

QUASISTATIONARY DISTRIBUTIONS AND FLEMING–VIOT PROCESSES IN FINITE SPACES

AMINE ASSELAH,* *Université Paris-Est*

PABLO A. FERRARI,** *Universidade de São Paulo and Universidad de Buenos Aires*

PABLO GROISMAN,** *** *Universidad de Buenos Aires*

Abstract

Consider a continuous-time Markov process with transition rates matrix Q in the state space $\Lambda \cup \{0\}$. In the associated Fleming–Viot process N particles evolve independently in Λ with transition rates matrix Q until one of them attempts to jump to state 0. At this moment the particle jumps to one of the positions of the other particles, chosen uniformly at random. When Λ is finite, we show that the empirical distribution of the particles at a fixed time converges as $N \rightarrow \infty$ to the distribution of a single particle at the same time conditioned on not touching $\{0\}$. Furthermore, the empirical profile of the unique invariant measure for the Fleming–Viot process with N particles converges as $N \rightarrow \infty$ to the unique quasistationary distribution of the one-particle motion. A key element of the approach is to show that the two-particle correlations are of order $1/N$.

Keywords: Quasistationary distribution; Fleming–Viot process

2010 Mathematics Subject Classification: Primary 60K35

Secondary 60J25

1. Introduction

Let Λ be a finite or countable state space, whose elements are also called *sites*. Let $Q = (q(x, y), x, y \in \Lambda \cup \{0\})$ be the transition rates matrix of an irreducible continuous-time Markov process on $\Lambda \cup \{0\}$. The transition matrix for the process stopped on $\{0\}$ is $\bar{Q}(x, y) = Q(x, y)$ for $x \in \Lambda$, and $\bar{Q}(0, y) = 0$ for any $y \in \Lambda$.

Consider the process on $\Lambda \cup \{0\}$ generated by Q , with initial law μ , and denote by $T_t\mu$ its law at time t conditioned on not having touched $\{0\}$ up to time t . In other words, for all $x \in \Lambda$,

$$T_t\mu(x) = \frac{\sum_{y \in \Lambda} \mu(y) \exp(t\bar{Q})(y, x)}{1 - \sum_{y \in \Lambda} \mu(y) \exp(t\bar{Q})(y, 0)}.$$

Let \mathcal{M} be the space of probability measures on Λ . Then $\{T_t, t \geq 0\}$ is a semigroup on \mathcal{M} and $T_t\mu$ is the unique solution to the Kolmogorov forward equations: $T_0\mu(x) = \mu(x)$, and, for any $x \in \Lambda$,

$$\frac{d}{dt} T_t\mu(x) = \sum_{y \in \Lambda} q(y, x) T_t\mu(y) + \sum_{y \in \Lambda} q(y, 0) T_t\mu(y) T_t\mu(x). \quad (1)$$

Received 23 February 2010; revision received 6 December 2010.

* Postal address: LAMA, Université Paris-Est, CNRS UMR 8050, 61 Avenue General de Gaulle, 94010 Creteil Cedex, France.

** Postal address: DM-FCEN, Universidad de Buenos Aires, Pabellon 1, Ciudad Universitaria, 1428 Buenos Aires, Argentina.

*** Email address: pgroisma@dm.uba.ar

A quasistationary distribution (QSD) for Q is a probability measure ν on Λ that is invariant under $\{T_t, t \geq 0\}$, that is,

$$T_t \nu = \nu \quad \text{for all } t \geq 0.$$

Our goal is to approximate the conditioned process by a particle system. This is useful from a theoretical point of view (in [4] the QSD is constructed as the limit of the invariant measures of the processes that approximate the conditioned evolution) as well as for simulation purposes, since the conditioned evolution cannot be simulated by the rejection method for large times.

1.1. The associated Fleming–Viot process

For each integer $N > 1$, the Fleming–Viot process with N particles is a continuous-time Markov process $\xi_t \in \Lambda^N, t \geq 0$; let $\xi_t(i)$ denote the position of the i th particle at time t . The generator \mathcal{L}^N acts on functions $f: \Lambda^N \rightarrow \mathbb{R}$ as

$$\mathcal{L}^N f(\xi) = \sum_{i=1}^N \sum_{x \in \Lambda \setminus \{\xi(i)\}} \left[q(\xi(i), x) + q(\xi(i), 0) \frac{\sum_{j \neq i}^N \mathbf{1}_{\{\xi(j)=x\}}}{N-1} \right] (f(\xi^{i,x}) - f(\xi)), \quad (2)$$

where $\xi^{i,x}(i) = x$, and, for $j \neq i, \xi^{i,x}(j) = \xi(j)$. We set $E_\xi^N[f(\xi_t)] = \exp(t\mathcal{L}^N)f(\xi)$. In words, each particle moves independently of the others as a continuous-time Markov process with transition rates matrix Q , but when it attempts to jump to state 0, it jumps to the position of one of the other particles chosen uniformly at random.

Denote by $\eta(\xi, x)$ the number of ξ particles at site x , and by $m(\xi)$ the empirical measure induced by a configuration $\xi \in \bigcup_N \Lambda^N$:

$$\eta(\xi, x) := \sum_{i=1}^N \mathbf{1}_{\{\xi(i)=x\}} \quad \text{and} \quad m(\xi) := \frac{\sum_{x \in \Lambda} \eta(\xi, x) \delta_x}{\sum_{x \in \Lambda} \eta(\xi, x)}. \quad (3)$$

We also use $m_x(\xi)$ to denote $m(\xi)(x)$ and $q(x, x) = -\sum_{y \in \Lambda \cup \{0\} \setminus \{x\}} q(x, y)$. With this notation, the time derivative of $E_\xi^N[m_x(\xi_t)]$ is easily seen to be

$$\frac{dE_\xi^N[m_x(\xi_t)]}{dt} = \sum_{y \in \Lambda} q(y, x) E_\xi^N[m_y(\xi_t)] + \frac{N}{N-1} \sum_{y \in \Lambda} q(y, 0) E_\xi^N[m_y(\xi_t) m_x(\xi_t)]. \quad (4)$$

There is a natural Fleming–Viot process associated to any given Markov process with absorbing states. This appears for the first time in [2] for Brownian motion absorbed at the boundary of a bounded domain. In the associated Fleming–Viot process N Brownian particles evolve independently until one of them reaches the boundary, which plays the role of state 0. This model and generalizations of it were studied in several papers; see, e.g. [1], [2], [5], [6], and [7], which dealt with diffusions in bounded or unbounded domains. These works had to address the serious problem of nonexplosion of the number of hits of the boundary, and this required sophisticated analysis. On the other hand, Ferrari and Marić [4] considered a countable space, and a Doeblin-type condition was imposed on the transition rates matrix Q . Here we continue the analysis of [4] without their strong Doeblin-type condition but with a finite space. The starting point is still the similarity between (1) and (4). It is obvious that we require a control of the correlations between occupation numbers. We mention that a sophisticated approach, formulated in the proof of Theorem 1.4 of [2], was used to study the same object (for N Brownian motions on a bounded domain), without the need for controlling such correlations.

Another approach given in [5], still for N Brownian motions on a bounded domain, was used to study the limit of the empirical measure through the hydrodynamics limit technology. Both approaches in [2] and [5] are sophisticated, whereas here we present a simple, short, and natural way to link quasistationary measures with the law of the empirical measure of the Fleming–Viot process under its invariant measure. This latter approach is adapted to finite-state Markov chains.

Once the Fleming–Viot process is considered, the issues to address are, of course, the well definiteness of the process and its ergodicity. Neither issue is straightforward to resolve in general state spaces. In countable spaces the well definiteness is immediate but not the ergodicity (see [4]). The conditions on the original process that guarantee the ergodicity of the Fleming–Viot process is still an open problem. Once this is solved, we are interested in the approximation of the conditioned process by its associated Fleming–Viot process.

1. Does the Fleming–Viot process approximate the conditioned evolution as $N \rightarrow \infty$ for each $t \geq 0$?
2. Does the empirical profile of the invariant measure of the Fleming–Viot process approximate a QSD? In the case of the existence of more than one QSD, which QSD is chosen by the Fleming–Viot process?

These two questions have been answered affirmatively for Brownian motion (the empirical profile of the invariant measure converges in this case to the first eigenfunction of the Laplacian with homogeneous Dirichlet boundary conditions) and more general diffusions (see [1], [2], and [5]). For countable spaces, under the condition that

$$\sum_{z \in \Lambda} \inf_{x \in \Lambda \setminus \{z\}} q(x, z) > \max_{x \in \Lambda} q(x, 0),$$

Ferrari and Marić [4] proved the existence of a unique stationary measure λ^N for the Fleming–Viot process and the existence of a measure ν on Λ such that

$$\lim_{N \rightarrow \infty} \int_{\Lambda^N} (m_x(\xi) - \nu(x))^2 d\lambda^N(\xi) = 0.$$

Furthermore, ν is the unique QSD of Q .

The main goal of this paper is to prove that the Fleming–Viot process approximates the conditioned evolution as well as the QSD under the only hypothesis that Λ is *finite* (Theorem 2). The result is not included in [4] and requires control of the two-particle correlations uniformly in the initial configuration.

Assume that μ is close to $m(\xi)$, and look at (1) and (4). A natural approach to show that $E_\xi^N[m(\xi_t)]$ is close to $T_t\mu$ is to establish that the occupation numbers of two distinct sites, at time t , become independent when N tends to ∞ (the so-called *propagation of chaos*). For this purpose, Ferrari and Marić [4] estimated the correlation of two ξ -particles, for Λ countable as the unique assumption.

Proposition 1. ([4, Proposition 3.1].) *Let μ be any probability measure on Λ , and let $\mu^{\otimes N}$ be the product probability on Λ^N . Then, there is a constant κ such that, for any $x, y \in \Lambda$ and $t > 0$,*

$$\left| \int E_\xi^N[m_x(\xi_t)m_y(\xi_t)] d\mu^{\otimes N}(\xi) - \int E_\xi^N[m_x(\xi_t)] d\mu^{\otimes N}(\xi) \int E_\xi^N[m_y(\xi_t)] d\mu^{\otimes N}(\xi) \right| \leq \frac{e^{\kappa t}}{N}.$$

In contrast to [4], our argument requires a control of the correlations under the invariant measure of the Fleming–Viot process. To that end, we obtain bounds that hold uniformly on the initial distribution of the particles. This result also holds for countable Λ with no further assumptions.

Proposition 2. *For each $t > 0$ and any $x, y \in \Lambda$,*

$$\sup_{\xi \in \Lambda^N} |E_{\xi}^N[m_x(\xi_t)m_y(\xi_t)] - E_{\xi}^N[m_x(\xi_t)]E_{\xi}^N[m_y(\xi_t)]| \leq \frac{\mathbf{1}_{\{x=y\}}}{N} + \frac{2}{N}(e^{2Ct} - 1). \tag{5}$$

Our proof of Proposition 2 also shows that any finite number of particles evolve independently in the limit $N \rightarrow \infty$. A similar result was obtained by Grigorescu and Kang [6], following their approach in [5].

1.2. Finite state spaces

In the rest of the paper we consider a finite Λ . In this case, for each $N \geq 2$, the Fleming–Viot process is an irreducible pure-jump Markov process on the finite state space Λ^N . Hence, it is ergodic, that is, there exists a unique stationary measure for the process, and starting from any measure, the process converges to the stationary measure. We still denote this measure by λ^N .

When Λ is finite, Darroch and Seneta [3] proved that $T_t\mu$ converges exponentially fast to a probability measure ν , uniformly in the initial measure. The measure ν is the unique QSD of Q .

Theorem 1. ([3].) *Assume that Λ is finite and that the process on Λ with rates $\{q(x, y), x, y \in \Lambda\}$ is irreducible. Then there exists $\theta > 0$ such that*

$$\sup_{\mu \in \mathcal{M}} \|T_t\mu - \nu\| \leq e^{-\theta t}. \tag{6}$$

In Theorem 1 we have used the total variation norm

$$\|\mu - \nu\| = \sum_{x \in \Lambda} |\mu(x) - \nu(x)|.$$

The asymptotic independence of Proposition 2 naturally implies the convergence of the empirical means in the Fleming–Viot process to the conditioned distribution $T_t m(\xi)$, uniformly in ξ . Moreover, since Λ is finite, $T_t m(\xi)$ is close to the unique QSD, uniformly in ξ , as implied by (6). These two facts imply the following, which is our main result.

Theorem 2. *Assume that Λ is finite. Then,*

$$\lim_{N \rightarrow \infty} \int_{\Lambda^N} \|m(\xi) - \nu\| d\lambda^N(\xi) = 0. \tag{7}$$

Remark 1. Note that (7) readily implies (using the fact that $0 \leq m_x \leq 1$ for all $x \in \Lambda$) that, for any subset $U \subset \Lambda$,

$$\lim_{N \rightarrow \infty} \int \prod_{x \in U} m_x d\lambda^N = \prod_{x \in U} \nu(x).$$

The rest of the paper is organized as follows. In Section 2 we construct the process *à la Harris* following [4]. We use this construction to estimate the correlations and prove Proposition 2 in Section 3. Finally, in Section 4 we prove Theorem 2.

2. Graphical construction

We use the Harris graphical method [8] to construct the process. This representation is then used to prove the asymptotic independence property. A realization of the process $(\xi_t, t \geq 0)$ is a deterministic function of a realization of a marked (Poisson) point process. All initial conditions ξ use the same realization of the marked point process. The construction allows us to explore the process backwards in time (in the absence of the useful duality tool in particle systems) and control the two-point correlation function with a branching process at a finite time.

Let $C := \max_{x \in \Lambda} q(x, 0)$ be the (maximum) absorption rate, and let

$$\bar{q} := \sup_{x \in \Lambda} \sum_{y \in \Lambda \setminus \{x\}} q(x, y); \quad p(x, y) := \frac{q(x, y)}{\bar{q}}, \quad y \neq x;$$

$$p(x, x) := 1 - \sum_{y \in \Lambda \setminus \{x\}} p(x, y).$$

To each particle i , we associate two independent marked Poisson processes (ω_i^I, ω_i^V) , which we respectively call the *internal* and *voter* point processes, described as follows.

- The internal process is defined on $\mathbb{R} \times \Lambda^\Lambda$ with intensity measure $\bar{q} dt d\gamma(F)$, where

$$d\gamma(F) = \prod_{x \in \Lambda} p(x, F(x)) \quad \text{for all } F \in \Lambda^\Lambda,$$

that is, γ is the joint distribution of independent random variables with marginal distributions $\{p(x, \cdot), x \in \Lambda\}$, so that to each state x the (random) F assigns a state $y = F(x)$ with probability $p(x, y)$. If (t, F) is an internal marked time associated to particle i and at time $t -$ particle i is at site x , then at time t , particle i jumps to site $F(x)$. The function F associated to this time is called the *mark*. This gives the correct rate $q(x, y) = \bar{q} p(x, y)$ for jumps from x to y .

- The voter point process is defined on $\mathbb{R} \times (\{1, \dots, N\} \setminus \{i\}) \times \{0, 1\}^\Lambda$ with measure intensity $C dt d\beta_i(j) d\gamma'(\zeta)$, where β_i is the uniform probability on $\{1, \dots, N\} \setminus \{i\}$ and γ' is the joint distribution of independent Bernoulli random variables with parameters $q(x, 0)/C, x \in \Lambda$. A voter marked time is (t, j, ζ) , where j corresponds to a reincarnation label and ζ takes into account the position-dependent rate: if the i th particle is at position x at time $t -$, it jumps to the position of particle j at time t only if $\zeta(x) = 1$, yielding the correct rate $q(x, 0)/(N - 1)$.

We call $\omega = ((\omega_i^I, \omega_i^V), i \in \{1, \dots, N\})$ an independent and identically distributed sequence of stationary marked point processes associated with labeled particles. Finally, for any subset of labels $a \subset \{1, \dots, N\}$, we denote by ω_a the processes associated with labels a . For any real numbers $s < t$, we denote by $\omega_a[s, t)$ and $\omega_a[s, t]$ the projections of the marked times in the time period $[s, t)$ and $[s, t]$, respectively.

We construct $\{\xi_t, t \geq 0\}$ in such a way that ξ_t is a function of the initial configuration ξ_0 and the time marks $\omega[0, t], t \geq 0$. Fix an initial configuration $\xi_0 \in \Lambda^N$, and $t > 0$. There is, almost surely, a finite number of time marks within $[0, t]$, say K ; let $\{b_k, 0 \leq k \leq K\}$ be the ordered time realizations with $b_0 = 0$. We build ξ_t inductively as follows.

- At time $b_0 = 0$, the configuration is ξ_0 .

- Assume that ξ_{b_k} is known. For $t \in [b_k, b_{k+1})$, set $\xi_t = \xi_{b_k}$. We now describe $\xi_{b_{k+1}}$.
 - If b_{k+1} corresponds to an internal time of particle i and mark F , we move particle i to $F(x)$, where $x = \xi_{b_k}(i)$. This move occurs with rate $\bar{q}q(x, F(x))/\bar{q} = q(x, F(x))$.
 - If b_{k+1} corresponds to a voter time of particle i and mark (j, ζ) , we move particle i to the position of particle j if $\zeta(x) = 1$, where $x = \xi_{b_k}(i)$. This move occurs with rate

$$C \frac{q(x, 0)}{C} \frac{1}{N-1} = \frac{q(x, 0)}{N-1}.$$

It is easy to check that $\{\xi_t, t \geq 0\}$, as constructed above, has generator given by (2); see [8].

By translation invariance of the law of ω , if we use the marks $\omega[-t, 0]$ instead of the marks $\omega[0, t]$, the configuration so obtained has the same law of ξ_t as constructed above. We abuse notation and call ξ_t the configuration constructed with the marks $\omega[-t, 0]$. For each particle label i , we build simultaneously a set of labels $\psi^i(t)$ of particles which could potentially influence $\xi_t(i)$ (that is, the set of all other particles that may have interacted with particle i up to time t). This set is constructed backwards in time (from time 0 to time $-t$), also as a function of $\omega[-t, 0]$. First, the process $t \mapsto \psi^i(t)$ may only change at the time realizations of the voter process ω^V , and it changes as follows. Let $-v$ be the largest time realization of $\omega_i^V[-t, 0]$, and let (j, ζ) be its associated mark. Then, for $0 \leq s < v$, we set $\psi^i(s) = \{i\}$, and $\psi^i(v) = \{i, j\}$ (regardless of the values of ζ). For $s < t$, assume that $\psi^i(s)$ is built, let $-v$ be the largest time realization of $\omega_{\psi^i(s)}^V[-t, -s)$, and let (j, ζ) be its associated mark. Then, for all $u \in [s, v)$,

$$\psi^i(u) = \psi^i(s) \quad \text{and} \quad \psi^i(v) = \psi^i(s) \cup \{j\}.$$

Note that, for any $t > 0$, $\psi^i(t)$ is $\sigma(\omega^V[-t, 0])$ -measurable and that, for any subset of labels $a \subset \{1, \dots, N\}$ containing i , we have

$$\{\psi^i(t) = a\} \in \sigma(\omega_a^V[-t, 0]) \quad \text{and} \quad \{\psi^i(t) = a, \xi_t(i) = x\} \in \sigma(\omega_a[-t, 0]). \tag{8}$$

The next lemma says that the sets of labels associated to two different particles intersect with probability of order $1/N$ times an exponential factor in t . Since t is fixed and N goes to ∞ , the bound is sufficient to establish the $1/N$ decay of two-particle correlation at any fixed time t .

Lemma 1. *For i, j distinct labels and $t > 0$,*

$$P(\psi^i(t) \cap \psi^j(t) \neq \emptyset) \leq \frac{1}{N-1} (e^{2Ct} - 1). \tag{9}$$

Proof. First, we show that the rate of growth of $\psi^i(t)$ is at most exponential. It is clear from the construction of ψ^i that its rate of growth at time t is at most $C\psi^i(t)$ and that it grows by adding one label (from $\{1, \dots, N\} \setminus \psi^i(t)$) uniformly at random. Thus,

$$E[|\psi^i(t)| \mid \sigma(\omega[-t, 0])] \leq 1 + C \int_0^t |\psi^i(s)| ds \implies E[|\psi^i(t)|] \leq \exp(Ct). \tag{10}$$

Second, we show that, for two distinct labels i, j ,

$$P(\psi^i(t) \cap \psi^j(t) \neq \emptyset) \leq \frac{2C}{N-1} \int_0^t E[\psi^i(s)]E[\psi^j(s)] ds. \tag{11}$$

Note that

$$P(\psi^i(t) \cap \psi^j(t) \neq \emptyset) = \int_0^t E \left[\frac{dP(\psi^i(s) \cap \psi^j(s) \neq \emptyset \mid \sigma(\omega[-s, 0]))}{ds} \right] ds$$

and

$$\begin{aligned} & E \left[\frac{dP(\psi^i(s) \cap \psi^j(s) \neq \emptyset \mid \sigma(\omega[-s, 0]))}{ds} \right] \\ &= \frac{2C}{N-1} E[\mathbf{1}_{\{\psi^i(s) \cap \psi^j(s) = \emptyset\}} \mid \psi^i(s) \parallel \psi^j(s)] \\ &= \frac{2C}{N-1} \sum_{a \cap b = \emptyset} |a| |b| P(\psi^i(s) = a, \psi^j(s) = b) \\ &= \frac{2C}{N-1} \sum_{a \cap b = \emptyset} |a| |b| P(\psi^i(s) = a) P(\psi^j(s) = b) \\ &\leq \frac{2C}{N-1} E[|\psi^i(s)|] E[|\psi^j(s)|]. \end{aligned} \tag{12}$$

In (12) we used the fact that, for two nonoverlapping subsets of labels a and b , $\{\psi^i(t) = a\}$ and $\{\psi^j(t) = b\}$ are independent by (8).

This concludes the proof of (11). Now (9) follows from (10) and (11).

3. Proof of Proposition 2

We need to show that, for any $x, y \in \Lambda$, any time $t \geq 0$, and initial configuration ξ ,

$$|E[\eta(\xi_t)(x)\eta(\xi_t)(y)] - E[\eta(\xi_t)(x)]E[\eta(\xi_t)(y)]| \leq 2Ne^{2Ct}. \tag{13}$$

Here and throughout this section, we use E and P to denote E_ξ^N and P_ξ^N , respectively. Using (3), the difference in the expectations on the left-hand side of (13) is

$$\sum_{i \leq N} \sum_{j \leq N} [P(\xi_t(i) = x, \xi_t(j) = y) - P(\xi_t(i) = x)P(\xi_t(j) = y)]. \tag{14}$$

For a subset a , $\{\xi_t(i) = x, \psi^i(t) = a\}$ is $\sigma(\omega_a[-t, 0])$ -measurable, by (8). Thus, for two nonoverlapping subsets of labels a and b ,

$$\begin{aligned} & P(\psi^i(t) = a, \psi^j(t) = b, \xi_t(i) = x, \xi_t(j) = y) \\ &= P(\psi^i(t) = a, \xi_t(i) = x)P(\psi^j(t) = b, \xi_t(j) = y). \end{aligned}$$

Compute a generic term on the right-hand side of (14) with $i \neq j$:

$$\begin{aligned} P(\{\xi_t(i) = x, \xi_t(j) = y\}) &= P(\psi^i(t) \cap \psi^j(t) \neq \emptyset, \xi_t(i) = x, \xi_t(j) = y) \\ &\quad + \sum_{a \cap b = \emptyset} P(\psi^i(t) = a, \psi^j(t) = b, \xi_t(i) = x, \xi_t(j) = y) \\ &= P(\psi^i(t) \cap \psi^j(t) \neq \emptyset, \xi_t(i) = x, \xi_t(j) = y) \\ &\quad + \sum_{a \cap b = \emptyset} P(\psi^i(t) = a, \xi_t(i) = x)P(\psi^j(t) = b, \xi_t(j) = y). \end{aligned} \tag{15}$$

To compute $P(\xi_t(i) = x)P(\xi_t(j) = y)$, we can think of two independent marked point processes driving the evolution (we use a tilde to denote the independent copy), and we have a decomposition similar to (15) for $i \neq j$:

$$\begin{aligned}
 P(\xi_t(i) = x)P(\xi_t(j) = y) &= P(\psi^i(t) \cap \tilde{\psi}^j(t) \neq \emptyset, \xi_t(i) = x, \tilde{\xi}_t(j) = y) \\
 &\quad + \sum_{a \cap b = \emptyset} P(\psi^i(t) = a, \tilde{\psi}^j(t) = b, \xi_t(i) = x, \tilde{\xi}_t(j) = y) \\
 &= P(\psi^i(t) \cap \tilde{\psi}^j(t) \neq \emptyset, \xi_t(i) = x, \tilde{\xi}_t(j) = y) \\
 &\quad + \sum_{a \cap b = \emptyset} P(\psi^i(t) = a, \xi_t(i) = x)P(\psi^j(t) = b, \xi_t(j) = y).
 \end{aligned}
 \tag{16}$$

Subtracting (16) from (15) we obtain, for $i \neq j$,

$$\begin{aligned}
 &|P(\xi_t(i) = x, \xi_t(j) = y) - P(\{\xi_t(i) = x\}P(\xi_t(j) = y))| \\
 &\leq P(\psi^i(t) \cap \psi^j(t) \neq \emptyset) + P(\psi^i(t) \cap \tilde{\psi}^j(t) \neq \emptyset) \\
 &\leq \frac{2}{N-1}(e^{2Ct} - 1),
 \end{aligned}$$

by Lemma 1 (we have used the fact that the lemma also holds for $\psi^i(t) \cap \tilde{\psi}^j(t)$). Thus, by summing over i and $j \in \{1, \dots, N\}$, and noting that there are N diagonal terms which bring a factor N when $i = j$, we obtain the desired bound.

4. Proof of Theorem 2

Note that (7) follows from (6), combined with the following property: for any positive time t ,

$$\lim_{N \rightarrow \infty} \sup_{\xi \in \Lambda^N} E_\xi^N [\|m(\xi_t) - T_t m(\xi)\|] = 0.$$

Indeed,

$$\begin{aligned}
 &\int_{\Lambda^N} \|m(\xi) - \nu\| \, d\lambda^N(\xi) \\
 &= \int_{\Lambda^N} E_\xi^N [\|m(\xi_t) - \nu\|] \, d\lambda^N(\xi) \\
 &\leq \int_{\Lambda^N} E_\xi^N [\|m(\xi_t) - T_t m(\xi)\|] \, d\lambda^N(\xi) + \int_{\Lambda^N} \|T_t m(\xi) - \nu\| \, d\lambda^N(\xi) \\
 &\leq \sup_{\xi \in \Lambda^N} E_\xi^N [\|m(\xi_t) - T_t m(\xi)\|] + \sup_{\mu \in \mathcal{M}} \|T_t \mu - \nu\|.
 \end{aligned}$$

Thus, we first estimate $E_\xi^N [\|m(\xi_t) - T_t m(\xi)\|]$. It is more convenient to work with the l^2 -norm, rather than the total variation norm. For a function $\varphi: \Lambda \rightarrow \mathbb{R}$, we denote its l^2 -norm as

$$\|\varphi\|_2 = \left(\sum_{x \in \Lambda} \varphi^2(x) \right)^{1/2}.$$

By the Cauchy–Schwarz inequality, note that if μ and ν are probabilities on Λ ,

$$\|\mu - \nu\|_2 \leq \|\mu - \nu\| \leq \sqrt{|\Lambda|} \|\mu - \nu\|_2.$$

To estimate $E_\xi^N [\|m(\xi_t) - T_t m(\xi)\|_2]$, note that

$$E_\xi^N [\|m(\xi_t) - T_t m(\xi)\|_2] \leq E_\xi^N [\|m(\xi_t) - E_\xi^N [m(\xi_t)]\|_2] + \|E_\xi^N [m(\xi_t)] - T_t m(\xi)\|_2. \tag{17}$$

Taking $y = x$ in (5) we obtain

$$E[(m_x(\xi_t) - E[m_x(\xi_t)])^2] \leq \frac{2e^{2Ct}}{N}. \tag{18}$$

By (18) and Jensen’s inequality, we have

$$[E_\xi^N \|m(\xi_t) - E_\xi^N [m(\xi_t)]\|_2]^2 \leq E_\xi^N [\|m(\xi_t) - E_\xi^N [m(\xi_t)]\|_2^2] \leq \frac{2|\Lambda|e^{2Ct}}{N}.$$

The second term in (17) is dealt with in the following lemma.

Lemma 2. For any $T > 0$,

$$\lim_{N \rightarrow \infty} \max_{0 \leq t \leq T} \max_{\xi \in \Lambda^N} \|E_\xi^N [m(\xi_t)] - T_t m(\xi)\|_2 = 0.$$

Proof. We introduce some simplifying notation:

$$u_x(t) = E_\xi^N [m_x(\xi_t)] \quad \text{and} \quad v_x(t) = T_t m(\xi)(x).$$

We show that there is a constant B such that, for any $t > 0$,

$$\frac{d}{dt} \|u(t) - v(t)\|_2^2 \leq B \|u(t) - v(t)\|_2^2 + \frac{4e^{2Ct} \sum_y q(y, 0)}{N}. \tag{19}$$

Since $\|u(0) - v(0)\|_2 = 0$, the result follows at once by means of Gronwall’s inequality.

We fix $t > 0$, and we often omit to display the time dependence. From (4),

$$\frac{du_x}{dt} = \sum_{y \in \Lambda} q(y, x) u_y + \sum_{y \in \Lambda} q(y, 0) u_x u_y + R_x(\xi, t), \tag{20}$$

where

$$R_x(\xi, t) = \sum_{y \in \Lambda} q(y, 0) \left[\frac{N}{N-1} E_\xi^N [m_y(\xi_t) m_x(\xi_t)] - E_\xi^N [m_y(\xi_t)] E_\xi^N [m_x(\xi_t)] \right].$$

Proposition 2 implies that

$$\sup_{x \in \Lambda} \sup_{\xi} |R_x(\xi, t)| \leq \frac{2e^{2Ct} \sum_y q(y, 0)}{N}.$$

On the other hand, from (1),

$$\frac{dv_x}{dt} = \sum_{y \in \Lambda} q(y, x) v_y + \sum_{y \in \Lambda} q(y, 0) v_x v_y. \tag{21}$$

Subtracting (21) from (20), we obtain

$$\frac{d(u_x - v_x)}{dt} = \sum_{y \in \Lambda} q(y, x)(u_y - v_y) + \sum_{y \in \Lambda} q(y, 0)(u_x u_y - v_x v_y) + R_x(\xi, t).$$

Now,

$$\begin{aligned} \frac{d}{dt} \frac{1}{2} \|u(t) - v(t)\|_2^2 &= \sum_x \frac{d(u_x - v_x)}{dt} (u_x - v_x) \\ &= \sum_{x \in \Lambda} \sum_{y \in \Lambda} q(x, y)(u_y - v_y)(u_x - v_x) \\ &\quad + \sum_{x \in \Lambda} \sum_{y \in \Lambda} q(y, 0)(u_x u_y - v_x v_y)(u_x - v_x) \\ &\quad + \sum_x R_x(u_x - v_x). \end{aligned} \tag{22}$$

We deal with each term on the right-hand side of (22) separately. First,

$$\left| \sum_{x \in \Lambda} \sum_{y \in \Lambda} q(x, y)(u_y - v_y)(u_x - v_x) \right| \leq \left(\sum_{x, y \in \Lambda} q^2(x, y) \right)^{1/2} \|u - v\|_2^2. \tag{23}$$

To deal with the second term, observe that $u_x u_y - v_x v_y = v_x(u_y - v_y) + u_y(u_x - v_x)$ and, hence, the second term in (22) equals

$$\sum_{x \in \Lambda} \sum_{y \in \Lambda} q(y, 0)v_x(u_y - v_y)(u_x - v_x) + \sum_{x \in \Lambda} \sum_{y \in \Lambda} q(y, 0)u_y(u_x - v_x)(u_x - v_x). \tag{24}$$

Since v is bounded, the first term in (24) can be treated as (23) with $q(x, y)$ replaced by $\tilde{q}(x, y) := q(y, 0)v_x$. The second term in (24) equals

$$\|u - v\|_2^2 \sum_y q(y, 0)u_y \leq \|u - v\|_2^2 \sup_y q(y, 0).$$

Finally, taking into account the fact that $\sum_x v_x = \sum_x u_x = 1$, the last term of (22) can be bounded by

$$2 \sup_x |R_x| \leq \frac{4e^{2Ct} \sum_y q(y, 0)}{N}.$$

Collecting all these computations, we obtain (19).

Acknowledgements

This work was initiated during the semester *Interacting Particle Systems* at IHP, Paris. We would like to thank the staff and the organizers for a wonderful work atmosphere. AA acknowledges the support of the French Ministry of Education through the ANR BLAN07-2184264 grant. PAF was partially supported by FAPESP. PG and PAF were partially supported by the Universidad de Buenos Aires under grant X447, by ANPCYT PICT 2008-315, and by CONICET PIP 0613.

References

- [1] BIENIEK, M., BURDZY, K. AND FINCH, S. (2009). Non-extinction of a Fleming–Viot particle model. Preprint. Available at <http://arxiv.org/abs/0905.1999v1>.
- [2] BURDZY, K., HOŁYST, R. AND MARCH, P. (2000). A Fleming–Viot particle representation of the Dirichlet Laplacian. *Commun. Math. Phys.* **214**, 679–703.
- [3] DARROCH, J. N. AND SENETA, E. (1967). On quasi-stationary distributions in absorbing continuous-time finite Markov chains. *J. Appl. Prob.* **4**, 192–196.
- [4] FERRARI, P. A. AND MARIĆ, N. (2007). Quasi stationary distributions and Fleming–Viot processes in countable spaces. *Electron. J. Prob.* **12**, 684–702.
- [5] GRIGORESCU, I. AND KANG, M. (2004). Hydrodynamic limit for a Fleming–Viot type system. *Stoch. Process. Appl.* **110**, 111–143.
- [6] GRIGORESCU, I. AND KANG, M. (2006). Tagged particle limit for a Fleming–Viot type system. *Electron. J. Prob.* **11**, 311–331.
- [7] GRIGORESCU, I. AND KANG, M. (2011). Immortal particle for a catalytic branching process *Prob. Theory Relat. Fields* 29pp.
- [8] HARRIS, T. E. (1978). Additive set-valued Markov processes and graphical methods. *Ann. Prob.* **6**, 355–378.