from the school held in Saint-Flour, August 18–September 4, 1991, Edited by P. L. Hennequin. MR-1242574

[8] P. Donnelly and T. G. Kurtz. Genealogical processes for Fleming-Viot models with selection and recombination. *The Annals of Applied Probability*, 9(4):1091–1148, 1999. MR-1728556

[9] P. Donnelly and T. G. Kurtz. Particle representations for measure-valued population models. *The Annals of Probability*, 27(1):166–205, 1999. MR-1681126

[10] D. G. Drubin and W. J. Nelson. Origins of cell polarity. *Cell*, 84:335–344, 1996.

[11] S. N. Ethier and T. G. Kurtz. The infinitely-many-neutral-alleles diffusion model. *Advances in Applied Probability*, 13(3):429–452, 1981. MR-0615945

[12] S. N. Ethier and T. G. Kurtz. *Markov processes : Characterization and Convergence*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1986. MR-0838085

[13] S. N. Ethier and T. G. Kurtz. Fleming-Viot processes in population genetics. *Siam Journal on Control and Optimization*, 31(2):345–386, 1993. MR-1205982

[14] W. J. Ewens. *Mathematical population genetics. I*, volume 27 of *Interdisciplinary Applied Mathematics*. Springer-Verlag, New York, second edition, 2004. MR-2026891

[15] S. Feng. *The Poisson-Dirichlet distribution and related topics*. Probability and its Applications (New York). Springer, Heidelberg, 2010. Models and asymptotic behaviors. MR-2663265

[16] P. C. Fife. *Mathematical aspects of reacting and diffusing systems*, volume 28 of *Lecture Notes in Biomathematics*. Springer-Verlag, Berlin, 1979. MR-0527914

[17] W. Fleming and M. Viot. Some measure-valued Markov processes in population genetics theory. *Indiana University Mathematics Journal*, 28:817–843, 1979. MR-0542340

[18] A. Gierer and H. Meinhardt. A theory of biological pattern formation. *Kybernetik*, 12:30–39, 1972.

[19] A. Gupta. Stochastic model for cell polarity. *The Annals of Applied Probability*, 22(2):827–859, 2012. MR-2953570

[20] J. E. Irazoqui, A. S. Gladfelter, and D. J. Lew. Scaffold-mediated symmetry breaking by cdc42p. *Nature Cell Biology*, 5:1062–1070, 2003.

[21] A. Joffe and M. Métivier. Weak convergence of sequences of semimartingales with applications to multitype branching processes. *Advances in Applied Probability*, 18(1):20–65, 1986. MR-0827331

[22] H.-W. Kang and T. G. Kurtz. Separation of time-scales and model reduction for stochastic reaction networks. *The Annals of Applied Probability (to appear)*, 2012.

[23] G. S. Katzenberger. Solutions of a stochastic differential equation forced onto a manifold by a large drift. *The Annals of Probability*, 19(4):1587–1628, 1991. MR-1127717

[24] H. K. Khalil. *Nonlinear systems*. Macmillan Publishing Company, New York, 1992. MR-1201326

[25] M. Kimura. Solution of a process of random genetic drift with a continuous model. *Proceedings of the National Academy of Sciences*, 41(3):144–150, 1955.

[26] M. Kimura and J. Crow. The number of alleles that can be maintained in a finite population. *Genetics*, 49:725–738, 1964.

[27] J. F. C. Kingman. Random discrete distribution. *Journal of the Royal Statistical Society. Series B.*, 37:1–22, 1975. With a discussion by S. J. Taylor, A. G. Hawkes, A. M. Walker, D. R. Cox, A. F. M. Smith, B. M. Hill, P. J. Burville, T. Leonard and a reply by the author. MR-0368264

[28] A. Lotka. *Elements of Physical Biology*. The Williams and Watkins company, Baltimore, 1925.

[29] P. A. P. Moran. Random processes in genetics. *Mathematical Proceedings of the Cambridge Philosophical Society*, 54(01):60–71, 1958. MR-0127989

[30] R. M. Nisbet and W. S. C. Gurney. *Modeling fluctuating populations*. Wiley, 1982.

[31] K. Oelschläger. On the derivation of reaction-diffusion equations as limit dynamics of systems of moderately interacting stochastic processes. *Probability Theory and Related Fields*, 82(4):565–586, 1989. MR-1002901

[32] E. Seneta. *Non-negative matrices and Markov chains*. Springer Series in Statistics. Springer, New York, 2006. MR-2209438

[33] M. Sohrmann and M. Peter. Polarizing without a c(l)ue. *Trends Cell Biology*, 13:526–533, 2003.

[34] T. Takaku, K. Ogura, H. Kumeta, N. Yoshida, and F. Inagaki. Solution structure of a novel cdc42 binding module of bem1 and its interaction with ste20 and cdc42. *Journal of Biological Chemistry*, 285(25):19346–19353, 2010.

[35] H. R. Thieme. *Mathematics in population biology*. Princeton Series in Theoretical and Computational Biology. Princeton University Press, Princeton, NJ, 2003. MR-1993355

[36] J. M. Varah. A lower bound for the smallest singular value of a matrix. *Linear Algebra and Applications*, 11:3–5, 1975. MR-0371929

[37] P. Verhulst. Notice sur la loi que la population poursuit dans son accroissement. *Correspondance Mathématique et Physique*, 10:113–121, 1838.

[38] V. Volterra. Fluctuations in the abundance of a species considered mathematically. *Nature*, 118:558–560, 1926.

[39] O. Weiner, P. Neilsen, G. Prestwich, M. Kirschner, L. Cantley, and H. Bourne. A ptdinsp(3)- and rho gtpase-mediated positive feedback loop regulates neutrophil polarity. *Nature Cell Biology*, 4(5):509–13, 2002.

[40] S. Wright. Evolution in Mendelian populations. *Genetics*, 16(2):97–159, 1931.