Resenhas IME-USP 1993, Vol. 1, No. 2/3, 173 - 181.

Recent Results on Finite and Infinite Systems of Interacting Diffusions[1]

J.T.Cox[2]

Syracuse University

Abstract: We review some recent results concerning the long time behavior of finite and infinite systems of interacting diffusions. After presenting the basic ergodic theory of the infinite systems, we observe that the long time behavior of the finite systems differs drastically from that of the infinite system, at least from a naive point of view. A theorem comparing the finite and infinite systems on different time scales is presented which gives a more complete picture of what is going on.

Key words: Interacting particle systems, finite systems approximation, interacting diffusions, Wright-Fisher process.

Introduction

The purpose of this note is to review recent work from [CG3], [S4], and [CGS] on the long time behavior of finite and infinite systems of interacting diffusions. These are Markov processes $x(t) = \{x_i(t), i \in \Lambda\} \in I^\Lambda$, where $I \subset \mathbb{R}$ is an interval, and $\Lambda \subset \mathbb{Z}^d$ (the d-dimensional integer lattice). They are defined via stochastic differential equations:

$$
(1) \qquad \begin{aligned} &dx_i(t) = \Big[\sum_{j\in\Lambda} a(i,j)x_j(t) - x_i(t)\Big]dt + \sqrt{g(x_i(t))}\,dw_i(t), \quad i \in \Lambda\,, \\ &x(0) \in I^\Lambda\,. \end{aligned}
$$

The ingredients in the above system are:

- A matrix $a(i,j)$ on $\Lambda \times \Lambda$ such that

$$
(2a) \qquad a(i,j) \geq 0, \qquad \sum_{j\in\Lambda} a(i,j) = 1 \quad \forall i \in \Lambda.
$$

- A function $g : I \to [0,\infty)$ such that

(2b) g is locally Lipschitz and vanishes at finite endpoints of I.

- A collection $\{w_i(t), i \in \Lambda\}$ of independent Brownian motions on $\mathbb{R}$.

We will call $x(t)$ an *infinite system* if the index set Λ is infinite, and a *finite system* otherwise. We are particularly interested in comparison theorems relating the long time behavior of infinite systems and large finite systems.

[1]Presented at V CLAPEM, São Paulo, Brasil, June 1993.

[2]Partly supported by the U.S. Army Research Office through the Mathematical Sciences Institute of Cornell University, by the National Science Foundation, and by the National Security Agency.

The first term in (1) describes how the coordinates $x_i(t)$ interact, while the second term is a kind of self fluctuation term. The interaction term is linear in the coordinates, and in the absence of the fluctuation term (set $g \equiv 0$), it is easy to see that (1) has solution

$$x_i(t) = \sum_{j \in \Lambda} a_t(i,j)x_j(0),$$

where $a_t(i,j) = e^{-t}\sum_{n=0}^{\infty} t^n\, a^{(n)}(i,j)/n!$, $t \geq 0$. Thus the interaction term causes *averaging* of the coordinates, and in fact drives the coordinates to total *consensus*. On the other hand, if the interaction term is not present (set $a(i,j) = 1$ if $i = j$, and 0 otherwise), then the coordinates $x_i(t)$ are simply *independent* martingales. In this case the coordinates just fluctuate independently of one another, there is no tendency to agreement. When both terms are present there is a tension between these two opposing forces.

Some well known examples of interacting diffusions are:

Example 1. $I = [0,1]$, $g(\theta) = c\theta(1-\theta)$, c a constant. This is the Wright-Fisher stepping stone model of mathematical genetics, and is the subject of a beautiful series of papers by T. Shiga and his co-workers (see [S1–S3], [NS],[SU]). The model is mathematically tractable because of the existence of a nice "dual" process.

Example 2. $I = [0,\infty)$, $g(\theta) = c\theta$, c a constant. This is a branching diffusion model, or "superprocess" over a Markov chain on Λ with transition matrix $a(i,j)$ (see [D]). The model is tractable because branching techniques can be brought to bear.

For simplicity, and because results are most complete in this case, we will assume from now on that

$$(3) \qquad I = [0,1], \qquad g(0) = g(1) = 0, \qquad g > 0 \text{ on } (0,1).$$

Remark. For unbounded unbounded I, growth conditions on $g(\theta)$ must be imposed (see [S4] and [CGS]).

The Infinite Systems $x(t)$

Let $\Lambda = \mathbb{Z}^d$, and assume in addition to (2a) that $a(i,j)$ is *irreducible*, and that

$$(4) \qquad a(i,j) = a(0,j-i) \quad \forall i,j \in \mathbb{Z}^d.$$

It follows from results of [SS] that for every $x(0) \in I^{\mathbb{Z}^d}$ (endowed with the product topology) there exists a unique strong solution $x(t)$ of (1) such that $x(t)$ is a continuous, strong Markov process, with Feller semigroup $S(t)$ and generator $\mathfrak{G}$ acting on C^2 functions which depend on finitely many coordinates according to

$$\mathfrak{G}f(x) = \frac{1}{2}\sum_{i \in \mathbb{Z}^d} g(x)\frac{\partial^2 f}{\partial x_i^2} + \sum_{i \in \mathbb{Z}^d}\Big[\sum_{j \in Zd} a(i,j) - \delta(i,j)\Big]\frac{\partial f}{\partial x_i}.$$

It turns out that the limiting behavior of $x(t)$ as $t \to \infty$ is very strongly influenced by the symmetrized matrix

$$\widehat{a}(i,j) = \frac{a(i,j) + a(j,i)}{2}.$$

In order to state precise results some additional notation is needed. Let $\mathcal{T}$ denote the collection of probability measures on $I^{\mathbb{Z}^d}$ which are translation invariant, and let $\mathcal{I}$ denote the set of probability measures on $I^{\mathbb{Z}^d}$ which are invariant for $x(t)$, i.e. , all μ such that $\mu S(t) = \mu$ for all $t \geq 0$. We use $\mathcal{L}$ for law, δ_θ for the unit point mass at the element $x_i \equiv \theta$, and Γ_{ext} for the set of extreme points of a convex set Γ. $\Rightarrow$ will denote weak convergence. In particular, if $\mu, \mu_1, \mu_2, \ldots$ are probability measures on $I^{\mathbb{Z}^d}$, then $\mu_n \Rightarrow \mu$ as $n \to \infty$ means that for every bounded continuous function f on $I^{\mathbb{Z}^d}$ which depends on only *finitely* many coordinates, $E^{\mu_n} f(x) \to E^{\mu} f(x)$.

If the matrix $\widehat{a}(i,j)$ is *recurrent*, then the interaction or averaging effect in (1) wins out, and the coordinates $x_i(t)$ tend to consensus in the limit $t \to \infty$. On the other hand, if $\widehat{a}(i,j)$ is *transient*, then the fluctuation term in (1) is dominant, and differences in the coordinates persist. More precisely, we have

Theorem 1.

(a) Assume $\widehat{a}(i,j)$ is recurrent. Then for all initial states $x(0)$, $x(t)$ clusters as $t \to \infty$. That is,

$$x_i(t) - x_j(t) \Rightarrow 0 \text{ as } t \to \infty \quad \forall i,j \in \mathbb{Z}^d. \tag{5}$$

Furthermore, if $\mathcal{L}(x(0)) \in \mathcal{T}$, and $Ex_0(0) = \theta$, then

$$\mathcal{L}(x(t)) \Rightarrow (1-\theta)\delta_0 + \theta\delta_1 \text{ as } t \to \infty. \tag{6}$$

(b) Assume $\widehat{a}(i,j)$ is transient. Then for all $\theta \in I$, with $\mathcal{L}(x(0)) = \delta_\theta$, the weak limit $\nu_\theta = \lim_{t\to\infty} \mathcal{L}(x(t))$ exists. More generally, if $\mathcal{L}(x(0)) \in \mathcal{T}$ is shift-ergodic, with $Ex_0(0) = \theta$, then

$$\mathcal{L}(x(t)) \Rightarrow \nu_\theta \text{ as } t \to \infty. \tag{7}$$

The measures ν_θ are translation invariant, shift-ergodic, and

$$I_{ext} = \{\nu_\theta, \theta \in I\}. \tag{8}$$

Theorem 1 was first proved in [S1–S2] for the Wright-Fisher stepping stone case $g(\theta) = c\theta(1-\theta)$ using *duality* techniques. Part (a) of Theorem 1 was proved for *general* $g(\theta)$ in [NS] (under some mild restrictions) and in [CG3]. Part (b) was obtained for general $g(\theta)$ in [CG3], except for the weaker version of (8), $(\mathcal{I} \cap \mathcal{T})_{ext} = \{\nu_\theta, \theta \in I\}$. The full strength of (8) has only recently been obtained by Shiga (see [S5]). The duality method used to attack the Wright-Fisher stepping stone case does not handle general the case of general $g(\theta)$; however, a beautiful coupling method invented by Liggett and Spitzer (see [LS]) can be used (see [CG3], [S4]).

The Finite Systems $x^N(t)$

For $N = 1, 2, \ldots$, let $\Lambda_N = (-N, N]^d \cap \mathbb{Z}^d$ be viewed as a *torus*, and let $a^N(i,j) = \sum_{k \in \mathbb{Z}^d} a(i, j+2Nk)$, $i, j \in \Lambda_N$. The finite systems $x^N(t) = \{x_i^N(t), i \in \Lambda_N\}$ are defined via the system

$$
(9) \qquad \begin{aligned} dx_i^N(t) &= \Big[\sum_{j \in \Lambda_N} a^N(i,j) x_j^N(t) - x_i^N(t) \Big] dt + \sqrt{g(x_i^N(t))}\, dw_i(t), \quad i \in \Lambda_N, \\ x^N(0) &\in I^{\Lambda_N}. \end{aligned}
$$

Observe that each $x^N(t)$ is really a finite-dimensional diffusion. Furthermore, the $x^N(\cdot)$ are *finite versions* of $x(\cdot)$ in the sense that for *fixed* t,

$$
(10) \qquad \mathcal{L}(x^N(t)) \Rightarrow \mathcal{L}(x(t)) \text{ as } N \to \infty.
$$

Remark. Let us make clear the meaning of the convergence in (10). For $x^N \in \Lambda_N$ let $\pi_N : I^{\Lambda_N} \to I^{\mathbb{Z}^d}$ be the periodic extension operator $(\pi_N x^N)_j = x_i^N$ where $i \in \Lambda_N$, $i = j \mod (2N)$. Also, let π_N denote the induced operator mapping probability measures on I^{Λ_N} to probability measures on $I^{\mathbb{Z}^d}$. If μ^N is a probability measure on I^{Λ_N}, $N = 1, 2, \ldots$, and μ is a probability measure on $I^{\mathbb{Z}^d}$, then $\mu^N \Rightarrow \mu$ as $N \to \infty$ means $\pi_N \mu^N \Rightarrow \mu$ as $N \to \infty$. That is, the convergence in (10) refers to convergence on finite windows of $\mathbb{Z}^d$.

The long time behavior of the finite systems, at least from a naive point of view, is rather simple. If, for instance, $x_i^N(\theta) \equiv \theta$ say, then for fixed N,

$$
\sum_{i \in \Lambda_N} x_i^N(t)
$$

is a nonnegative martingale, and must converge a.s. as $t \to \infty$. ¿From this fact it is easy to see that for fixed N,

$$
(11) \qquad \mathcal{L}(x^N(t)) \Rightarrow (1-\theta)\delta_{\mathbf{0}} + \theta\delta_{\mathbf{1}} \text{ as } t \to \infty,
$$

no matter whether $\widehat{a}(i,j)$ is transient or recurrent. Compare this with Theorem 1, especially (7). This brings up the natural

Question. What relationship exists between the ν_θ of Theorem 1(b) and the large finite systems $x^N(t)$?

To answer this question we must allow N and t to tend to infinity together. Presumably, if N is large and t is large *but not too large*, then one expects $\mathcal{L}(x^N(t)) \approx \mathcal{L}(x(t)) \approx \nu_\theta$. But how large is "not too large?"

The Finite Systems Scheme

We will adopt the finite systems scheme of [CG1], and define the following objects.

(i) The time scale $\beta_N = (2N)^d$.

(ii) The empirical densities

$$\Theta^N(t) = |\Lambda_N|^{-1} \sum_{i \in \Lambda_N} x_i^N(t) \,.$$

(iii) The rescaled process of empirical densities

$$Z_N(t) = \Theta^N(t\beta_N) \,.$$

(iv) The diffusion $Z(t)$ on I, defined for the case $\widehat{a}(i,j)$ transient, by

$$dZ(t) = \sqrt{g^*(Z(s))}\, dw(s) \,, \qquad Z(0) = \rho \,.$$

where $w(t)$ is a Brownian motion on $I\!R$ and g^* is the function

$$g^*(\theta) = E^{\nu_\theta} g(x_0).$$

It can be shown that the function $g^*(\theta)$ is Lipschitz on I, and hence $Z(t)$ is well defined. The probability transition function of $Z(t)$ will be denoted $Q_t(\rho, d\theta)$.

The Comparison Theorem

In view of (6) and (11), it is not surprising that if $\widehat{a}(i,j)$ is recurrent, then the finite systems cluster no matter how N and t tend to infinity. So we will not discuss this case further, except to point out that the *rate* of clustering is an interesting and subtle issue, and is partly addressed in [CGS]. The following result, taken from [CGS], is our answer to the basic question concerning the comparison of large finite and infinite systems.

Theorem 2. Assume $\widehat{a}(i,j)$ is transient. Suppose that for $N = 1, 2, \ldots$, $\mathcal{L}(x^N(0))$ is homogeneous, and $\Theta^N(0) \Rightarrow \theta$ as $N \to \infty$. If $t_N \uparrow \infty$ and $t_N/\beta_N \to s \in [0, \infty]$, then

$$\Theta^N(t_N) \Rightarrow Z(s), \qquad Z(0) = \theta \,, \tag{12}$$

and

$$\mathcal{L}(x^N(t_N)) \Rightarrow \int_I Q_s(\theta, d\rho)\nu_\rho. \tag{13}$$

The convergence in (12) shows that the density process $\Theta^N(\cdot)$ varies smoothly on the time scale β_N. In fact, $\Theta^N(\cdot\, \beta_N) \Rightarrow Z(\cdot)$ on path space. The following special cases help interpret the convergence in (13).

$s = 0$: This is the case $t_N \ll \beta_N$, and the right-hand side of (13) reduces to ν_θ. Thus, the finite systems $x^N(t_N)$ look (locally) like the infinite system $x(t_N)$, namely both are described by ν_θ. This extends (10).

$s = \infty$: This is the case $t_N \gg \beta_N$, and the right-hand side of (13) reduces to $(1-\theta)\delta_0 + \theta\delta_1$ Here we see that on this time scale the finite systems *know* they are finite systems, and cluster. This extends (11).

$s \in (0, \infty)$: Here $t_N \approx s\beta_N$, and in this intermediate phase, very roughly speaking, the law of the finite system $x^N(t_N)$ is ν_ρ, where $\rho = \Theta^N(t_N)$. Here we have the picture that the global density $\Theta^N(\cdot)$ varies on the time scale β_N, and the law of the process $x^N(\cdot)$ is slaved to $\Theta^N(\cdot)$. As $\Theta^N(\cdot)$ *diffuses* through I, $\mathcal{L}(x^N(\cdot))$ diffuses through the invariant measures ν_ρ.

Results of this type were proved for other interacting systems in [CG1] and [DG1]. For instance, a version of Theorem 2 holds for the one-dimensional supercritical contact process. In this case we set $\Theta^N(t) = \sum_{i\in\Lambda_N} 1\{x_i^N(t) = 1\}$, $I = \{0,1\}$, $\nu_0 = \delta_0$, $\nu_1 =$ the upper invariant measure, $\beta_N \approx e^{\gamma N}$ for some constant γ, and $Z(t) =$ the Markov chain on I which jumps at rate one from 1 to 0, and is absorbed at 0. See Theorem 3 of [CG1] for details.

The processes in [CGS1] and [DG1] all enjoy special properties which make them mathematically tractable: duality, branching independence, or mean-field independence. Theorem 2 shows that the finite systems scheme holds more generally, and is *not* just an artifact of special properties like duality and independence.

The mapping $g \to g^*$

Another result from [CGS] gives information about the mapping $g \to g^*$. There is only one case of Theorem 2 in which g^* can be exactly calculated, and that is the Wright-Fisher stepping stone case. Letting $g_o(\theta) = \theta(1-\theta)$, explicit calculations can be made to show that

$$(cg_o)^* = \frac{1}{1 + c\widehat{A}(0,0)}\,(cg_o)\,,$$

where $\widehat{A}(i,jj) = \int_0^\infty \widehat{a}_{2t}(i,j)\,dt$. Thus, up to a scaling factor, g_o is a *fixed point* of the mapping $g \to g^*$. Surprisingly, it can be shown that there is no other such fixed point, highlighting the fundamental importance of this example. For the following result, and more, see [BCGH] and [CGS].

Theorem 3. Assume $\widehat{a}(i,j)$ is transient. If g satisfies (3), and for every constant $c > 0$, $(cg)^*$ is a multiple of g, then g must be a multiple of g_o.

Some remarks on proofs

There are several features of the interacting diffusion systems which seem "essential" for the comparison theorem. They are:

(i) A good ergodic theorem for the infinite system: If μ has "spatial density" θ then $\mu S(t) \to \nu_\theta$ as $t \to \infty$.

(ii) A very strong Feller-type property of the infinite system: If $\mu_n \Rightarrow \mu$ and $\mu S(t_n) \to \nu$, then $\mu_n S(t_n) \Rightarrow \nu$ for all $t_n \to \infty$.
(iii) A good comparison of finite/infinite systems, uniform over initial states, up to times $l_N \to \infty$ slowly.
(iv) $\Theta^N(\cdot)$ varies on the time scale β_N: if $l_N << \beta_N$, then

$$|\Theta^N(t\beta_N) - \Theta^N(t\beta_N + l_N)| \Rightarrow 0.$$

It is easy to see why (iv) holds with $\beta_N = |\Lambda_N| = (2N)^d$. ¿From the definition of $\Theta^N(\cdot)$ and (9),

$$d\Theta^N(t) = |\Lambda_N|^{-1} \sum_{i\in\Lambda_N} \sqrt{g(x_i^N(t))}\, dw_i(t),$$

and hence $\Theta^N(t)$ is a martingale with increasing process

$$[\Theta^N](t) = |\Lambda_N|^{-2} \int_0^t \sum_{i\in\Lambda_N} g(x_i^N(s))\, ds.$$

This shows that $\Theta^N(\cdot)$ does not vary over time scales smaller than β_N, since if $l_N \ll \beta_N$, then

$$\begin{aligned} E|\Theta^N(t\beta_N) - \Theta^N(t\beta_N + l_N)|^2 &= |\Lambda_N|^{-2} \int_{t\beta_N}^{t\beta_N + l_n} \sum_{i\in\Lambda_N} g(x_i^N(s))\, ds \\ &\le \frac{l_N \|g\|_\infty}{\beta_N} \end{aligned}$$

which tends to 0 as $N \to \infty$. Furthermore, $Z^N(t) = \Theta^N(t\beta_N)$ is a martingale, with increasing process

$$[Z_N](t) = |\Lambda_N|^{-1} \int_0^t \sum_{i\in\Lambda_N} g(x_i^N(s\beta_N))\, ds.$$

$\beta_N = (2N)^d$ is exactly the right time scale such that $[Z_N](t)$ can converge to a nondegenerate limit (as shown in [CGS]).

Finally, we state the coupling that is the main tool of [CGS]. Given two initial states $x(0)$, $y(0)$, a bivariate process $(x(t), y(t))$ can be constructed via

$$\begin{aligned} dx_i(t) &= [\sum_{j\in Zd} a(i,j)x_j(t) - x_i(t)]\, dt + \sqrt{g(x_i(t))}\, dw_i(t), \\ dy_i(t) &= [\sum_{j\in Zd} a(i,j)y_j(t) - y_i(t)]\, dt + \sqrt{g(y_i(t))}\, dw_i(t). \end{aligned}$$

The crucial point is that the same collection $\{w_i(t)\}$ of Brownian motions is used for both $x(t)$ and $y(t)$. It is straightforward to show that if $\mathcal{L}(x(0), y(0))$ is translation invariant, then $\frac{d}{dt}E|x_i(t) - y_i| \leq 0$. Hence $E|x_i(t) - y_i|$ is decreasing in t, and, in some cases, tends to zero as $t \to \infty$. In fact, this coupling is one of the key tools in proving Theorem 1. It is also possible to couple the infinite and finite systems into a bivariate process $(x(t), x^N(t))$, and obtain useful comparison estimates from the coupling.

References

[BCGH] J.-B.Baillon, Ph. Clément, A. Greven, F. den Hollander, *On the attracting orbit of a non-linear transformation arising from renormalization of hierarchically interacting diffusions*, Preprint (1993).

[CG1] J.T. Cox and A. Greven, *On the long term behavior of some finite particle systems*, Probab. Th. Rel. Fields **85** (1990), 195-237.

[CG2] J.T. Cox and A. Greven, *On the long term behavior of some finite particle systems: a critical dimension example*, Random Walks, Brownian Motion, and Interacting Particle Systems, R. Durrett and H. Kesten eds. (1991), Birkhäuser.

[CG3] J.T. Cox and A. Greven, *Ergodic theorems for infinite systems of interacting diffusions*, Ann. Probab. (1994).

[CGS] J.T. Cox, A. Greven, T. Shiga, *Finite and infinite systems of interacting diffusions*, In preparation. (1993).

[DG1] D.A. Dawson and A. Greven, *Multiple time scale analysis of interacting diffusions*, Prob. Th. Rel. Fields (1993).

[DG2] D.A. Dawson and A. Greven, *Models of Interacting Diffusions: Multiple Time Scale Phenomena, Phase Transition and Patterns of Cluster Formation*, Prob. Th. Rel. Fields (1993).

[D] J.-D. Deuschel, *Algebraic L_2 decay of critical processes on the lattice.*, Prerint.

[KS] N. Konno and T. Shiga, *Stochastic partial differential equations for some measure-valued diffusions*, Prob. Th. Rel. Fields (1988), 201-225.

[LS] T.M. Liggett and F. Spitzer, *Ergodic theorems for coupled random walks and other systems with locally interacting components*, Z. Wahrsch. verw. Gebiete **56** (1981), 443-468.

[NS] M. Notohara and T. Shiga, *Convergence to gentically uniform state in stepping stone models of population genetics*, J. Math. Bio. **10** (1980), 281-294.

[S1] T. Shiga, *An interacting system in population genetics I*, J. Math. Kyoto Univ. **20** (1980), 213-243.

[S2] T. Shiga, *An interacting system in population genetics II*, J. Math. Kyoto Univ. **20** (1980), 723-733.

[S3] T. Shiga, *A certain class of infinite-dimensional diffusion processes arising in population genetics*, J. Math. Soc. Japan (1987), 17-25.

[S4] T. Shiga, *Ergodic theorems and exponential decay of sample paths for certain interacting diffusion systems*, Osaka J. Math. (1992), 789-807.

[S5] T. Shiga, *Stationary distribution problem for interacting diffusions*, CRM Proceedings, Conference on Measure-valued Processes, Stochastic Partial Differential Equations and Interacting Systems (199?).

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

[SU] T. Shiga and K. Uchiyama, *Stationary states and their stability of the stepping stone model involving mutation and selection*, Prob. Th. Rel. Fields (1986), 87-117.

J.T.Cox
Mathematics Department
Syracuse University
Syracuse, New York 13244
U S A