The Annals of Probability
1991, Vol. 19, No. 4, 1587–1628

SOLUTIONS OF A STOCHASTIC DIFFERENTIAL EQUATION FORCED ONTO A MANIFOLD BY A LARGE DRIFT

BY G. S. KATZENBERGER

University of Wisconsin

We consider a sequence of $\mathbb{R}^d$-valued semimartingales $\{X_n\}$ satisfying

$$X_n(t) = X_n(0) + \int_0^t \sigma_n(X_n(s-))\,dZ_n(s) + \int_0^t F(X_n(s-))\,dA_n(s),$$

where $\{Z_n\}$ is a "well-behaved" sequence of $\mathbb{R}^e$-valued semimartingales, σ_n is a continuous $d \times e$ matrix-valued function, F is a vector field whose deterministic flow has an asymptotically stable manifold of fixed points Γ, and A_n is a nondecreasing process which asymptotically puts infinite mass on every interval. Many Markov processes with lower dimensional diffusion approximations can be written in this form. Intuitively, if $X_n(0)$ is close to Γ, the drift term $F\,dA_n$ forces X_n to stay close to Γ, and any limiting process must actually stay on Γ. If $X_n(0)$ is only in the domain of attraction of Γ under the flow of F, then the drift term immediately carries X_n close to Γ and forces X_n to stay close to Γ. We make these ideas rigorous, give conditions under which $\{X_n\}$ is relatively compact in the *Skorohod topology* and give a stochastic integral equation for the limiting process(es).

1. Introduction. To introduce our topic, consider the following small random perturbation of a dynamical system:

$$\phi_\varepsilon(x,t) = x + \varepsilon W(t) + \int_0^t F(\phi_\varepsilon(x,s))\,ds,$$

where W is d-dimensional Brownian motion and F is a vector field on $\mathbb{R}^d$. If F is locally Lipschitz, Gronwall's inequality implies that $\phi_\varepsilon \to \phi_0$, as $\varepsilon \to 0$, uniformly on compact subsets of $\mathbb{R}^d \times [0,\infty)$. Many authors, going back to Wentzell and Freidlin (1969), have considered such systems in a neighborhood of an isolated stable equilibrium point of the deterministic system ϕ_0. We are interested in the system near a manifold of stable equilibria.

Suppose that ϕ_0 has an asymptotically stable manifold of fixed points Γ. By this we mean that if the deterministic system ϕ_0 is started at a point x near Γ, then $\phi_0(x,t)$ converges as $t \to \infty$ to a point $\Phi(x) \in \Gamma$. The expected behavior of ϕ_ε on its natural time scale is to follow the trajectories of ϕ_0 as they converge to points of Γ. In particular, if we start ϕ_ε at a point on Γ, then in the limit as $\varepsilon \to 0$ the system just sits there. To capture the interesting behavior of the random system, we need to speed up time so that the random term does not go away in the limit. Using the scaling properties of Brownian motion, the

Received November 1989; revised July 1990.
AMS 1980 *subject classifications*. Primary 60H10; secondary 60J60, 60J70.
Key words and phrases. Stochastic differential equation, semimartingale, diffusion, diffusion approximation, manifold, flow.

process $\psi_\varepsilon(x,t) = \phi_\varepsilon(x, t/\varepsilon^2)$ satisfies

$$\psi_\varepsilon(x,t) = x + W(t) + \frac{1}{\varepsilon^2}\int_0^t F(\psi_\varepsilon(x,s))\,ds,$$

where W is a different d-dimensional Brownian motion than that appearing previously. Intuitively, when ε is small the drift term $\varepsilon^{-2}F(\psi_\varepsilon)\,dt$ very forcefully pushes ψ_ε toward Γ. If ψ_ε is started near Γ, the drift term easily overpowers the tendency of the Brownian motion to carry ψ_ε away from Γ. However, there is nothing preventing the Brownian motion from carrying the system tangentially along Γ. This suggests that if $y \in \Gamma$, then $\psi_\varepsilon(y, \cdot\,)$ should converge as $\varepsilon \to 0$ to some diffusion process on Γ. More generally, if x is in the domain of attraction of Γ under the deterministic system ϕ_0, then $\psi_\varepsilon(x,t)$ should follow $\phi_0(x, t/\varepsilon^2)$ until it is close to Γ, at which time ψ_ε is free to diffuse on Γ as before. Note that the convergence of $\phi_0(x, t/\varepsilon^2)$ to $\Phi(x)$ is very fast for ε small. Thus $\psi_\varepsilon(x, \cdot\,)$, in the limit as $\varepsilon \to 0$, should get "zapped" instantly to $\Phi(x)$, then diffuse nicely on Γ. This can be summarized by saying that the process

$$\psi_\varepsilon(x,t) + \phi_0(x, t/\varepsilon^2) + \Phi(x)$$

converges to a diffusion process on Γ. For t close to 0, $\psi_\varepsilon(x,t) - \phi_0(x, t/\varepsilon^2)$ should be close to 0; for t bounded away from 0, $\phi_0(x, t/\varepsilon^2) - \Phi(x)$ is close to 0.

We consider processes like ψ_ε in the sense that they are the sum of a well-behaved term and a "large" drift term which pushes toward a manifold. More precisely, we consider a sequence $\{X_n\}$ of cadlag $\mathbb{R}^d$-valued semimartingales satisfying

$$X_n(t) = X_n(0) + \int_0^t \sigma_n(X_n(s-))\,dZ_n(s) + \int_0^t F(X_n(s-))\,dA_n(s),$$

where $\{Z_n\}$ is a well-behaved sequence of cadlag $\mathbb{R}^e$-valued semimartingales, σ_n is a continuous $d \times e$ matrix-valued function, F is a vector field whose deterministic flow has an asymptotically stable manifold of fixed points Γ and A_n is a cadlag nondecreasing process which asymptotically puts infinite mass on every interval. Thus, in the preceding example, Brownian motion plays the role of the first integral term and the large drift plays the role of the second integral term. In the example, $A_n(t)$ would be $\varepsilon_n^{-2}t$, where $\varepsilon_n \to 0$ as $n \to \infty$.

Many Markov processes with lower dimensional diffusion approximations can be written in the general form given previously. Examples of such processes include population genetics models [see Ethier and Nagylaki (1980, 1988)], certain queueing systems [see Reiman (1983), Johnson (1983) and Peterson (1985)] and critical branching processes [see Kurtz (1978) and Joffe and Métivier (1986)].

Heuristically, if $X_n(0)$ is close to Γ, the drift term $F\,dA_n$ forces X_n to stay close to Γ, and any limiting process must actually stay on Γ. More generally, if $X_n(0)$ is in the domain of attraction of Γ under the deterministic flow of F, then the drift term immediately carries X_n close to Γ and forces X_n to stay close to Γ. We make these ideas rigorous, give conditions under which $\{X_n\}$ (when appropriately stopped) is relatively compact in the Skorohod topology and give a stochastic integral equation that any limiting process must satisfy.

What we mean by the "deterministic flow" of F depends on the jump behavior of A_n. We consider two cases: the asymptotically continuous case, when

$$\sup_{0<s\leq t} \Delta A_n(s) \Rightarrow 0,$$

where $\Delta A_n(t) = A_n(t) - A_n(t-)$ is the jump of A_n at time t, and the counting process case, when A_n is a counting process. In the asymptotically continuous case, the relevant flow is the usual continuous flow of F,

$$\phi(x,t) = x + \int_0^t F(\phi(x,s))\,ds.$$

In the counting process case, the relevant flow is the discrete dynamical system consisting of iterates of $x + F(x)$.

The counting process case is much more general than a first glance suggests. In particular, if A_n is a pure jump process with well-behaved jump sizes which are bounded and bounded away from 0, then the system can be reformulated as a process driven by a counting process. The counting process case is particularly useful in proving diffusion approximations for discrete-time processes, in which case $A_n(t)$ is typically $\lfloor \alpha_n t \rfloor$, where $\lfloor \cdot \rfloor$ denotes the greatest integer function and α_n is a sequence of real numbers going to ∞.

The basic strategy we use is to find a stochastic differential equation (SDE) for X_n with well-behaved coefficients and driving processes. We then apply results of Kurtz and Protter (1991) on convergence of stochastic integrals to get the results. In finding a "nice" SDE for X_n, we need to eliminate the large drift term in an appropriate way. To accomplish this in the asymptotically continuous case, we first find a C^2 function Φ: $\mathbb{R}^d \to \Gamma$ such that $\Phi(y) = y$ for $y \in \Gamma$ and $\partial\Phi(x)F(x) = 0$ for $x \in \mathbb{R}^d$. Applying Itô's formula to $\Phi(X_n)$ yields

$$\Phi(X_n(t)) = \Phi(X_n(0)) + \int_0^t \partial\Phi\, \sigma_n\, dZ_n + \tfrac{1}{2}\sum_{ijkl} \int_0^t \partial_{ij}\Phi\, \sigma_n^{ik}\sigma_n^{jl}\, d[Z_n^k, Z_n^l] + \eta_n(t),$$

where η_n represents jump correction terms. We then show [under appropriate conditions on $X_n(0)$] that $\eta_n \Rightarrow 0$ and $d(X_n, \Gamma) \Rightarrow 0$, which implies $X_n - \Phi(X_n) \Rightarrow 0$. Then X_n is a solution of

$$X_n(t) = X_n(0) + \int_0^t \partial\Phi\, \sigma_n\, dZ_n + \tfrac{1}{2}\sum_{ijkl} \int_0^t \partial_{ij}\Phi\, \sigma_n^{ik}\sigma_n^{jl}\, d[Z_n^k, Z_n^l] + \varepsilon_n(t),$$

where $\varepsilon_n \Rightarrow 0$ as $n \to \infty$. The function Φ, which plays a critical role in this strategy, is the limit map of the flow of F. A result of Falconer (1983) provides the necessary smoothness of Φ.

The main advantages of this strategy over the semigroup approach utilized in Ethier and Nagylaki (1980, 1988), for example, are the local nature of the estimates involved and the ability to handle the situation where the "fast" and "slow" processes are not separated.

Section 2 provides the necessary background in the Skorohod space, integration against cadlag functions, stochastic integration and general notation used in the remainder of the paper. In Section 3 we investigate exponential stability of Γ under the appropriate deterministic systems and construct certain strong Liapounov functions used in showing that the stochastic system stays close to Γ. Section 4 contains a result on relative compactness of stochastic integrals. In Section 5 we give conditions under which $d(X_n, \Gamma) \Rightarrow 0$ and give a result used in showing the jump correction terms go to 0 in the asymptotically continuous case. Section 6 contains the main results in the asymptotically continuous case and Section 7 handles the counting process case. Section 8 is devoted to examples.

2. Preliminaries. A function $g: [0, \infty) \to E$ mapping $[0, \infty)$ into a Banach space E is said to be *cadlag* if it is right-continuous at every $t \in [0, \infty)$ and has a left limit $g(t-)$ at every $t \in (0, \infty)$. $D_E[0, \infty)$ denotes the space of all cadlag functions from $[0, \infty)$ to E with the *Skorohod topology* [see Ethier and Kurtz (1986) for the definition and properties of the Skorohod topology]. For $g \in D_E[0, \infty)$, we write $\Delta g(t) = g(t) - g(t-)$ for the jump of g at time t and write $g^\tau(t)$ for $g(t \wedge \tau)$. If $g \in D_E[0, \infty)$ has finite variation on bounded intervals, we say simply that g has finite variation and write g^c for the continuous part of g,

$$g^c(t) = g(t) - \sum_{0 < s \le t} \Delta g(s).$$

Let $\mathbb{M}(d, e)$ be the set of all $d \times e$ matrices with real entries. If $g \in D_{\mathbb{R}^e}[0, \infty)$ has finite variation and $f \in D_{\mathbb{M}(d, e)}[0, \infty)$, we define the integral

$$\int_s^t f\,dg$$

as the limit of sums of the form

$$\sum_{i=0}^{n-1} f(r_i)(g(r_{i+1}) - g(r_i)),$$

where $s = r_0 < r_1 < \cdots < r_n = t$ and the limit is as the mesh of the partition goes to 0. Note that f is evaluated at the left-hand endpoint of the interval $(r_i, r_{i+1}]$ and not at an arbitrary point in the interval, as is traditional with Riemann–Stieltjes integration. If μ_g is the $\mathbb{R}^e$-valued measure defined by $\mu_g([0, t]) = g(t) - g(0)$, then

$$\int_s^t f\,dg = \int_{(s, t]} f(r-)\mu_g(dr),$$

where the integral on the right is the usual Lebesgue integral. Furthermore,

$$\int_s^t f\,dg = \int_s^t f\,dg^c + \sum_{s<r\le t} f(r-)\,\Delta g(r),$$

where the integral $\int_s^t f\,dg^c$ is equal to the usual Riemann–Stieltjes integral. We often write

$$\int_s^t f(r-)\,dg(r) \quad \text{for } \int_s^t f\,dg.$$

Define

$$\int_s^t f(r)\,dg(r) = \int_s^t f(r-)\,dg(r) + \sum_{s<r\le t} \Delta f(r)\,\Delta g(r).$$

Note that $\int_s^t f(r)\,dg(r) = \int_{(s,t]} f(r)\mu_g(dr)$, so the notation is justified. We emphasize that $\int_s^t f\,dg$ equals $\int_s^t f(r-)\,dg(r)$ and not, in general, $\int_s^t f(r)\,dg(r)$. If f and g have no jumps in common, then these are the same.

If $f, g \in D_{\mathbb{R}}[0,\infty)$ are both of finite variation, then we can integrate by parts:

$$\begin{aligned} f(t)g(t) &= f(0)g(0) + \int_0^t f\,dg + \int_0^t g\,df + \sum_{0<s\le t} \Delta f(s)\,\Delta g(s) \\ &= f(0)g(0) + \int_0^t f(s-)\,dg(s) + \int_0^t g(s)\,df(s). \end{aligned}$$

If $g \in D_{\mathbb{R}}[0,\infty)$ is of finite variation and $G\colon \mathbb{R} \to \mathbb{R}$ is C^1, then

$$\begin{aligned} G(g(t)) &= G(g(0)) + \int_0^t (G' \circ g)\,dg \\ &\quad + \sum_{0<s\le t} \left[\Delta G(g(s)) - G'(g(s-))\,\Delta g(s)\right] \\ &= G(g(0)) + \int_0^t (G' \circ g)\,dg^c + \sum_{0<s\le t} \Delta G(g(s)), \end{aligned}$$

where $\Delta G(g(s)) = G(g(s)) - G(g(s-))$. This change of variables formula can be extended to higher dimensions in the obvious way.

We also need to integrate against semimartingales. A semimartingale is the sum of a local martingale and an adapted finite variation process. If X and Y are adapted stochastic processes, X has sample paths in $D_{\mathbb{M}(d,e)}[0,\infty)$, Y has sample paths in $D_{\mathbb{R}^e}[0,\infty)$ and Y is a semimartingale, then the integral $\int_s^t X\,dY$ is defined, as in the deterministic setting, as the limit of sums

$$\sum_{i=0}^{n-1} X(r_i)\bigl(Y(r_{i+1}) - Y(r_i)\bigr),$$

where $s = r_0 < r_1 < \cdots < r_n = t$, the limit being in probability as the mesh size goes to 0. Standard results in stochastic calculus imply that this limit exists. We use the same notation conventions for stochastic integrals as for the deterministic integrals. For cadlag real-valued semimartingales X and Y, the integration by parts formula is

$$X(t)Y(t) = X(0)Y(0) + \int_0^t X\,dY + \int_0^t Y\,dX + [X,Y](t),$$

where $[X, Y](t)$ is the cross variation of X and Y on $(0, t]$, defined as the limit of sums

$$\sum_{i=0}^{n-1} (X(t_{i+1}) - X(t_i))(Y(t_{i+1}) - Y(t_i)),$$

where $0 = t_0 < t_1 < \cdots < t_n = t$ and the limit is in probability as the mesh size goes to 0. This limit exists for any cadlag semimartingales. If X is an $\mathbb{R}^d$-valued semimartingale, we write

$$[X] = \sum_{i=1}^{d} [X^i, X^i].$$

There is also a stochastic change of variables formula, called Itô's formula. If Y is a real-valued cadlag semimartingale and $G\colon \mathbb{R} \to \mathbb{R}$ is C^2, then

$$\begin{aligned} G(Y(t)) &= G(Y(0)) + \int_0^t G'(Y)\,dY + \tfrac{1}{2}\int_0^t G''(Y)\,d[Y] \\ &\quad + \sum_{0<s\le t} \Big[\Delta G(Y(s)) - G'(Y(s-))\,\Delta Y(s) \\ &\qquad\qquad - \tfrac{1}{2}G''(Y(s-))(\Delta Y(s))^2\Big] \\ &= G(Y(0)) + \int_0^t G'(Y)\,dY + \tfrac{1}{2}\int_0^t G''(Y)\,d[Y]^c \\ &\quad + \sum_{0<s\le t} [\Delta G(Y(s)) - G'(Y(s-))\,\Delta Y(s)]. \end{aligned}$$

This also has the obvious generalization to higher dimensions. In general, Y is not of finite variation so Y^c need not exist, but $[Y]$ is always of finite variation. For a complete discussion of stochastic integrals against cadlag semimartingales, see Protter (1990).

We need a form of Gronwall's inequality.

LEMMA 2.1. *Let $f, g \in D_{\mathbb{R}}[0,\infty)$ with g nondecreasing and $g(0) = 0$. Let $\eta \ge 0$ and assume*

$$0 \le f(t) \le \eta + \int_0^t f(s-)\,dg(s), \qquad t \ge 0. \tag{2.1}$$

Then $f(t) \le \eta e^{g(t)}$ for all $t \ge 0$.

PROOF. Iterating (2.1),

$$\begin{aligned} f(t) &\le \eta + \eta \sum_{n=1}^{\infty} \int_0^t \int_{(0,s_1)} \cdots \int_{(0,s_{n-1})} dg(s_n) \cdots dg(s_2)\,dg(s_1) \\ &\le \eta + \eta \sum_{n=1}^{\infty} \frac{1}{n!} g(t)^n = \eta e^{g(t)}. \end{aligned}$$

□

Let $d(\cdot, \cdot)$ denote the Euclidean metric on $\mathbb{R}^d$. By a *neighborhood* of a set A we mean an open set containing A. For a C^1 function $f\colon \mathbb{R}^d \to \mathbb{R}^e$, ∂f denotes the total derivative of f, that is, the $e \times d$ matrix of partial derivatives, $(\partial_j f^i)$. For nonnegative functions f and g, we write $f \ll g$ to indicate that $f \le cg$ for some constant c, and we write $f \sim g$ to indicate that $f \ll g$ and $g \ll f$. A function f is said to be LC^k if f is C^k and all of its kth-order derivatives are locally Lipschitz.

3. Deterministic results. In this section we study relevant deterministic systems. Let $U \subset \mathbb{R}^d$ be open and $F\colon U \to \mathbb{R}^d$ be a C^1 vector field. Assume $\Gamma = \{x | F(x) = 0\}$ is a C^0 submanifold of U of dimension m. Let $G\colon U \times [0, \infty) \to \mathbb{R}^d$ be continuous with $G(x, 0) = 0$ for $x \in U$ and $G(y, \xi) = 0$ for $y \in \Gamma$ and $\xi \ge 0$. Assume that for every compact $K \subset U$ and $\delta < \infty$, there exists $C = C(K, \delta) < \infty$ such that

$$(3.1) \qquad |G(x,\xi) - G(y,\xi)| \le C\xi |x-y|, \quad \text{for } x, y \in K \text{ and } 0 \le \xi \le \delta.$$

Usually, $G(x, \xi) = \xi F(x)$. Let $\mathscr{I}$ be the following class of integrators:

$$\mathscr{I} = \{a \in D_{\mathbb{R}}[0,\infty) | a \text{ is nondecreasing and } a(0) = 0\}.$$

For $\delta \ge 0$, let

$$\mathscr{I}(\delta) = \Big\{a \in \mathscr{I} \Big| \sup_{t \ge 0} \Delta a(t) \le \delta\Big\},$$

$$\mathscr{I}_\infty(\delta) = \{a \in \mathscr{I}(\delta) | a(\infty) = \infty\}.$$

For $\delta > 0$, let $D(\delta) = \{z \in \mathbb{C} |\, |\delta z + 1| < 1\}$ and let $D(0) = \{z \in \mathbb{C} | \operatorname{Re} z < 0\}$, where $\mathbb{C}$ is the complex numbers and $\operatorname{Re} z$ denotes the real part of z. Note that $D(\delta) \subset D(\eta)$ if $\delta \ge \eta$.

For $a \in \mathscr{I}$, define the flow of (F, G) driven by a to be the solution of

$$(3.2) \qquad \psi_a(x,t) = x + \int_0^t F(\psi_a(x,s))\, da^c(s) + \sum_{0 < s \le t} G(\psi_a(x, s-), \Delta a(s)).$$

When $G(x, \xi) = \xi F(x)$, (3.2) reduces to

$$(3.3) \qquad \psi_a(x,t) = x + \int_0^t F(\psi_a(x, s-))\, da(s).$$

The form of (3.2) is invariant under smooth coordinate transformations, while (3.3) is not. In light of (3.1) and Lemma 2.1, the standard Picard existence and Gronwall uniqueness proofs work to provide existence and uniqueness of ψ_a up until it leaves U. We say a set $S \subset U$ is $\mathscr{I}(\delta)$-invariant if S is ψ_a-invariant for every $a \in \mathscr{I}(\delta)$. For $a \in \mathscr{I}$, let

$$U_a = \Big\{x \in U \Big| \lim_{t \to \infty} \psi_a(x,t) \text{ exists and is in } \Gamma\Big\}$$

and, for $x \in U_a$, let $\Phi_a(x) = \lim_{t\to\infty} \psi_a(x,t)$. Note that $\Gamma \subset U_a$ and Φ_a is the identity on Γ.

The main results of this section are contained in Proposition 3.5. We show under eigenvalue conditions on ∂F that Γ is exponentially stable under ψ_a and we construct strong Liapounov functions for ψ_a used in the analysis of the stochastic equations.

First we consider the fundamental linear system driven by $a \in \mathscr{I}(\delta)$.

LEMMA 3.1. *Let $\delta \geq 0$ and $A \in \mathbb{M}(d, d)$ with the spectrum of A in $D(\delta)$. For $a \in \mathscr{I}$, let*

$$\Psi_a(t, s) = I + \int_s^t A\Psi_a(u-, s)\, da(u), \qquad t \geq s \geq 0.$$

Then there exists $Q \in \mathrm{GL}(d)$ and $\beta > 0$ with

$$|Q^{-1}\Psi_a(t, s)Q| \leq \exp[-\beta(a(t) - a(s))]$$

for all $a \in \mathscr{I}(\delta)$ and $t \geq s \geq 0$.

PROOF. If $\delta = 0$, then there exists $\delta' > 0$ such that the spectrum of A is in $D(\delta')$, so we can assume that $\delta > 0$. By considering $A' = \delta A$, we can take $\delta = 1$. Let $a \in \mathscr{I}(1)$ and let $B = I + A$. The spectral radius of B, $\rho(B)$, is less than 1. Examining the real Jordan canonical form implies $\rho(B) = \inf_{Q \in \mathrm{GL}(d)} |Q^{-1}BQ|$, so there exists $Q \in \mathrm{GL}(d)$ with $|Q^{-1}BQ| < 1$. For $x \in \mathbb{R}^d$, define $\|x\| = |Q^{-1}x|$. The operator norm of B relative to this norm is $\alpha \equiv \|B\| = |Q^{-1}BQ| < 1$. Note that

$$\Psi_a(t, s) = \exp[A(c(t) - c(s))] \prod_{s < r \leq t} (I + \Delta a(r)A),$$

where c is the continuous part of a. Let $\beta = 1 - \alpha$ so $\|e^{Au}\| = \|e^{Bu}\| e^{-u} \leq e^{-\beta u}$, for $u \geq 0$. Then $\|\exp[A(c(t) - c(s))]\| \leq \exp[-\beta(c(t) - c(s))]$. For $0 \leq \xi \leq 1$,

$$\|I + \xi A\| = \|\xi B + (1 - \xi)I\| \leq \xi\alpha + (1 - \xi) = 1 - \beta\xi \leq e^{-\beta\xi},$$

so that

$$\left\| \prod_{s < r \leq t} (I + \Delta a(r)A) \right\| \leq \prod_{s < r \leq t} \|I + \Delta a(r)A\| \leq \exp\left(-\beta \sum_{s < r \leq t} \Delta a(r)\right),$$

completing the proof. $\square$

Next we show local exponential stability of Γ under ψ_a when Γ is a linear subspace of $\mathbb{R}^d$.

LEMMA 3.2. *Assume U is a neighborhood of 0 and $\Gamma = N \cap U$ for some linear subspace $N \subset \mathbb{R}^d$. Assume $\delta \geq 0$, the matrix $\partial F(0)$ has $d - m$ eigenvalues in $D(\delta)$ and*

$$\lim_{x \to 0} \sup_{0 < \xi \leq \delta} \frac{|G(x, \xi) - \xi F(x)|}{\xi\, d(x, \Gamma)} = 0.$$

Then there exists a bounded $\mathscr{I}(\delta)$-invariant neighborhood W of 0, a C^2 function h: $W \to [0, \infty)$ and a constant $\beta > 0$ such that $\overline{W} \subset U$ and the following hold:

(i) *$h(\psi_a(x, t)) \le e^{-\beta a(t)} h(x)$, for all $a \in \mathscr{I}(\delta)$, $x \in W$ and $t \ge 0$;*

(ii) *$h \sim \gamma^2 w^{-3}$, $|\partial h| \ll \gamma w^{-3} + \gamma^2 w^{-4}$ and $|\partial^2 h| \ll w^{-3} + \gamma^2 w^{-5}$ on W, where $\gamma(x) = d(x, \Gamma)$ and $w(x) = d(x, W^c)$;*

(iii) *if $a \in \mathscr{I}_\infty(\delta)$, then $W \subset U_a$ and Φ_a is continuous on W.*

Moreover, 0 has a local base of $\mathscr{I}(\delta)$-invariant neighborhoods.

REMARK. Item (ii) implies that $1/h$, extended to be 0 outside W, is C^2 on $\mathbb{R}^d - \Gamma$.

PROOF OF LEMMA 3.2. If $\delta = 0$, then there exists $\delta' > 0$ such that $\partial F(0)$ has $d - m$ eigenvalues in $D(\delta')$. Thus [setting $G(x, \xi) = \xi F(x)$ if $\delta = 0$] we can take $\delta > 0$. By considering $\tilde{F} = \delta F$ and $\tilde{G}(x, \xi) = G(x, \delta\xi)$, take $\delta = 1$. Let $A = \partial F(0)$ and P be the range of A. N is the null space of A and $\mathbb{R}^d = P \oplus N$. Let π_P and π_N be the projections corresponding to this decomposition. For $a \in \mathscr{I}$, let Ψ_a be as in Lemma 3.1. Note that $\Psi_a = \pi_P \Psi_a \pi_P + \pi_N$ so P is Ψ-invariant. Also $\rho((I + A)|_P) < 1$, so Lemma 3.1 implies the existence of an invertible linear operator Q on P and a constant $\alpha > 0$ such that

$$|Q^{-1} \Psi_a(t, s)|_P Q| \le \exp[-\alpha(a(t) - a(s))], \tag{3.4}$$

for $a \in \mathscr{I}(1)$ and $t \ge s \ge 0$. Extend Q to $\mathbb{R}^d$ by $Q = Q\pi_P + \pi_N$ and define the vector norm $\|\cdot\|$ on $\mathbb{R}^d$ by $\|x\| = |Q^{-1}x|$. Then (3.4) becomes

$$\|\pi_P \Psi_a(t, s)\| \le \exp[-\alpha(a(t) - a(s))]. \tag{3.5}$$

Let $\theta(x) = F(x) - Ax$ and $\phi(x, \xi) = G(x, \xi) - \xi Ax$. Fix $\varepsilon > 0$ such that $\alpha - 2\varepsilon e^\alpha > 0$ and let $\beta = \alpha - 2\varepsilon e^\alpha$. Then $\partial\theta(0) = 0$ and there exists a convex neighborhood $V \subset U$ of 0 such that $\|\partial\theta\| \le \varepsilon$ on V, $\|G(x, \xi) - \xi F(x)\| \le \varepsilon \xi \|\pi_P x\|$ on $V \times [0, 1]$ and $V = \pi_P V + \pi_N V$. Since θ vanishes on N, for $x \in V$ and $0 \le \xi \le 1$,

$$\|\theta(x)\| = \left\| \int_0^1 \partial\theta(\pi_N x + t\pi_P x)\pi_P x \, dt \right\| \le \varepsilon \|\pi_P x\| \tag{3.6}$$

and

$$\|\phi(x, \xi)\| \le \|G(x, \xi) - \xi F(x)\| + \xi \|\theta(x)\| \le 2\varepsilon\xi\|\pi_P x\|. \tag{3.7}$$

Applying the integration by parts formula to $\Psi_a(t, r)\psi_a(x, r)$ yields

$$\begin{aligned} \psi_a(x, t) = {} & \Psi_a(t, s)\psi_a(x, s) + \int_s^t \Psi_a(t, r)\theta(\psi_a(x, r)) \, da^c(r) \\ & + \sum_{s < r \le t} \Psi_a(t, r)\phi(\psi_a(x, r-), \Delta a(r)). \end{aligned} \tag{3.8}$$

Let $\tau_a(x) = \inf\{t \ge 0 | \psi_a(x,t) \notin V\}$. Applying π_P to (3.8) and using (3.5)–(3.7) yields

$$\|\pi_P\psi_a(x,t)\| \le \exp[-\alpha(a(t) - a(s))]\|\pi_P\psi_a(x,s)\| + 2\varepsilon e^{\alpha}\int_s^t \exp[-\alpha(a(t) - a(r-))]\|\pi_P\psi_a(x,r-)\|\, da(r),$$

for $s \le t \le \tau_a(x)$ and $a \in \mathscr{I}(1)$. Then Lemma 2.1 (Gronwall's inequality) implies

$$(3.9)\quad \|\pi_P\psi_a(x,t)\| \le \exp[-\beta(a(t) - a(s))]\|\pi_P\psi_a(x,s)\|, \qquad s \le t \le \tau_a(x).$$

Applying π_N to (3.8), noting that $\pi_N\Psi_a(t,r) = \pi_N$ and using (3.6), (3.7) and (3.9),

$$(3.10)\quad \begin{aligned} &\|\pi_N\psi_a(x,t) - \pi_N\psi_a(x,s)\| \\ &\quad\le 2\varepsilon\|\pi_P\psi_a(x,s)\|\int_s^t \exp[-\beta(a(r-) - a(s))]\, da(r) \\ &\quad\le \frac{2\varepsilon e^{\beta}}{\beta}\|\pi_P\psi_a(x,s)\|(1 - \exp[-\beta(a(t) - a(s))]), \end{aligned}$$

for $s \le t \le \tau_a(x)$. Letting $M = 2\varepsilon e^{\beta}/\beta$, setting $s = 0$ and adding M times (3.9) to (3.10) gives

$$(3.11)\quad \|\pi_N\psi_a(x,t)\| + M\|\pi_P\psi_a(x,t)\| \le \|\pi_N x\| + M\|\pi_P x\|, \qquad t \le \tau_a(x).$$

Fix $\rho > 0$ small enough that the closure of

$$W_0 \equiv \{x \,|\, \|\pi_N x\| + M\|\pi_P x\| < \rho\}$$

is contained in V. Then (3.11) implies that $\tau_a(x) = \infty$ for $x \in W_0$ and that W_0 is $\mathscr{I}(1)$-invariant.

Let $f(x) = \rho^{-1}\|\pi_N x\|$, $g(x) = M\rho^{-1}\|\pi_P x\|$, $v = g + \sqrt{f^2 + g^2}$, $u = (1 - v)^3 \vee 0$ and $W = \{x | u(x) > 0\}$. On W define $h = g^2/u$. Note that $f + g \le v$ so $W \subset W_0$. If $x_1, y_1, x_2, y_2 \ge 0$ with $y_1 \le y_2$ and $x_1 + y_1 \le x_2 + y_2$, then

$$y_1 + \sqrt{x_1^2 + y_1^2} \le y_2 + \sqrt{x_2^2 + y_2^2}.$$

This, (3.9) and (3.11) imply that $v(\psi_a(x,t)) \le v(x)$ and $h(\psi_a(x,t)) \le e^{-2\beta a(t)}h(x)$ for $a \in \mathscr{I}(1)$, $x \in W$ and $t \ge 0$.

Note that

$$h = \frac{g^2\left(1 - g + \sqrt{f^2 + g^2}\right)^3(1 + 2g - f^2)^3}{\left((1 - f^2)^2 - 4g^2\right)^3}.$$

If $(1 - f^2)^2 \le 4g^2$ then $v \ge \frac{1}{2}|1 - f^2| + \frac{1}{2}|1 + f^2| \ge 1$, so $(1 - f^2)^2 - 4g^2 > 0$ on W. Since f^2 and g^2 are C^{∞}, to show that h is C^2 it suffices to show that the numerator is C^2. This is left to the reader.

It is clear that $u \sim w^3$ and $g \sim \gamma$ on W, implying the first estimate in item (ii). The other estimates are straightforward.

For $a \in \mathscr{I}_\infty(1)$ and $t \geq s \geq 0$, (3.9) and (3.10) imply

$$(3.12)\quad \|\psi_a(x,t) - \psi_a(x,s)\| \leq C\|\pi_P \psi_a(x,s)\| \leq C\|\pi_P x\| e^{-\beta a(s)} \to 0$$

uniformly on W as $s \to \infty$. Thus $\psi_a(x,t)$ converges uniformly on W to a continuous function $\Phi_a(x) \in \overline{W}$. But $\pi_P \psi_a(x,t) \to 0$ so $\Phi_a(x) \in \Gamma$ and $W \subset U_a$. □

Henceforth, take $G(x,\xi) = \xi F(x)$ so that (3.3) holds. This is not necessary, but simplifies the discussion [see Katzenberger (1990)]. The next lemma provides local results without the assumption that Γ is a subspace.

LEMMA 3.3. *Let $y_0 \in \Gamma$ and $\delta \geq 0$. Assume Γ is C^2 and $\partial F(y_0)$ has $d - m$ eigenvalues in $D(\delta)$. Then there exists a bounded $\mathscr{I}(\delta)$-invariant neighborhood W of y_0, a C^2 function h: $W \to [0,\infty)$ and a constant $\beta > 0$ such that $\overline{W} \subset U$ and items* (i)–(iii) *of Lemma* 3.2 *hold. Moreover, y_0 has a local base of $\mathscr{I}(\delta)$-invariant neighborhoods.*

PROOF. For notational convenience take $y_0 = 0$. Let $A = \partial F(0)$, let N be the tangent space of Γ at 0 and let P be the range of A. Then N is the null space of A and $\mathbb{R}^d = P \oplus N$. Let π_P and π_N be the projections corresponding to this decomposition. All the eigenvalues of $A|_P$ have negative real parts so $A|_P$ is invertible. Then, since Γ is C^2, there exist sets $V_P \subset P \cap U$ and $V_N \subset N \cap U$ and a C^2 function ϕ: $V_N \to V_P$ such that the following hold:

(i) V_P is a neighborhood of 0 in P and V_N is a neighborhood of 0 in N.
(ii) $V_P + V_N \subset U$.
(iii) For $v \in V_N$, there exists a unique element $\phi(v)$ in V_P such that $v + \phi(v) \in \Gamma$.

Let $V_0 = V_P + V_N$. Then $\Gamma \cap V_0$ is the graph of ϕ. Extend ϕ to $V_N \oplus P$ by defining $\phi(x) = \phi(\pi_N x)$. Then $\eta(x) = x + \phi(x)$ is a C^2 diffeomorphism of $V_N \oplus P$ onto itself with $\eta^{-1}(x) = x - \phi(x)$. Moreover, η carries V_N onto $\Gamma \cap V_0$. Let $\theta_a(x,t) = \eta^{-1}(\psi_a(\eta(x),t))$. Then θ_a solves

$$\theta_a(x,t) = x + \int_0^t \tilde{F}(\theta_a(x,s))\,da^c(s) + \sum_{0<s\leq t} \tilde{G}(\theta_a(x,s-),\Delta a(s)),$$

where

$$\tilde{F}(x) = \partial\eta^{-1}(\eta(x))F(\eta(x)),$$
$$\tilde{G}(x,\xi) = \eta^{-1}(\eta(x) + \xi F(\eta(x))) - x.$$

Note that $\partial\eta^{-1}(0) = I$, so in a neighborhood $V_1 \subset \eta^{-1}(V_0)$ of 0, $\partial\eta^{-1}(\eta(x))$ is invertible. Then $\{x \in V_1 | \tilde{F}(x) = 0\} = N \cap V_1$. Moreover, $\partial\tilde{F}(0) = \partial F(0)$ and, for x near 0 and $0 \leq \xi \leq \delta$, $|\tilde{G}(x,\xi) - \xi\tilde{F}(x)| \ll \xi|\eta(x)|d(x,N)$, implying

$$\lim_{x\to 0}\ \sup_{0<\xi\leq\delta} \frac{|\tilde{G}(x,\xi) - \xi\tilde{F}(x)|}{\xi\, d(x,N)} = 0.$$

Then $\tilde{F}$ and $\tilde{G}$ restricted to V_1 satisfy the conditions of Lemma 3.2. The result now follows from that lemma and the smoothness of η. □

The following lemma is used to patch together the local Liapounov functions constructed in Lemma 3.2 and Lemma 3.3.

LEMMA 3.4. *Assume that* h_1: $W_1 \to [0,\infty)$ *and* h_2: $W_2 \to [0,\infty)$ *are* C^2 *functions with* $h_i \sim \gamma^2 w_i^{-3}$, $|\partial h_i| \ll \gamma w_i^{-3} + \gamma^2 w_i^{-4}$ *and* $|\partial^2 h_i| \ll w_i^{-3} + \gamma^2 w_i^{-5}$ *on* W_i, *where* $\gamma(x) = d(x,\Gamma)$ *and* $w_i(x) = d(x, W_i^c)$. *Let* $g_i = 1/h_i$, *extended to be* 0 *outside* W_i, *and let* $W = W_1 \cup W_2$. *Then the function* h: $W \to [0,\infty)$ *defined by*

$$h(x) = \begin{cases} 0, & \text{for } x \in W \cap \Gamma, \\ \dfrac{1}{g_1(x) + g_2(x)}, & \text{for } x \in W - \Gamma, \end{cases}$$

is C^2 *with* $h \sim \gamma^2 w^{-3}$, $|\partial h| \ll \gamma w^{-3} + \gamma^2 w^{-4}$ *and* $|\partial^2 h| \ll w^{-3} + \gamma^2 w^{-5}$ *on* W, *where* $w(x) = d(x, W^c)$.

The proof is technical but straightforward and so is omitted.

Next we extend the local results of Lemma 3.3 to a neighborhood of Γ.

PROPOSITION 3.5. *Let* $\delta \ge 0$. *Assume* Γ *is* C^2 *and, for every* $y \in \Gamma$, *the matrix* $\partial F(y)$ *has* $d - m$ *eigenvalues in* $D(\delta)$. *Then there exists an* $\mathscr{I}(\delta)$-*invariant neighborhood* $V \subset U$ *of* Γ *such that the following hold*:

(i) *If* $a \in \mathscr{I}_\infty(\delta)$, *then* $V \subset U_a$, U_a *is open and* Φ_a *is continuous on* U_a.

(ii) *For every compact* $K \subset V$ *there exists an* $\mathscr{I}(\delta)$-*invariant neighborhood* $V_K \subset V$ *of* K, *a* C^2 *function* h: $V_K \to [0,\infty)$ *and a constant* $\beta > 0$ *such that* $\bar{V}_K$ *is a compact subset of* U,

$$h(\psi_a(x,t)) \le e^{-\beta a(t)} h(x), \quad \text{for all } a \in \mathscr{I}(\delta),\, x \in V_K,\, t \ge 0,$$

and $h(x) \sim d(x,\Gamma)^2$ *and* $|\partial h(x)| \ll d(x,\Gamma)$ *on compact subsets of* V_K.

PROOF. For $y \in \Gamma$, let W_y and h_y be as in Lemma 3.3 and let $V = \bigcup_{y \in \Gamma} W_y$. Then for $a \in \mathscr{I}_\infty(\delta)$, we have $V \subset U_a$, so

$$U_a = \bigcup_{t \ge 0} \{x \in U | \psi_a(x,t) \in V\}.$$

For fixed t, the map $x \mapsto \psi_a(x,t)$ is continuous, so U_a is open. Lemma 3.3 implies that Φ_a is continuous on V. Let $K \subset U_a$ be compact and note that there exists $T < \infty$ with $\psi_a(K,T) \subset V$. Let $b(t) = a(t+T) - a(T)$. Then $\Phi_a(x) = \Phi_b(\psi_a(x,T))$ is continuous on K, so Φ_a is continuous on U_a.

Let $K \subset V$ and let $\{y_i | 1 \le i \le n\} \subset \Gamma$ with $K \subset V_K \equiv \bigcup_{i=1}^n W_i$, using the subscript i in place of y_i. Define $g_i = 1/h_i$, extended to be 0 outside of W_i, and define $h\colon V_K \to [0,\infty)$ by

$$h(x) = \begin{cases} 0, & \text{for } x \in V_K \cap \Gamma, \\ \dfrac{1}{g_1(x) + g_2(x) + \cdots + g_n(x)}, & \text{for } x \in V_K - \Gamma. \end{cases}$$

Applying Lemma 3.4 $(n-1)$ times implies that h is C^2 and proves the estimates on h and ∂h, while Lemma 3.3 provides the exponential decay of h along ψ_a. □

The following theorem of Falconer (1983) provides the smoothness of the limit maps which we need in studying the stochastic equations.

THEOREM [Falconer (1983)]. *Let $f\colon U \to \mathbb{R}^d$ be an LC^k mapping where $U \subset \mathbb{R}^d$ is open and $k \geq 0$. Suppose that $\Gamma = \{x \in U | f(x) = x\}$ is a C^1 submanifold of U of dimension m and that for every $y \in \Gamma$, the matrix $\partial f(y)$ has $d-m$ eigenvalues in $\{z \in \mathbb{C} | \, |z| < 1\}$. Write $f^{(n)}$ for the nth iterate of f. Then*

$$U_f = \left\{ x \in U \,\middle|\, \lim_{n\to\infty} f^{(n)}(x) \text{ exists and is in } \Gamma \right\}$$

is a neighborhood of Γ and $f^{(\infty)} = \lim_{n\to\infty} f^{(n)}$ is C^k on U_f.

COROLLARY 3.6. *Assume $k \geq 1$, F is LC^k on U, Γ is C^2 and, for every $y \in \Gamma$, the matrix $\partial F(y)$ has $d-m$ eigenvalues in $D(0)$. Let $c(t) = t$. Then Φ_c is C^k on U_c.*

PROOF. Let $f(x) = \psi_c(x,1)$. Standard results on smoothness of dynamical systems imply that f is LC^k. Proposition 3.5 implies that Γ is the fixed point set of f restricted to U_c. For $y \in \Gamma$, the derivative $\partial f(y) = \exp(\partial F(y))$ has $d-m$ eigenvalues in the open unit disk. Then Falconer's theorem provides the conclusion. □

COROLLARY 3.7. *Assume $k \geq 1$, F is LC^k on U, Γ is C^1 and, for every $y \in \Gamma$, the matrix $\partial F(y)$ has $d-m$ eigenvalues in $D(1)$. Let $d(t)$ be the greatest integer function $\lfloor t \rfloor$. Then Φ_d is C^k on U_d.*

PROOF. Falconer's theorem applies immediately to $f(x) = \psi_d(x,1) = x + F(x)$. □

REMARKS. In the preceding corollaries, ψ_c is the solution of

$$\psi_c(x,t) = x + \int_0^t F(\psi_c(x,s))\,ds$$

and $\psi_d(x,k)$ is just the kth iterate of $x + F(x)$.

4. Convergence of stochastic integrals. This section contains results on the convergence and relative compactness of stochastic integrals. We use results due to Kurtz and Protter (1991). Jakubowski, Mémin and Pagès (1989) contains similar results. The symbol $\Rightarrow$ indicates convergence in distribution.

For $\delta > 0$, define h_δ: $[0,\infty) \to [0,\infty)$ by

$$h_\delta(r) = \begin{cases} 0, & \text{for } r \le \delta, \\ 1 - \delta/r, & \text{for } r \ge \delta. \end{cases}$$

Define J_δ: $D_{\mathbb{R}^d}[0,\infty) \to D_{\mathbb{R}^d}[0,\infty)$ by

$$J_\delta(g)(t) = \sum_{0<s\le t} h_\delta(|\Delta g(s)|)\,\Delta g(s).$$

J_δ is continuous in the Skorohod topology and the map $g \mapsto g - J_\delta(g)$ simply truncates the jumps of g which are larger than δ in magnitude.

CONDITION 4.1 [Kurtz and Protter (1991)]. For $n \ge 1$, let Y_n be a $\{\mathscr{G}_t^n\}$-semimartingale with sample paths in $D_{\mathbb{R}^d}[0,\infty)$. Assume that for some $\delta > 0$ (allowing $\delta = \infty$) and every $n \ge 1$ there exist stopping times $\{\tau_n^k | k \ge 1\}$ and a decomposition of $Y_n - J_\delta(Y_n)$ into a local martingale M_n plus a finite variation process F_n such that $P[\tau_n^k \le k] \le 1/k$ and

$$\sup_{n\ge 1} E\big[[M_n](t \wedge \tau_n^k) + T_{t\wedge\tau_n^k}(F_n)\big] < \infty,$$

for every $t \ge 0$ and $k \ge 1$, where $T_t(\cdot)$ denotes total variation on the interval $[0,t]$.

REMARK. Condition 4.1 is satisfied if $Y_n - J_\delta(Y_n)$ has a decomposition $M_n + F_n$ such that both M_n and F_n have bounded jumps and $\{T_t(F_n) | n \ge 1\}$ is stochastically bounded for each $t \ge 0$. Condition 4.1 is an assumption of Theorem 2.2 of Kurtz and Protter (1991), which we use to provide convergence of stochastic integrals.

CONDITION 4.2. For $n \ge 1$, let Y_n be a $\{\mathscr{G}_t^n\}$-semimartingale with sample paths in $D_{\mathbb{R}^d}[0,\infty)$. Assume that for some $\delta > 0$ (allowing $\delta = \infty$) and every $n \ge 1$ there exist stopping times $\{\tau_n^k | k \ge 1\}$ and a decomposition of $Y_n - J_\delta(Y_n)$ into a local martingale M_n plus a finite variation process F_n such that $P[\tau_n^k \le k] \le 1/k$, $\{[M_n](t \wedge \tau_n^k) + T_{t\wedge\tau_n^k}(F_n) | n \ge 1\}$ is uniformly integrable for every $t \ge 0$ and $k \ge 1$ and

$$\lim_{\gamma\to 0} \limsup_{n\to\infty} P\Big[\sup_{0\le t\le T} \big(T_{t+\gamma}(F_n) - T_t(F_n)\big) > \varepsilon\Big] = 0, \tag{4.1}$$

for every $\varepsilon > 0$ and $T > 0$.

REMARKS. Condition 4.2 implies Condition 4.1. Under the uniform integrability condition, (4.1) is equivalent to requiring that $\{T_t(F_n)\}$ be relatively compact in $D_{\mathbb{R}}[0,\infty)$ and have continuous limit points [see Theorem 3.8.6(c) of Ethier and Kurtz (1986)].

PROPOSITION 4.3. *If $\{Y_n\}$ is a sequence of $\mathbb{R}^d$-valued cadlag semimartingales satisfying Condition* 4.1 *and $Y_n \Rightarrow Y$, then $[Y_n^i, Y_n^j] \Rightarrow [Y^i, Y^j]$ and*

$\{[Y_n^i, Y_n^j]\}$ *satisfies Condition* 4.1 (*with* $M_n = 0$). *Moreover, if* Y *is continuous, then* $\{[Y_n^i, Y_n^j]\}$ *satisfies Condition* 4.2 (*with* $M_n = 0$).

PROOF. Theorem 2.2 of Kurtz and Protter (1991) implies that Y is a semimartingale and

$$[Y_n^i, Y_n^j] = Y_n^i Y_n^j - \int Y_n^i \, dY_n^j - \int Y_n^j \, dY_n^i \Rightarrow [Y^i, Y^j].$$

Let $0 < \delta < \infty$ and $F_n = [Y_n^i, Y_n^j] - J_\delta([Y_n^i, Y_n^j])$. Then $T_t(F_n) \le [Y_n](t)$ so $\{T_t(F_n)\}$ is stochastically bounded for each $t \ge 0$. Let

$$C_k = \inf\left\{C > 1 \,\middle|\, \sup_{n \ge 1} P[T_k(F_n) \ge C - 1] \le 1/k\right\}$$

and $\tau_n^k = \inf\{t > 0 | T_t(F_n) \ge C_k\}$. Then $P[\tau_n^k \le k] \le 1/k$ and $T_{t \wedge \tau_n^k}(F_n) \le C_k + \delta$, so $\{[Y_n^i, Y_n^j]\}$ satisfies Condition 4.1.

Trivially, $\{T_{t \wedge \tau_n^k}(F_n) | n \ge 1\}$ is uniformly integrable for each $t \ge 0$ and $k \ge 1$. Note that $[Y_n] \Rightarrow [Y]$ and $T_t(F_n) - T_s(F_n) \le [Y_n](t) - [Y_n](s)$, for all $t \ge s \ge 0$. Thus $\{T_t(F_n)\}$ is relatively compact in $D_{\mathbb{R}}[0, \infty)$. If Y is continuous, then $[Y]$ is also continuous, so the limit points of $\{T_t(F_n)\}$ are continuous. □

Next we consider relative compactness of stochastic integrals with bounded, not necessarily convergent, integrands.

PROPOSITION 4.4. *For* $n \ge 1$, *let* (H_n, Y_n) *be a* $\{\mathscr{G}_t^n\}$-*adapted process with sample paths in* $D_{\mathbb{M}(d,e) \times \mathbb{R}^e}[0, \infty)$. *Assume that* Y_n *is a* $\{\mathscr{G}_t^n\}$-*semimartingale*, $\{Y_n\}$ *is relatively compact and satisfies Condition* 4.2 *and*

$$\sup_{0 < t \le T} |\Delta Y_n(t)| \Rightarrow 0. \tag{4.2}$$

Let $\hat{H}_n(t) = \sup_{0 \le s \le t} |H_n(s)|$ *and assume that* $\{\hat{H}_n(t)\}$ *is stochastically bounded for each* $t \ge 0$. *Then* $\{(Y_n, \int H_n \, dY_n)\}$ *is relatively compact in* $D_{\mathbb{R}^e \times \mathbb{R}^d}[0, \infty)$ *and satisfies Condition* 4.2.

PROOF. Let δ, M_n, F_n and τ_n^k be as in Condition 4.2. (4.2) implies

$$\int H_n \, dJ_\delta(Y_n) \Rightarrow 0,$$

so it suffices to consider $\int H_n \, dY_n^\delta$, where $Y_n^\delta = Y_n - J_\delta(Y_n)$. Condition 4.2 and (4.2) imply $[Y_n] - [M_n] \Rightarrow 0$, so Proposition 4.3 implies that $\{[M_n]\}$ is relatively compact and

$$\lim_{\gamma \to 0} \limsup_{n \to \infty} P\left[\sup_{0 \le t \le T} ([M_n](t + \gamma) - [M_n](t)) > \varepsilon\right] = 0,$$

for all $\varepsilon > 0$ and $T > 0$. We use Theorem 3.8.6(c) of Ethier and Kurtz (1986) to show relative compactness. Let $\mu_n(L) = \inf\{t \geq 0 | \hat{H}_n(t) \geq L\}$. Then, for $k \geq t$,

$$P\left[\left|\int_0^t H_n\, dM_n\right| > L_0\right] \leq P\left[\tau_n^k \leq k\right] + P\left[\hat{H}_n(t) \geq L\right]$$

$$+ \frac{1}{L_0^2} E\left[\left|\int_0^t H_n\, dM_n^{\tau_n^k \wedge \mu_n(L)}\right|^2\right]$$

$$\leq \frac{1}{k} + P\left[\hat{H}_n(t) \geq L\right] + \frac{L^2}{L_0^2} E\left[[M_n](t \wedge \tau_n^k)\right].$$

Taking the supremum over $n \geq 1$, letting $L_0 \to \infty$, then $L \to \infty$ and $k \to \infty$ we see that $\{\int_0^t H_n\, dM_n\}$ is stochastically bounded for each $t \geq 0$.

We now verify a form of the so-called Aldous condition. Let τ_n be a $\{\mathscr{G}_t^n\}$-stopping time with $\tau_n \leq T - \gamma$ a.s., where $T > \gamma > 0$. Then, for $0 \leq u \leq \gamma$ and $k \geq T$,

$$E\left[\left|\int_{\tau_n}^{\tau_n + u} H_n\, dM_n\right|^2 \wedge 1\right]$$

$$\leq P\left[\tau_n^k \leq k\right] + P\left[\hat{H}_n(T) \geq L\right] + E\left[\left|\int_{\tau_n}^{\tau_n + u} H_n\, dM_n^{\tau_n^k \wedge \mu_n(L)}\right|^2\right]$$

$$\leq \frac{1}{k} + P\left[\hat{H}_n(T) \geq L\right] + L^2 E\left[[M_n]^{\tau_n^k}(\tau_n + \gamma) - [M_n]^{\tau_n^k}(\tau_n)\right]$$

$$\leq \frac{1}{k} + P\left[\hat{H}_n(T) \geq L\right] + L^2 \varepsilon$$

$$+ L^2 E\left[[M_n]^{\tau_n^k}(T);\ \sup_{0 \leq t \leq T} ([M_n](t + \gamma) - [M_n](t)) > \varepsilon\right].$$

Taking the lim sup as $n \to \infty$, letting $\gamma \to 0$ and using the uniform integrability on the last term, then letting $\varepsilon \to 0$, then $L \to \infty$ and $k \to \infty$, we see that Theorem 3.8.6 of Ethier and Kurtz (1986) applies, so $\{\int H_n\, dM_n\}$ is relatively compact in $D_{\mathbb{R}^d}[0, \infty)$. Similarly, $\{\int H_n\, dF_n\}$ is relatively compact. Both of these are asymptotically continuous, so their sum is relatively compact and the sequence of pairs, $(Y_n, \int H_n\, dY_n)$, is relatively compact. That this sequence satisfies Condition 4.2 follows from the preceding estimates. □

5. Convergence of X_n to Γ. In this section we describe the precise setting for the remainder of the paper and show, using the deterministic results, that under suitable conditions $d(X_n, \Gamma) \Rightarrow 0$ in the Skorohod topology. Let U, F and Γ be as in Section 3, so that $U \subset \mathbb{R}^d$ is open, $F\colon U \to \mathbb{R}^d$ is a C^1 vector field and $\Gamma = F^{-1}(0)$ is a C^0 submanifold of U of dimension m.

For $n \geq 1$, let $(\Omega^n, \mathscr{F}^n, \{\mathscr{F}_t^n\}_{t \geq 0}, P)$ be a filtered probability space, let Z_n be an $\mathbb{R}^e$-valued cadlag $\{\mathscr{F}_t^n\}$-semimartingale with $Z_n(0) = 0$ and let A_n be a real-valued cadlag $\{\mathscr{F}_t^n\}$-adapted nondecreasing process with $A_n(0) = 0$. Let $\sigma_n\colon U \to \mathbb{M}(d, e)$ be continuous with $\sigma_n \to \sigma$ uniformly on compact subsets of U. Let X_n be an $\mathbb{R}^d$-valued cadlag $\{\mathscr{F}_t^n\}$-semimartingale satisfying

$$X_n(t) = X_n(0) + \int_0^t \sigma_n(X_n)\, dZ_n + \int_0^t F(X_n)\, dA_n,$$

for all $t \le \lambda_n(K)$ and all compact $K \subset U$, where

$$\lambda_n(K) = \inf\{t \ge 0 | X_n(t-) \notin \mathring{K} \text{ or } X_n(t) \notin \mathring{K}\}.$$

We assume further that for every compact $K \subset U$, the sequence $\{Z_n^{\lambda_n(K)}\}$ is relatively compact and satisfies Condition 4.1, and

$$\sup_{0 < t \le T \wedge \lambda_n(K)} |\Delta Z_n(t)| \Rightarrow 0 \tag{5.1}$$

as $n \to \infty$, for every $T > 0$.

We need the following conditions.

(C5.1) For every compact $K \subset U$, $\{Z_n^{\lambda_n(K)}\}$ satisfies Condition 4.2.

(C5.2) For every $T > \varepsilon > 0$ and compact $K \subset U$,

$$\inf_{0 \le t \le T \wedge \lambda_n(K) - \varepsilon} \big(A_n(t+\varepsilon) - A_n(t)\big) \Rightarrow \infty$$

as $n \to \infty$, where the infimum of the empty set is taken to be ∞.

REMARKS. The filtration $\{\mathscr{F}_t^n\}$ is not assumed to be right-continuous and complete. Nevertheless, $\lambda_n(K)$ is an $\{\mathscr{F}_t^n\}$-stopping time. (5.1) implies that any limit point (in distribution) of $\{Z_n^{\lambda_n(K)}\}$ is continuous. Theorem 2.2 of Kurtz and Protter (1991) implies that any such limit point is a semimartingale. Proposition 4.3 implies that $\{[Z_n^i, Z_n^j]^{\lambda_n(K)}\}$ is relatively compact and satisfies Condition 4.2. Moreover, if $Z_n^{\lambda_n(K)} \Rightarrow Z$ along some subsequence of the positive integers, then $[Z_n^i, Z_n^j]^{\lambda_n(K)} \to [Z^i, Z^j]$ along the same subsequence. Condition (C5.2) requires that, asymptotically, dA_n puts infinite mass on every interval. Katzenberger (1990) handles the more general equation

$$X_n(t) = X_n(0) + \int_0^t \sigma_n(X_n)\, dZ_n + \int_0^t F(X_n)\, dA_n^c + \sum_{0 < s \le t} G(X_n(s-), \Delta A_n(s)).$$

The results and proofs there are similar but slightly more complicated.

THEOREM 5.1. *Let $\delta \ge 0$, assume Γ is C^2 and that, for every $y \in \Gamma$, the matrix $\partial F(y)$ has $d - m$ eigenvalues in $D(\delta)$. Assume* (C5.1), (C5.2) *and*

$$P\left[\sup_{0 < t \le T \wedge \lambda_n(K)} \Delta A_n(t) \le \delta\right] \to 1,$$

for every $T > 0$ and compact $K \subset U$. Then $d(X_n^{\lambda_n(K)}, \Gamma) \Rightarrow 0$ in $D_{\mathbb{R}}[0, \infty)$.

We need a lemma.

LEMMA 5.2. *For $n \geq 1$, let Y_n be a real-valued cadlag process, let B_n be a cadlag nondecreasing process with $B_n(0) = 0$ and let τ_n be a random time. Assume the following*:

(i) $\{Y_n^{\tau_n}\}$ *is relatively compact in* $D_{\mathbb{R}}[0, \infty)$.
(ii) *For every* $T \geq 0$, $\sup_{0 < t \leq T \wedge \tau_n} |\Delta Y_n(t)| \Rightarrow 0$ *as* $n \to \infty$.
(iii) *For every* $T \geq 0$ *and* $\varepsilon > 0$, $\inf_{0 \leq t \leq T \wedge \tau_n - \varepsilon}(B_n(t + \varepsilon) - B_n(t)) \Rightarrow \infty$ *as* $n \to \infty$.

Then

$$\sup_{0 \leq t \leq T \wedge \tau_n} e^{-B_n(t)} \left| \int_0^t e^{B_n} \, dY_n \right| \Rightarrow 0, \tag{5.2}$$

for every $T \geq 0$.

REMARK. This lemma does not assume that Y_n is a semimartingale. Nevertheless, the integral in (5.2) makes sense as an appropriate limit of sums (or can be defined by integration by parts) since B_n is a finite variation process. We apply this lemma only when Y_n is a semimartingale and B_n is adapted.

PROOF OF LEMMA 5.2. By redefining $B_n(t)$ for $t > \tau_n$ to be $B_n(\tau_n) + n(t - \tau_n)$ and replacing Y_n with $Y_n^{\tau_n}$, we can take $\tau_n = \infty$. Moreover, if suffices to prove (5.2) when Y_n converges in distribution to a continuous process Y. By the Skorohod representation theorem we can assume that all of the random variables are on the same probability space and $Y_n \to Y$ uniformly on bounded time intervals a.s. We take $Y_n(0) = 0$. Integrating by parts,

$$e^{-B_n(t)} \int_0^t e^{B_n} \, dY_n = e^{-B_n(t)} \left[Y_n(t) + \int_0^t (Y_n(t) - Y_n(s)) \, d(e^{B_n(s)}) \right]. \tag{5.3}$$

Now

$$\sup_{0 \leq t \leq T} e^{-B_n(t)} |Y_n(t)| \leq \sup_{0 \leq t \leq T} |Y_n(t) - Y(t)| + \sup_{0 \leq t \leq T} e^{-B_n(t)} |Y(t)|,$$

which $\to 0$ in probability. Also,

$$\sup_{0 \leq t \leq T} e^{-B_n(t)} \int_0^t |Y_n(t) - Y_n(s)| \, d(e^{B_n(s)})$$

$$\leq 2 \sup_{0 \leq t \leq T} |Y_n(t) - Y(t)| + \sup_{0 \leq t \leq T} e^{-B_n(t)} \int_0^t |Y(t) - Y(s)| \, d(e^{B_n(s)}).$$

The first term on the right-hand side $\to 0$ a.s. as $n \to \infty$. The second term is

$$\leq \sup_{\substack{s,t\leq T\\ |s-t|\leq\varepsilon}} |Y(t) - Y(s)| + \sup_{\varepsilon\leq t\leq T} e^{-B_n(t)}\int_0^{t-\varepsilon} |Y(t) - Y(s)|\, d(e^{B_n(s)})$$

$$\leq \sup_{\substack{s,t\leq T\\ |s-t|\leq\varepsilon}} |Y(t) - Y(s)| + 2\sup_{0\leq t\leq T} |Y(t)| \sup_{\varepsilon\leq t\leq T} \exp\big[-(B_n(t) - B_n(t-\varepsilon))\big].$$

The second term $\to 0$ in probability as $n \to \infty$ and the first $\to 0$ a.s. as $\varepsilon \to 0$. □

Proof of Theorem 5.1. Let $V \subset U$ be as in Proposition 3.5. Then it suffices to prove the result for compact $K \subset V$. Fix a compact $K \subset V$, let V_K, h and $\beta > 0$ be as in Proposition 3.5 and write λ_n for $\lambda_n(K)$. It suffices to show that $h(X_n^{\lambda_n}) \Rightarrow 0$ on $\{X_n(0) \in K\}$, so we may as well assume that $X_n(0) \in K$ a.s. To simplify notation, we write $h_n(t)$ for $h(X_n(t))$, $\sigma_n(t)$ for $\sigma_n(X_n(t))$ and so on. Integrating by parts,

$$d(e^{\beta A_n} h_n) = e^{\beta A_n}\, dh_n + h_n\, d(e^{\beta A_n}) + \Delta h_n\, \Delta e^{\beta A_n}. \tag{5.4}$$

Itô's formula yields

$$d(e^{\beta A_n}) = \beta e^{\beta A_n}\, dA_n^c + \Delta e^{\beta A_n} \tag{5.5}$$

and

$$\begin{aligned} dh_n &= \partial h_n\, \sigma_n\, dZ_n + \partial h_n\, F_n\, dA_n^c + \tfrac{1}{2}\sum_{ij} \partial_{ij} h_n\, d\big[U_n^i, U_n^j\big]^c \\ &\quad + \Delta h_n - \partial h_n\, \sigma_n\, \Delta Z_n, \end{aligned} \tag{5.6}$$

where $U_n(t) = \int_0^t \sigma_n\, dZ_n$. Integrating (5.4) and using (5.5) and (5.6), we get that $h_n(t)$ equals

$$e^{-\beta A_n(t)} h_n(0) \tag{5.7}$$

$$+ e^{-\beta A_n(t)}\int_0^t e^{\beta A_n}\, \partial h_n\, \sigma_n\, dZ_n \tag{5.8}$$

$$+ e^{-\beta A_n(t)}\int_0^t e^{\beta A_n}(\partial h_n\, F_n + \beta h_n)\, dA_n^c \tag{5.9}$$

$$+ \tfrac{1}{2} e^{-\beta A_n(t)} \sum_{ij} \int_0^t e^{\beta A_n}\, \partial_{ij} h_n\, d\big[U_n^i, U_n^j\big]^c \tag{5.10}$$

$$\begin{aligned} &+ e^{-\beta A_n(t)} \sum_{0<s\leq t} e^{\beta A_n(s-)}\big[e^{\beta \Delta A_n(s)} h_n(s) - h_n(s-) \\ &\qquad - \partial h_n(s-)\sigma_n(s-)\, \Delta Z_n(s)\big]. \end{aligned} \tag{5.11}$$

Since $h \geq 0$, it suffices to show that the supremum over $0 \leq t \leq T \wedge \lambda_n$ of the positive part of each of these terms goes to 0. Proposition 3.5 implies

$\partial h F \leq -\beta h$ on V_K so (5.9) is less than or equal to 0 for $t \leq \lambda_n$. For (5.7),

$$\sup_{0 \leq t \leq T \wedge \lambda_n} e^{-\beta A_n(t)} h(X_n(0)) \leq h(X_n(0)) \Rightarrow 0.$$

For (5.8), let $Y_n(t) = \int_0^{t \wedge \lambda_n} \partial h_n \, \sigma_n \, dZ_n$. By Proposition 4.4, $\{Y_n\}$ is relatively compact, so Lemma 5.2 implies

$$\sup_{0 \leq t \leq T \wedge \lambda_n} e^{-\beta A_n(t)} \left| \int_0^t e^{\beta A_n} \, dY_n \right| \Rightarrow 0.$$

The supremum of (5.10) over $0 \leq t \leq T \wedge \lambda_n$ is

$$\leq C \sup_{0 \leq t \leq T \wedge \lambda_n} e^{-\beta A_n(t)} \int_0^t e^{\beta A_n} \, d[Z_n],$$

which $\Rightarrow 0$ by Proposition 4.3 and Lemma 5.2.

Now consider (5.11). Note that

$$\begin{aligned} & e^{\beta\xi} h(x + z + \xi F(x)) - h(x) - \partial h(x) z \\ (5.12) \qquad & \quad = e^{\beta\xi} h(x + \xi F(x)) - h(x) \\ & \qquad + e^{\beta\xi}[h(x + z + \xi F(x)) - h(x + \xi F(x)) - \partial h(x) z] \\ & \qquad + (e^{\beta\xi} - 1) \, \partial h(x) z. \end{aligned}$$

The set $K' = \{x + \xi F(x) | x \in K, 0 \leq \xi \leq \delta\}$ is a compact subset of V_K, since V_K is $\mathscr{I}(\delta)$-invariant. Let $\varepsilon > 0$ be small enough that $H = \{x | d(x, K') \leq \varepsilon\}$ is contained in V_K. If $\xi \leq \delta$, $x \in K$ and $|z| \leq \varepsilon$, then $x + z \in H$ and $x + \xi F(x) + tz \in H$ for all $t \in [0, 1]$. Then the first term on the right-hand side of (5.12) is less than or equal to 0, the magnitude of the third term is bounded by a constant times $\xi |z| \, d(x, \Gamma)$ and the magnitude of the second term is

$$= e^{\beta\xi} \left| \int_0^1 (\partial h(x + \xi F(x) + tz) - \partial h(x)) z \, dt \right|$$

$$\leq C|z|(\xi |F(x)| + |z|) \leq C|z|^2 + C|z| \xi \, d(x, \Gamma).$$

Thus, for $\xi \leq \delta$, $x \in K$ and $|z| \leq \varepsilon$,

$$\begin{aligned} (5.13) \qquad & \left[e^{\beta\xi} h(x + z + \xi F(x)) - h(x) - \partial h(x) z\right]^+ \\ & \quad \ll |z|^2 + |z| \xi \, d(x, \Gamma) \ll |z|^2 + |z| \xi. \end{aligned}$$

Let $M = \sup_{n \geq 1} \sup_{x \in K} |\sigma_n(x)|$ and let Λ_n be the event where

$$\sup_{0 < t \leq T \wedge \lambda_n} \Delta A_n(t) \leq \delta \quad \text{and} \quad \sup_{0 < t \leq T \wedge \lambda_n} |\Delta Z_n(t)| \leq \frac{\varepsilon}{M}.$$

Then $P[\Lambda_n] \to 1$. On Λ_n and for $t \in [0, T \wedge \lambda_n]$ we have $\Delta A_n(t) \leq \delta$, $X_n(t-) \in K$ and $|\sigma_n(t-) \Delta Z_n(t)| \leq \varepsilon$. Thus, on Λ_n, the supremum of (5.11) over

$[0, T \wedge \lambda_n]$ is

$$\leq C \sup_{0<t\leq T\wedge\lambda_n} |\Delta Z_n(t)| \sup_{0\leq t\leq T\wedge\lambda_n} e^{-\beta A_n(t)} \int_0^t e^{\beta A_n}\, dA_n$$

$$+ C \sup_{0\leq t\leq T\wedge\lambda_n} e^{-\beta A_n(t)} \int_0^t e^{\beta A_n}\, d[Z_n]$$

$$\leq C \sup_{0<t\leq T\wedge\lambda_n} |\Delta Z_n(t)| + C \sup_{0\leq t\leq T\wedge\lambda_n} e^{-\beta A_n(t)} \int_0^t e^{\beta A_n}\, d[Z_n].$$

The first part goes to 0 by assumption and the second part goes to 0 by Proposition 4.3 and Lemma 5.2. This completes the proof. □

The following result is used only in Section 6, but its proof is closely related to that of Theorem 5.1 and so is included in this section.

PROPOSITION 5.3. *Assume Γ is C^2 and, for every $y \in \Gamma$, the matrix $\partial F(y)$ has $d - m$ eigenvalues in $D(0)$. Assume $K \subset U$ is compact,*

$$\sup_{0<t\leq T\wedge\lambda_n(K)} \Delta A_n(t) \Rightarrow 0,$$

for every $T > 0$, and $d(X_n^{\lambda_n(K)}, \Gamma) \Rightarrow 0$. Then

$$\sum_{0<t\leq T\wedge\lambda_n(K)} d(X_n(t-), \Gamma)^2 (\Delta A_n(t))^2 \Rightarrow 0,$$

for every $T > 0$.

PROOF. By replacing U with a slightly smaller set, we can assume that there exists $\delta > 0$ such that, for every $y \in \Gamma$, the matrix $\partial F(y)$ has $d - m$ eigenvalues in $D(\delta)$. Let $V \subset U$ be as in Proposition 3.5. We can assume that $K \subset V$. Let V_K, h and $\beta > 0$ be as in Proposition 3.5, and we write λ_n for $\lambda_n(K)$. Since $h(x) \sim d(x, \Gamma)^2$ on V_K, it suffices to show that $\{\int_0^{T\wedge\lambda_n} h(X_n)\, dA_n\}$ is stochastically bounded. Assume that $X_n(0) \in K$ a.s.

For $0 \leq s \leq T \wedge \lambda_n$,

$$e^{\beta A_n(s)} \int_s^{T\wedge\lambda_n} e^{-\beta A_n}\, dA_n \leq \alpha_n,$$

where $\alpha_n = e^{\beta\xi_n}/\beta$ and $\delta_n = \sup_{0<t\leq T\wedge\lambda_n} \Delta A_n(t)$.

We use the notation conventions of the proof of Theorem 5.1 and label the terms (5.7)–(5.11) as r_n, u_n, v_n, w_n and y_n, respectively. Note that $v_n(t) \leq 0$ for $t \leq \lambda_n$. For r_n,

$$\int_0^{T\wedge\lambda_n} |r_n|\, dA_a \leq \alpha_n h_n(0) \Rightarrow 0. \tag{5.14}$$

For w_n, interchanging the order of integration,

$$\begin{aligned}\int_0^{T\wedge\lambda_n}|w_n^2|\,dA_n &\le C\int_0^{T\wedge\lambda_n}e^{-\beta A_n(t-)}\int_{(0,t)}e^{\beta A_n(s-)}\,s[Z_n](s)\,dA_n(t)\\ &\le C\alpha_n[Z_n](T\wedge\lambda_n),\end{aligned}\tag{5.15}$$

which is stochastically bounded.

Now consider y_n. Let $\varepsilon > 0$ be small enough that $H = \{x \mid d(x, K) \le 2\varepsilon\}$ is contained in V_K. If necessary, make $\delta > 0$ smaller so that $\sup_{x\in K}|F(x)| \le \varepsilon/\delta$. As in (5.13), for $\xi \le \delta$, $x \in K$ and $|z| \le \varepsilon$,

$$\begin{aligned}&\left[e^{\beta\xi}h(x+z+\xi F(x)) - h(x) - \partial h(x)z\right]^+\\ &\qquad \ll |z|^2 + |z|\xi\, d(x,\Gamma) \ll |z|^2 + \xi^2 h(x).\end{aligned}$$

Let $M = \sup_{n\ge 1}\sup_{x\in K}|\sigma_n(x)|$ and let Λ_n be the event where $\delta_n \le \delta$ and

$$\sup_{0<t\le T\wedge\lambda_n}|\Delta Z_n(t)| \le \frac{\varepsilon}{M}.$$

Then $P[\Lambda_n] \to 1$. On Λ_n and for $t \in [0, T\wedge\lambda_n]$ we have $\Delta A_n(t) \le \delta$, $X_n(t-) \in K$ and $|\sigma_n(t-)\,\Delta Z_n(t)| \le \varepsilon$, so

$$\begin{aligned}\int_0^{T\wedge\lambda_n}y_n\,dA_n &\le C\alpha_n\sum_{0<s<T\wedge\lambda_n}\left[|\Delta Z_n(s)|^2 + h_n(s-)(\Delta A_n(s))^2\right]\\ &\le C\alpha_n[Z_n](T\wedge\lambda_n) + C\alpha_n\delta_n\int_0^{T\wedge\lambda_n}h_n\,dA_n.\end{aligned}$$

Thus

$$\int_0^{T\wedge\lambda_n}y_n\,dA_n \le a_n + b_n\int_0^{T\wedge\lambda_n}h_n\,dA_n,\tag{5.16}$$

where $\{a_n\}$ is stochastically bounded and $b_n \Rightarrow 0$.

Before dealing with u_n, we need to estimate $\int(|v_n| + |y_n|)\,dA_n$. Note that for $\xi \le \delta$ and $x \in K$,

$$\begin{aligned}|e^{\beta\xi}h(x+\xi F(x)) - h(x)| &\le (e^{\beta\xi}-1)h(x) + |h(x+\xi F(x)) - h(x)|\\ &\ll \xi h(x),\end{aligned}$$

so that, for $\xi \le \delta$, $x \in K$ and $|z| \le \varepsilon$,

$$|e^{\beta\xi}h(x+z+\xi F(x)) - h(x) - \partial h(x)z| \ll |z|^2 + \xi h(x).$$

Then, on Λ_n,

$$\int_0^{T\wedge\lambda_n}(|v_n| + |y_n|)\,dA_n \le C\alpha_n[Z_n](T\wedge\lambda_n) + C\alpha_n\int_0^{T\wedge\lambda_n}h_n\,dA_n.$$

This, (5.14) and (5.15) give an a priori estimate on $\int |u_n|\, dA_n$:

$$\int_0^{T\wedge\lambda_n} |u_n|\, dA_n \le \int_0^{T\wedge\lambda_n} (h_n + |r_n| + |v_n| + |w_n| + |y_n|)\, dA_n$$
$$\le c_n + d_n \int_0^{T\wedge\lambda_n} h_n\, dA_n,$$

where $\{c_n\}$ and $\{d_n\}$ are stochastically bounded. This is too weak to be useful directly, but it is used below to bound $\int u_n\, dA_n$.

Let $Y_n(t) = \int_0^t \partial h_n \sigma_n\, dZ_n$, so that $u_n(t) = e^{-\beta A_n(t)} \int_0^t e^{\beta A_n}\, dY_n$. Theorem 2.2 of Kurtz and Protter (1990) implies that $Y_n^{\lambda_n} \Rightarrow 0$. Integrating by parts,

$$u_n(t) = e^{-\beta A_n(t)} \left[Y_n(t) + \int_0^t (Y_n(t) - Y_n(s))\, d(e^{\beta A_n(s)}) \right]$$

so $u_n^{\lambda_n} \Rightarrow 0$. Let $V_n(t) = e^{\beta A_n(t)} u_n(t) = \int_0^t e^{\beta A_n}\, dY_n$. Integrating by parts,

$$\int_0^{T\wedge\lambda_n} \beta u_n\, dA_n = \int_0^{T\wedge\lambda_n} \beta e^{-\beta A_n} V_n\, dA_n$$

$$\begin{aligned} &= -u_n(T\wedge\lambda_n) + \int_0^{T\wedge\lambda_n} e^{-\beta A_n}\, dV_n \\ &\quad + \sum_{0<t\le T\wedge\lambda_n} u_n(t-)\big(e^{-\beta\Delta A_n(t)} - 1 + \beta\Delta A_n(t)\big) \\ &\quad + \sum_{0<t\le T\wedge\lambda_n} \Delta Y_n(t)(e^{-\beta\Delta A_n(t)} - 1). \end{aligned} \tag{5.17}$$

Now, $u_n(T\wedge\lambda_n) \Rightarrow 0$ and $\int_0^{T\wedge\lambda_n} e^{-\beta A_n}\, dV_n = Y_n(T\wedge\lambda_n) \Rightarrow 0$. The last term on the right-hand side of (5.17) is

$$\le C \sum_{0<t\le T\wedge\lambda_n} |\partial h_n(t-)|\, |\Delta Z_n(t)|\, \Delta A_n(t)$$
$$\le C \sum_{0<t\le T\wedge\lambda_n} \left[|\Delta Z_n(t)|^2 + |\partial h_n(t-)|^2 (\Delta A_n(t))^2 \right]$$
$$\le C[Z_n](T\wedge\lambda_n) + C\delta_n \int_0^{T\wedge\lambda_n} h_n\, dA_n.$$

Finally, the third term on the right-hand side of (5.17) is

$$\le C \sum_{0<t\le T\wedge\lambda_n} |u_n(t-)| (\Delta A_n(t))^2 \le C\delta_n \int_0^{T\wedge\lambda_n} |u_n|\, dA_n$$
$$\le C\delta_n c_n + C\delta_n d_n \int_0^{T\wedge\lambda_n} h_n\, dA_n.$$

Thus

$$\int_0^{T\wedge\lambda_n} u_n\, dA_n \le a'_n + b'_n \int_0^{T\wedge\lambda_n} h_n\, dA_n, \tag{5.18}$$

where $\{a'_n\}$ is stochastically bounded and $b'_n \Rightarrow 0$. Combining (5.14)–(5.16) with (5.18),

$$\int_0^{T\wedge\lambda_n} h_n\, dA_n \le \int_0^{T\wedge\lambda_n} (|r_n| + u_n + |w_n| + y_n)\, dA_n$$
$$\le a_n'' + b_n'' \int_0^{T\wedge\lambda_n} h_n\, dA_n,$$

where $\{a_n''\}$ is stochastically bounded and $b_n'' \Rightarrow 0$. This completes the proof. □

6. Asymptotically continuous case. Assume, in addition to the assumptions of Section 5, that

$$\sup_{0<t\le T\wedge\lambda_n(K)} \Delta A_n(t) \Rightarrow 0, \tag{6.1}$$

for all $T > 0$ and all compact $K \subset U$. Let $\psi(x,t)$ be the solution of

$$\psi(x,t) = x + \int_0^t F(\psi(x,s))\, ds,$$

$U_\Gamma = \{x \in U | \lim_{t\to\infty} \psi(x,t) \text{ exists and is in } \Gamma\}$ and $\Phi(x) = \lim_{t\to\infty} \psi(x,t)$.

Theorem 6.1 is self-contained in the sense that is does not use any previous results of this paper, but assumes many of their conclusions. Theorem 6.2 combines Theorem 6.1 with results from Sections 3, 4 and 5, producing a result which assumes no a priori knowledge of the behavior of X_n, other than the convergence of $X_n(0)$ to Γ. Theorem 6.3 provides the big picture. It has the same assumptions as Theorem 6.2 except that the limit of $X_n(0)$ is only assumed to be in U_Γ and not Γ itself. The conclusions of Theorem 6.3 include the initial behavior of X_n, its "instant" translation along ψ to Γ. This is a boundary layer type result similar to those found in the singular perturbations literature.

THEOREM 6.1. *Assume U_Γ is a neighborhood of Γ, Φ is C^2 on U_Γ, $K \subset U$ is compact, $X_n(0) \Rightarrow X(0) \in \Gamma$, $d(X_n^{\lambda_n(K)}, \Gamma) \Rightarrow 0$ and*

$$\sum_{0<t\le T\wedge\lambda_n(K)} d(X_n(t-),\Gamma)^2 (\Delta A_n(t))^2 \Rightarrow 0, \tag{6.2}$$

for every $T > 0$. Then the sequence of triples $\{(X_n^{\lambda_n(K)}, Z_n^{\lambda_n(K)}, \lambda_n(K))\}$ is relatively compact in $D_{\mathbb{R}^d\times\mathbb{R}^e}[0,\infty) \times [0,\infty]$. If (X, Z, λ) is a limit point of this sequence then (X, Z) is a continuous semimartingale, $X(t) \in \Gamma$ for every t a.s., $\lambda \ge \inf\{t \ge 0 | X(t) \notin \mathring{K}\}$ a.s. and

$$X(t) = X(0) + \int_0^{t\wedge\lambda} \partial\Phi(X)\sigma(X)\, dZ + \frac{1}{2}\sum_{ijkl} \int_0^{t\wedge\lambda} \partial_{ij}\Phi(X)\sigma^{ik}(X)\sigma^{jl}(X)\, d[Z^k, Z^l]. \tag{6.3}$$

REMARKS. If A_n is continuous or

$$\left\{ \sum_{0<t\le T\wedge\lambda_n(K)} (\Delta A_n(t))^2 \right\}$$

is stochastically bounded in n for each $T > 0$ then (6.2) is automatically satisfied. If $Z_n^{\lambda_n(K)} \Rightarrow Z$, $\lambda_n(K) \Rightarrow \lambda$ and (6.3) has only one solution X which stays on Γ, then $X_n^{\lambda_n(K)} \Rightarrow X$. Note that (6.3) will have a unique solution (for a given Z and λ) if σ is locally Lipschitz and Φ is LC^2. Corollary 3.6 provides conditions for the required smoothness of Φ. Theorem 6.1 holds for more general coefficients σ_n [see Theorem 5.4 of Kurtz and Protter (1991)].

PROOF OF THEOREM 6.1. We can assume that $K \subset U_\Gamma$. Note that $\partial\Phi(x)F(x) = 0$ for $x \in U_\Gamma$. Using the notation conventions of the proof of Theorem 5.1, Itô's formula implies

$$\Phi(X_n(t)) = \Phi(X_n(0)) + \int_0^t \partial\Phi_n \sigma_n \, dZ_n + \frac{1}{2}\sum_{ijkl} \int_0^t \partial_{ij}\Phi_n \, \sigma_n^{ik}\sigma_n^{jl} \, d\left[Z_n^k, Z_n^l\right] + \eta_n(t), \tag{6.4}$$

where

$$\eta_n(t) = \sum_{0<s\le t} \left[\Delta\Phi_n(s) - \partial\Phi_n(s-)\,\Delta Y_n(s) - \frac{1}{2}\sum_{ij} \partial_{ij}\Phi_n(s-)\,\Delta Y_n^i(s)\,\Delta Y_n^j(s)\right] \tag{6.5}$$

and $Y_n(t) = \int_0^t \sigma_n \, dZ_n$. Let $\varepsilon > 0$ be so small that

$$H = \{x \mid d(x, K) \le 2\varepsilon\} \subset U_\Gamma,$$

and let $\delta > 0$ be so small that $\sup_{x\in K}|F(x)| \le \varepsilon/\delta$. Let

$$\rho(\alpha) = \max_{1\le i,j\le d} \max_{\substack{x,y\in H \\ |x-y|\le\alpha}} |\partial_{ij}\Phi(x) - \partial_{ij}\Phi(y)|. \tag{6.6}$$

Then, since Φ is C^2, $\rho(\alpha) \to 0$ as $\alpha \to 0$. Let $x \in K$, $|z| \le \varepsilon$ and $\xi \le \delta$ so that $x + tz + s\xi F(x) \in H$ for all $s, t \in [0,1]$. Then

$$\begin{aligned}
&\left|\Phi(x + z + \xi F(x)) - \Phi(x + \xi F(x)) - \partial\Phi(x)z - \frac{1}{2}\sum_{ij}\partial_{ij}\Phi(x)z^i z^j\right| \\
&\quad \le \left|\int_0^1\int_0^1 \sum_{ij}\left[\partial_{ij}\Phi(x + stz + s\xi F(x)) - \partial_{ij}\Phi(x)\right] t z^i z^j \, ds\, dt\right| \\
&\qquad + \left|\int_0^1\int_0^1 \partial_{ij}\Phi(x + stz + s\xi F(x))\xi z^i F^j(x)\, ds\, dt\right| \\
&\quad \ll \rho(|z| + \xi|F(x)|)|z|^2 + |z|\xi\, d(x,\Gamma)
\end{aligned}$$

and, since $\partial\Phi(x)F(x) = 0$,

$$\begin{aligned}
&|\Phi(x + \xi F(x)) - \Phi(x)| \\
&\quad \le \left|\int_0^1 [\partial\Phi(x + t\xi F(x)) - \partial\Phi(x)]\xi F(x)\, dt\right| \ll \xi^2 d(x,\Gamma)^2.
\end{aligned}$$

Thus, for $x \in K$, $|z| \leq \varepsilon$ and $\xi \leq \delta$,

$$\left|\Phi(x + z + \xi F(x)) - \Phi(x) - \partial\Phi(x)z - \tfrac{1}{2}\sum_{ij} \partial_{ij}\Phi(x)z^i z^j\right| \tag{6.7}$$
$$\ll \rho(|z| + \xi|F(x)|)|z|^2 + |z|\xi\, d(x, \Gamma) + \xi^2\, d(x, \Gamma)^2.$$

Let Λ_n be the event where

$$\sup_{0<t\leq T\wedge\lambda_n} \Delta A_n(t) \leq \delta \quad \text{and} \quad \sup_{0<t\leq T\wedge\lambda_n} |\Delta Y_n(t)| \leq \varepsilon,$$

and let

$$\alpha_n = \sup_{0<t\leq T\wedge\lambda_n} \left(|\Delta Y_n(t)| + |F(X_n(t-))|\,\Delta A_n(t)\right).$$

Then $P[\Lambda_n] \to 1$ and $\alpha_n \Rightarrow 0$. (6.7) implies that, on Λ_n,

$$\sup_{o\leq t\leq T\wedge\lambda_n} |\eta(t)|$$
$$\ll \rho(\alpha_n)[Z_n](T \wedge \lambda_n) + \sum_{0<s\leq T\wedge\lambda_n} d(X_n(s-), \Gamma)^2(\Delta A_n(s))^2$$
$$+ [Z_n](T \wedge \lambda_n)^{1/2}\left(\sum_{0<s\leq T\wedge\lambda_n} d(X_n(s-), \Gamma)^2(\Delta A_n(s))^2\right)^{1/2},$$

which $\Rightarrow 0$. Thus, $\eta_n^{\lambda_n} \Rightarrow 0$. Moreover, the assumption that $d(X_n^{\lambda_n}, \Gamma) \Rightarrow 0$ implies that $X_n^{\lambda_n} - \Phi(X_n^{\lambda_n}) \Rightarrow 0$ as well. Then (6.4) becomes

$$X_n(t) = X_n(0) + \int_0^t \partial\Phi_n\, \sigma_n\, dZ_n + \tfrac{1}{2}\sum_{ijkl} \int_0^t \partial_{ij}\Phi_n \sigma_n^{ik}\sigma_n^{jl}\, d\left[Z_n^k, Z_n^l\right] + \varepsilon_n(t),$$

where $\varepsilon_n^{\lambda_n} \Rightarrow 0$. Proposition 4.3 implies that $\{[Z_n^i, Z_n^j]\}$ is relatively compact and satisfies Condition 4.1. Then Theorem 5.4 of Kurtz and Protter (1991) yields the relative compactness of $\{(X_n^{\lambda_n}, Z_n^{\lambda_n}, \lambda_n)\}$. Let (X, Z, λ) be a limit point of this sequence. Theorem 2.2 of Kurtz and Protter (1991) and Proposition 4.3 imply that Z is a semimartingale and that (6.3) is satisfied. This in turn implies that (X, Z) is a semimartingale. The other assertions are elementary. □

THEOREM 6.2. *Assume that* Γ *is* C^2 *and, for every* $y \in \Gamma$, *the matrix* $\partial F(y)$ *has* $d - m$ *eigenvalues in* $D(0)$. *Assume* (C5.1) *and* (C5.2) *hold,* Φ *is* C^2 *(or* F *is* LC^2*) and* $X_n(0) \Rightarrow X(0) \in \Gamma$. *Then for every compact* $K \subset U$, *the conclusions of Theorem* 6.1 *hold.*

PROOF. This follows immediately from Proposition 3.5, Corollary 3.6, Theorem 5.1, Proposition 5.3 and Theorem 6.1. □

REMARK. Theorem 6.2 can also be proved using Proposition 4.4 in place of Theorem 5.4 of Kurtz and Protter (1991) to get the desired relative compactness.

THEOREM 6.3. *Assume that* Γ *is* C^2 *and, for every* $y \in \Gamma$, *the matrix* $\partial F(y)$ *has* $d - m$ *eigenvalues in* $D(0)$. *Assume* (C5.1) *and* (C5.2) *hold,* Φ *is* C^2 (*or* F *is* LC^2) *and* $X_n(0) \Rightarrow X(0) \in U_\Gamma$. *Let*

$$Y_n(t) = X_n(t) - \psi(X_n(0), A_n(t)) + \Phi(X_n(0))$$

and, for compact $K \subset U_\Gamma$, *let*

$$\mu_n(K) = \inf\{t \geq 0 | Y_n(t-) \notin \mathring{K} \text{ or } Y_n(t) \notin \mathring{K}\}.$$

Then, for every compact $K \subset U_\Gamma$, *the sequence* $\{(Y_n^{\mu_n(K)}, Z_n^{\mu_n(K)}, \mu_n(K))\}$ *is relatively compact in* $D_{\mathbb{R}^d \times \mathbb{R}^e}[0,\infty) \times [0,\infty]$. *If* (Y, Z, μ) *is a limit point of this sequence, then* (Y, Z) *is a continuous semimartingale,* $Y(t) \in \Gamma$ *for every* t *a.s.,* $\mu \geq \inf\{t \geq 0 | Y(t) \notin \mathring{K}\}$ *a.s. and*

$$\begin{aligned} Y(t) &= Y(0) + \int_0^{t\wedge\mu} \partial\Phi(Y)\sigma(Y)\,dZ \\ &\quad + \tfrac{1}{2}\sum_{ijkl}\int_0^{t\wedge\mu} \partial_{ij}\Phi(Y)\sigma^{ik}(Y)\sigma^{jl}(Y)\,d[Z^k, Z^l]. \end{aligned} \tag{6.8}$$

REMARK. Basically this theorem says that X_n follows the flow of F according to the clock $A_n(t)$ until X_n is close to Γ, then it stays close to Γ and moves according to the SDE given in (6.8). Notice that $\psi(X_n(0), A_n(t)) - \Phi(X_n(0))$ is small for t bounded away from 0. The theorem implies that $X_n(t) - \psi(X_n(0), A_n(t))$ is small for t close to 0. This is made precise in the proof.

PROOF OF THEOREM 6.3. Fix a compact $K \subset U_\Gamma$. We can assume that $X_n(0) \in K$ a.s. Let V be as in Proposition 3.5 (with $\delta = 0$). Then there exists $T > 0$ such that $\psi(K, T) \subset V$. Let V_K be a ψ-invariant neighborhood of $\psi(K, T)$, as guaranteed by Proposition 3.5. Note that V_K can be chosen so that $\overline{V}_k$ is a compact subset of U_Γ. Then

$$H = \overline{V}_K \cup \{\psi(x,t) | x \in K, 0 \leq t \leq T\}$$

is a compact ψ-invariant subset of U_Γ with $K \subset \mathring{H}$. By similar reasoning, there exists a compact ψ-invariant $L \subset U_\Gamma$ with $H \subset \mathring{L}$. Let

$$\begin{aligned} V_n(t) &= \psi(X_n(0), A_n(t)) \\ &= X_n(0) + \int_0^t F(V_n)\,dA_n^c + \sum_{0<s\leq t} \Delta V_n(s). \end{aligned}$$

Let $\beta = 2/d(H, L^c)$ and define

$$T_n = 1 \wedge \inf\left\{t > 0 \,\middle|\, \sup_{0\leq s\leq t} (\beta + A_n(s))|X_n(s) - V_n(s)| \geq 1\right\}.$$

Then T_n is an $\{\mathscr{F}_t^n\}$-stopping time with $T_n \leq \lambda_n(L)$. Let $U_n(t) = \int_0^t \sigma_n \, dZ_n$. Then

$$\begin{aligned}|X_n^{\lambda_n(L)}(t) - V_n^{\lambda_n(L)}(t)| \leq |U_n^{\lambda_n(L)}(t)| + C\int_0^{t\wedge\lambda_n(L)} |X_n - V_n| \, dA_n^c \\ + \sum_{0<s\leq t\wedge\lambda_n(L)} |F(X_n(s-))\,\Delta A_n(s) \\ - \psi(V_n(s-), \Delta A_n(s)) + V_n(s-)|.\end{aligned} \tag{6.9}$$

For $x, y \in L$ and $\xi \geq 0$,

$$\begin{aligned}|\xi F(x) - \psi(y,\xi) + y| &\leq \xi|F(x) - F(y)| + |\xi F(y) - \psi(y,\xi) + y| \\ &\ll \xi|x-y| + \xi^2.\end{aligned}$$

Then (6.9) becomes

$$\begin{aligned}&|X_n^{\lambda_n(L)}(t) - V_n^{\lambda_n(L)}(t)| \\ &\quad\leq |U_n^{\lambda_n(L)}(t)| + C\int_0^{t\wedge\lambda_n(L)} |X_n - V_n| \, dA_n + C \sum_{0<s\leq t\wedge\lambda_n(L)} (\Delta A_n(s))^2 \\ &\quad\leq \hat{U}_n(t) + C\delta_n(t) A_n(t\wedge\lambda_n(L)) + C\int_0^{t\wedge\lambda_n(L)} |X_n - V_n| \, dA_n,\end{aligned}$$

where

$$\hat{U}_n(t) = \sup_{0\leq s\leq t} |U_n^{\lambda_n(L)}(s)| \quad \text{and} \quad \delta_n(t) = \sup_{0<s\leq t\wedge\lambda_n(L)} \Delta A_n(s).$$

Lemma 2.1 (Gronwall's inequality) implies

$$\begin{aligned}|X_n^{\lambda_n(L)}(t) - V_n^{\lambda_n(L)}(t) &\leq \exp(CA_n(t\wedge\lambda_n(L))) \\ &\quad\times\big(\hat{U}_n(t) + C\delta_n(t)A_n(t\wedge\lambda_n(L))\big).\end{aligned}$$

Then, on $\{T_n < 1\}$,

$$\begin{aligned}1 &\leq \sup_{0\leq s\leq T_n} (\beta + A_n(s))|X_n(s) - V_n(s)| \\ &\leq (\beta + A_n(T_n))\exp(CA_n(T_n))\big(\hat{U}_n(T_n) + C\delta_n(T_n)A_n(T_n)\big).\end{aligned} \tag{6.10}$$

For $M > 0$, let $\gamma_n(M) = \inf\{t > 0 | A_n(t) \geq M\}$. Then (6.10) implies

$$1_{\{T_n < 1\wedge\gamma_n(M)\}} \leq (\beta + M)e^{CM}\big(\hat{U}_n(\gamma_n(M)) + CM\delta_n(\gamma_n(M))\big). \tag{6.11}$$

(C5.2) implies that $\gamma_n(M) \wedge \lambda_n(L) \Rightarrow 0$. Then the relative compactness of $U_n^{\lambda_n(L)}$ (by Proposition 4.4) and (6.1) imply that the right-hand side of (6.11) $\Rightarrow 0$, hence $P[T_n < 1;\ A_n(T_n) < M] \to 0$. Then (C5.2) implies that $A_n(T_n) \Rightarrow \infty$.

Let $N_0 = 1$ and, for $k \geq 1$, let

$$N_k = \min\Big\{N > N_{k-1} \,\Big|\, \inf_{n\geq N} P\big[A_n(T_n \wedge k^{-1}) \geq k\big] \geq 1 - 1/k\Big\}.$$

Each N_k is finite since $A_n(T_n \wedge \varepsilon) \Rightarrow \infty$, for fixed $\varepsilon > 0$. Define $\varepsilon_n = 1/k$, for $N_k \le n < N_{k+1}$, and let $\tau_n = T_n \wedge \varepsilon_n$. Then $\tau_n \Rightarrow 0$ and $A_n(\tau_n) \Rightarrow \infty$. By definition of T_n,

$$\sup_{0 \le t \le \tau_n} |X_n(t) - V_n(t)| \le \frac{1}{A_n(\tau_n)} + \sup_{0 < s \le \lambda_n(L) \wedge \varepsilon_n} (|\Delta X_n(s)| + |\Delta V_n(s)|),$$

which $\Rightarrow 0$. Then $X_n^{\tau_n} - V_n^{\tau_n} \Rightarrow 0$ in $D_{\mathbb{R}^d}[0, \infty)$.

$$|X_n(\tau_n) - \Phi(X_n(0))| \le |X_n(\tau_n) - V_n(\tau_n)| + |V_n(\tau_n) - \Phi(X_n(0))| \Rightarrow 0$$

so $X_n(\tau_n) \Rightarrow \Phi(X(0))$. Let

$$\tilde{X}_n(t) = X_n(t + \tau_n), \qquad \tilde{Z}_n(t) = Z_n(t + \tau_n) - Z_n(\tau_n),$$

$$\tilde{Y}_n(t) = Y_n(t + \tau_n), \qquad \tilde{A}_n(t) = A_n(t + \tau_n) - A_n(\tau_n)$$

and $\tilde{\lambda}_n(K) = \inf\{t \ge 0 | \tilde{X}_n(t-) \notin \mathring{K}$ or $\tilde{X}_n(t) \notin \mathring{K}\}$. For compact $K' \subset \mathring{L}$, $\tilde{\lambda}_n(K') \le \lambda_n(L)$ so $\tilde{X}_n$, $\tilde{Z}_n$ and $\tilde{A}_n$ satisfy the hypotheses of Theorem 6.2 on $\mathring{L}$. Then the result follows from $Y_n^{\tau_n} - Y_n(0) \Rightarrow 0$, $\tilde{Y}_n - \tilde{X}_n \Rightarrow 0$, $\tau_n \Rightarrow 0$ and Theorem 6.2. □

7. Counting process case. Throughout this section we assume A_n is a counting process. We also assume the following condition:

(C7.1) The process

$$\bar{Z}_n(t) = \sum_{0 < s \le t} \Delta Z_n(s)\, \Delta A_n(s)$$

exists, is an $\{\mathscr{F}_t^n\}$-semimartingale and, for every compact $K \subset U$, the sequence $\{\bar{Z}_n^{\lambda_n(K)}\}$ is relatively compact and satisfies Condition 4.1.

REMARKS. If Z_n is continuous, or if Z_n does not jump when A_n does, then $\bar{Z}_n = 0$ so the conditions are trivial. More importantly, If Z_n jumps only when A_n does, then $\bar{Z}_n = Z_n$ so the conditions are already assumed. This is typically the case when X_n is a step-interpolation of a discrete-time process.

Let $\psi(x, 0) = x$ and $\psi(x, k = 1) = \psi(x, k) + F(\psi(x, k))$ so that $\psi(x, k)$ is the kth iterate of $x + F(x)$. Let

$$U_\Gamma = \left\{x \in U \,\middle|\, \lim_{k \to \infty} \psi(x, k) \text{ exists and is in } \Gamma\right\}$$

and $\Phi(x) = \lim_{k \to \infty} \psi(x, k)$.

Theorem 7.1 is the counting process analog of Theorem 6.1, Theorem 7.2 corresponds to Theorem 6.2 and Theorem 7.3 to Theorem 6.3.

THEOREM 7.1. *Assume* (C7.1) *holds*, U_Γ *is a neighborhood of* Γ, Φ *is* C^2 *on* U_Γ, $X_n(0) \Rightarrow X(0) \in \Gamma$, $K \subset U$ *is compact and* $d(X_n^{\lambda_n(K)}, \Gamma) \Rightarrow 0$. *Then the sequence* $\{(X_n^{\lambda_n(K)}, Z_n^{\lambda_n(K)}, \lambda_n(K))\}$ *is relatively compact in* $D_{\mathbb{R}^d \times \mathbb{R}^e}[0, \infty) \times [0, \infty]$. *If* (X, Z, λ) *is a limit point of this sequence, then* (X, Z) *is a continuous* *semimartingale*, $X(t) \in \Gamma$ *for every* t *a.s.*, $\lambda \geq \inf\{t \geq 0 | X(t) \notin \mathring{K}\}$ *a.s. and*

$$X(t) = X(0) + \int_0^{t\wedge\lambda} \partial\Phi(X)\sigma(X)\,dZ + \frac{1}{2}\sum_{ijkl}\int_0^{t\wedge\lambda}\partial_{ij}\Phi(X)\sigma^{ik}(X)\sigma^{jl}(X)\,d[Z^k,Z^l]. \tag{7.1}$$

PROOF. We can assume that $K \subset U_\Gamma$. Using the notation conventions of the proof of Theorem 6.1, Itô's formula implies that (6.4) and (6.5) hold, where, as in Theorem 6.1, $Y_n(t) = \int_0^t \sigma_n\,dZ_n$. Let $K' = \{x + F(x) | x \in K\}$. Then K' is a compact subset of U_Γ. Let $\varepsilon > 0$ be small enough that

$$H = \{x | d(x, K') \leq \varepsilon\} \subset U_\Gamma.$$

Let ρ be as in (6.6), so that $\rho(\alpha) \to 0$ as $\alpha \to 0$. Note that $\Phi(x + F(x)) = \Phi(x)$ and $\partial\Phi(x + F(x))(I + \partial F(x)) = \partial\Phi(x)$. Then for $x \in K$ and $|z| \leq \varepsilon$,

$$\begin{aligned}\Phi(x + z + F(x)) - \Phi(x) - \partial\Phi(x)z - \tfrac{1}{2}\sum_{ij}\partial_{ij}\Phi(x)z^iz^j \\ = \sum_{ij}\int_0^1\int_0^t\big[\partial_{ij}\Phi(x + sz + F(x)) - \partial_{ij}\Phi(x)\big]z^iz^j\,ds\,dt \\ - \partial\Phi(x + F(x))\,\partial F(x)z.\end{aligned} \tag{7.2}$$

The absolute value of the integral term is bounded by $\rho(|z| + |F(x)|)|z|^2$. Let $\theta(x) = \partial\Phi(x + F(x))\,\partial F(x)$. Then θ is continuous on U_Γ and $\theta(y) = 0$ for $y \in \Gamma$. Let Λ_n be the event where $\sup_{0<t\leq T\wedge\lambda_n(K)}|\Delta Y_n(t)| \leq \varepsilon$ and let

$$\alpha_n = \sup_{0<t\leq T\wedge\lambda_n(K)}\big(|\Delta Y_n(t)| + |F(X_n(t-))|\big).$$

Then $P[\Lambda_n] \to 1$ and $\alpha_n \Rightarrow 0$. Equation (7.2) implies that, on Λ_n,

$$\sup_{0\leq t\leq T\wedge\lambda_n}|\eta(t)| \leq C\rho(\alpha_n)[Z_n](T\wedge\lambda_n) + \left|\int_0^{T\wedge\lambda_n}\theta(X_n)\sigma_n(X_n)\,d\overline{Z}_n\right|.$$

Note that $\rho(\alpha_n) \Rightarrow 0$ and $\theta(X_n^{\lambda_n}) \Rightarrow 0$, so Proposition 4.3 and Theorem 2.2 of Kurtz and Protter (1991) imply that $\eta_n^{\lambda_n} \Rightarrow 0$. The remainder of the proof follows that of Theorem 6.1. □

THEOREM 7.2. *Assume that* Γ *is* C^2 *and, for every* $y \in \Gamma$, *the matrix* $\partial F(y)$ *has* $d - m$ *eigenvalues in* $D(1)$. *Assume* (C5.1), (C5.2) *and* (C7.1) *hold*, Φ *is* C^2 (*or* F *is* LC^2) *and* $X_n(0) \Rightarrow X(0) \in \Gamma$. *Then, for every compact* $K \subset U$, *the conclusions of Theorem* 7.1 *hold*.

PROOF. This follows immediately from Proposition 3.5, Corollary 3.7, Theorem 5.1 and Theorem 7.1. □

THEOREM 7.3. *Assume that* Γ *is* C^2 *and, for every* $y \in \Gamma$, *the matrix* $\partial F(y)$ *has* $d - m$ *eigenvalues in* $D(1)$. *Assume* (C5.1), (C5.2) *and* (C7.1) *hold*, Φ *is*

C^2 (or F is LC^2) and $X_n(0) \Rightarrow X(0) \in U_\Gamma$. Let

$$Y_n(t) = X_n(t) - \psi(X_n(0), A_n(t)) + \Phi(X_n(0))$$

and, for compact $K \subset U_\Gamma$, let

$$\mu_n(K) = \inf\{t \geq 0 | Y_n(t-) \notin \mathring{K} \text{ or } Y_n(t) \notin \mathring{K}\}.$$

Then, for every compact $K \subset U_\Gamma$, the sequence $\{(Y_n^{\mu_n(K)}, Z_n^{\mu_n(K)}, \mu_n(K))\}$ is relatively compact in $D_{\mathbb{R}^d \times \mathbb{R}^e}[0,\infty) \times [0,\infty]$. If (Y, Z, μ) is a limit point of this sequence, then (Y, Z) is a continuous semimartingale, $Y(t) \in \Gamma$ for every t a.s., $\mu \geq \inf\{t \geq 0 | Y(t) \notin \mathring{K}\}$ a.s. and

$$
\begin{aligned}
Y(t) &= Y(0) + \int_0^{t\wedge\mu} \partial\Phi(Y)\sigma(Y)\,dZ \\
&\quad + \tfrac{1}{2}\sum_{ijkl} \int_0^{t\wedge\mu} \partial_{ij}\Phi(Y)\sigma^{ik}(Y)\sigma^{jl}(Y)\,d[Z^k, Z^l].
\end{aligned}
\tag{7.3}
$$

See the remark following the statement of Theorem 6.3.

PROOF OF THEOREM 7.3. Fix a compact $K \subset U_\Gamma$. We can assume that $X_n(0) \in K$ a.s. Let V be as in Proposition 3.5 (with $\delta = 1$). Then there exists $N > 0$ such that $\psi(K, N) \subset V$. Let V_K be a ψ-invariant neighborhood of $\psi(K, N)$, as guaranteed by Proposition 3.5. Note that V_K can be chosen so that $\overline{V}_K$ is a compact subset of U_Γ. Then

$$H = \overline{V}_K \cup \{\psi(x, k) | x \in K, 0 \leq k \leq N\}$$

is a compact ψ-invariant subset of U_Γ with $K \subset H$. By similar reasoning, there exists a compact ψ-invariant $L \subset U_\Gamma$ with $H \subset \mathring{L}$. Let

$$V_n(t) = \psi(X_n(0), A_n(t)) = X_n(0) + \int_0^t F(V_n)\,dA_n(s).$$

Let $\beta = 2/d(H, L^c)$ and define

$$T_n = 1 \wedge \inf\left\{t > 0 \,\Big|\, \sup_{0\leq s\leq t} (\beta + A_n(s))|X_n(s) - V_n(s)| \geq 1\right\}.$$

Then T_n is an $\{\mathscr{F}_t^n\}$-stopping time with $T_n \leq \lambda_n(L)$. Let $U_n(t) = \int_0^t \sigma_n\,dZ_n$. Then

$$
\begin{aligned}
|X_n^{\lambda_n(L)}(t) - V_n^{\lambda_n(L)}(t)| &\leq |U_n^{\lambda_n(L)}(t)| + \int_0^{t\wedge\lambda_n(L)} |F(X_n) - F(V_n)|\,dA_n \\
&\leq \hat{U}_n(t) + C\int_0^{t\wedge\lambda_n(L)} |X_n - V_n|\,dA_n,
\end{aligned}
\tag{7.4}
$$

where $\hat{U}_n(t) = \sup_{0\leq s\leq t} |U_n^{\lambda_n(L)}(s)|$. Lemma 2.1 (Gronwall's inequality) implies

$$|X_n^{\lambda_n(L)}(t) - V_n^{\lambda_n(L)}(t)| \leq \exp(CA_n(t \wedge \lambda_n(L)))\hat{U}_n(t).$$

As in the proof of Theorem 6.3, this implies that $A_n(T_n) \Rightarrow \infty$. Following the construction of τ_n in Theorem 6.3, we get

$$\sup_{0 \le t \le \tau_n} |X_n(t) - V_n(t)| \le \frac{1}{A_n(\tau_n)} + \sup_{0 < s \le \lambda_n(L) \wedge \varepsilon_n} |\Delta U_n(s)|,$$

which $\Rightarrow 0$. The remainder of the proof follows that of Theorem 6.3. □

To see how the preceding theorems can be applied to a more general setting, consider the system

$$(7.5) \quad X_n(t) = X_n(0) + \int_0^t \sigma_n(X_n)\, dZ_n + \sum_{0 < s \le t} G(X_n(s-), \Delta B_n(s)),$$

where B_n is a cadlag $\{\mathscr{F}_t^n\}$-adapted nondecreasing process with $B_n(0) = 0$ and $B_n^c \equiv 0$, so that B_n is a pure jump process. Suppose that there exists an $\{\mathscr{F}_t^n\}$-semimartingale δ_n and a counting process A_n such that

$$B_n(t) = \int_0^t \delta_n\, dA_n = \sum_{0 < s \le t} \delta_n(s-)\, \Delta A_n(s).$$

Then we can rewrite (7.5) as

$$\begin{bmatrix} X_n \\ \delta_n \end{bmatrix}(t) = \begin{bmatrix} X_n \\ \delta_n \end{bmatrix}(0) + \int_0^t \begin{bmatrix} \sigma_n(X_n) & 0 \\ 0 & 1 \end{bmatrix} d\begin{bmatrix} Z_n \\ \delta_n \end{bmatrix} + \int_0^t \begin{bmatrix} G(X_n, \delta_n) \\ 0 \end{bmatrix} dA_n,$$

which is of the form

$$\tilde{X}_n(t) = \tilde{X}_n(0) + \int_0^t \tilde{\sigma}_n(\tilde{X}_n)\, d\tilde{Z}_n + \int_0^t \tilde{F}(\tilde{X}_n)\, dA_n.$$

Thus Theorems 7.1, 7.2 and 7.3 apply to this situation. In particular, note that if δ_n is a constant and $\delta_n \to \delta > 0$, then the eigenvalue condition needed to apply Theorems 7.2 and 7.3 is that, for $y \in \Gamma$, the matrix $\partial_x G(y, \delta)$ has $d - m$ eigenvalues in $D(1)$.

8. Examples.

Diffusion. Let Φ and U_Γ be as in Section 6. Assume Γ is C^2 and, for every $y \in \Gamma$, the matrix $\partial F(y)$ has $d - m$ eigenvalues in $D(0)$. Let $\sigma_n \colon U \to \mathbb{M}(d, e)$ and $b_n \colon U \to \mathbb{R}^d$ be continuous with $\sigma_n \to \sigma$ and $b_n \to b$ uniformly on compact subsets of U. Let W be an e-dimensional Brownian motion and let X_n be a solution of

$$X_n(t) = X_n(0) + \int_0^t \sigma_n(X_n)\, dW + \int_0^t b_n(X_n)\, ds + \alpha_n \int_0^t F(X_n)\, ds,$$

where $\alpha_n \to \infty$ as $n \to \infty$. Theorems 6.2 and 6.3 apply immediately to this situation. However, if we require additional smoothness of σ_n, b_n and Φ, we can get convergence in probability. Assume $X_n(0) \to X(0) \in \Gamma$ in probability, Φ is LC^2 (or F is LC^3) and, for each compact $K \subset U$, there exists a constant

$L = L(K)$ such that

$$|\sigma_n(x) - \sigma_n(y)| \le L|x - y|,$$
$$|b_n(x) - b_n(y)| \le L|x - y|,$$

for all $x, y \in K$ and $n \ge 1$. Let X be the solution of

$$\begin{aligned} X(t) &= X(0) + \int_0^t \partial\Phi(X)\sigma(X)\,dW + \int_0^t \partial\Phi(X)b(X)\,ds \\ &\quad + \tfrac{1}{2}\sum_{ij}\int_0^t \partial_{ij}\Phi(X)a^{ij}(X)\,ds, \end{aligned} \tag{8.1}$$

where $a = (a^{ij}) = \sigma\sigma^T$. The coefficients in (8.1) are locally Lipschitz on U_Γ, so the solution of (8.1) is unique up until the first time it leaves U_Γ. For compact $K \subset U_\Gamma$, let $\lambda(K) = \inf\{t \ge 0 | X(t) \notin \mathring{K}\}$. Then we have the following theorem.

THEOREM. *$X_n^{\lambda(K)} \to X^{\lambda(K)}$ in $C_{\mathbb{R}^d}[0,\infty)$ in probability.*

REMARKS. Note that X_n is stopped at $\lambda(K)$, not $\lambda_n(K)$. $C_{\mathbb{R}^d}[0,\infty)$ is the space of continuous functions $f\colon [0,\infty) \to \mathbb{R}^d$ with the topology of uniform convergence on bounded time intervals. The convergence in Theorem 6.3 can also be shown to be in probability.

PROOF OF THE THEOREM. Condition (C5.2) is satisfied by $A_n(t) = \alpha_n t$. Let $H \subset U_\Gamma$ be compact with $K \subset \mathring{H}$. We write λ_n for $\lambda_n(H)$ and λ for $\lambda(H)$. Let $P_n = \partial\Phi\,\sigma_n$ and $P = \partial\Phi\,\sigma$. Theorem 6.2 implies that $\Phi(X_n^{\lambda_n}) - X_n^{\lambda_n} \to 0$, $P_n(X_n^{\lambda_n}) - P(X_n^{\lambda_n}) \to 0$ and $b_n(X_n^{\lambda_n}) - b_n(X_n^{\lambda_n}) \to 0$ in probability. Then (8.1) and Itô's formula applied to $\Phi(X_n)$ yield

$$\begin{aligned} |X_n^{\lambda_n \wedge \lambda}(t) - X^{\lambda_n \wedge \lambda}(t)| &\le \gamma_n(t) + \left|\int_0^{t\wedge\lambda_n\wedge\lambda}(P(X_n) - P(X))\,dW\right| \\ &\quad + C\int_0^{t\wedge\lambda_n\wedge\lambda}|X_n - X|\,ds, \end{aligned}$$

where γ_n is continuous, nondecreasing and adapted and $\gamma_n(t) \to 0$ in probability as $n \to \infty$. For $\varepsilon > 0$, let $\mu_n^\varepsilon = \lambda_n \wedge \lambda \wedge \inf\{t \ge 0 | \gamma_n(t) \ge \varepsilon\}$. Then by Doob's inequality,

$$\begin{aligned} &E\left[\sup_{0\le t\le T\wedge\mu_n^\varepsilon}|X_n(t) - X(t)|^2; \mu_n^\varepsilon > 0\right] \\ &\quad \le 2\varepsilon^2 + 8E\left[\int_0^{T\wedge\mu_n^\varepsilon}|P(X_n) - P(X)|^2\,ds; \mu_n^\varepsilon > 0\right] \\ &\qquad + 2C^2T\int_0^T E\left[\left|X_n^{\mu_n^\varepsilon} - X^{\mu_n^\varepsilon}\right|^2; \mu_n^\varepsilon > 0\right]ds \\ &\quad \le 2\varepsilon^2 + C_T\int_0^T E\left[\left|X_n^{\mu_n^\varepsilon} - X^{\mu_n^\varepsilon}\right|^2; \mu_n^\varepsilon > 0\right]ds. \end{aligned}$$

Applying Gronwall's inequality,

$$E\left[\sup_{0\le t\le T\wedge\mu_n^\varepsilon}|X_n(t)-X(t)|^2;\ \mu_n^\varepsilon>0\right]\le 2\varepsilon^2 e^{TC_T}.$$

Since $P[\mu_n^\varepsilon<\lambda_n\wedge\lambda]\to 0$, this implies $X_n^{\lambda_n\wedge\lambda}-X^{\lambda_n\wedge\lambda}\to 0$ in probability. Since $P[\lambda(K)<\lambda(H)\wedge\lambda_n(H)]\to 1$, we get that $X_n^{\lambda(K)}\to X^{\lambda(K)}$ in probability. □

Diffusion in shrinking domains with reflection at the boundary. Let ψ, Φ and U_Γ be as in Section 6. Assume that U_Γ is a neighborhood of Γ and Φ is C^2 on U_Γ. For $n\ge 1$, let $U_n\subset U_\Gamma$ be a neighborhood of Γ such that each flow line $\{\psi(x,t)|t\ge 0\}$ of F intersects ∂U_n in at most one point. Then, once a flow line enters $\overline{U}_n$, it is in U_n for all future times. Assume $U_{n+1}\subset U_n$, $\bigcap_{n\ge 1}U_n=\Gamma$ and $\sup_{x\in K\cap U_n}d(x,\Gamma)\to 0$ as $n\to\infty$ for all compact $K\subset U_\Gamma$. Let $\sigma_n\colon U\to\mathbb{M}(d,e)$ and $b_n\colon U\to\mathbb{R}^d$ be continuous with $\sigma_n\to\sigma$ and $b_n\to b$ uniformly on compact subsets of U. Let W be an e-dimensional Brownian motion and let (X_n,L_n) be a solution of

$$X_n(t)=X_n(0)+\int_0^t\sigma_n(X_n)\,dW+\int_0^t b_n(X_n)\,ds+\int_0^t F(X_n)\,dL_n$$

such that $X_n(t)\in\overline{U}_n$ for all $t\ge 0$ and L_n is a continuous nondecreasing process which increases only when $X_n(t)\in\partial U_n$, that is,

$$\int_0^t 1_{U_n}(X_n)\,dL_n=0.$$

Assume $X_n(0)\Rightarrow X(0)\in\Gamma$ [and $X_n(0)\in U_n$]. Theorem 6.1 implies that $\{X_n^{\lambda_n(K)}\}$ is relatively compact in $C_{\mathbb{R}^d}[0,\infty)$, and any limit point (X,λ) of $\{(X_n^{\lambda_n(K)},\lambda_n(K))\}$ satisfies

$$X(t)=X(0)+\int_0^{t\wedge\lambda}\partial\Phi(X)\sigma(X)\,dW+\int_0^{t\wedge\lambda}\partial\Phi(X)b(X)\,ds$$
$$+\tfrac12\sum_{ij}\int_0^{t\wedge\lambda}\partial_{ij}\Phi(X)a^{ij}(X)\,ds,$$

where $a=(a^{ij})=\sigma\sigma^T$.

Moran diploid model. Let $S=\{x\in\mathbb{M}(r,r)|x=x^T$ and $\mathbf{1}^Tx\mathbf{1}=1\}$ and $S_0=\{x\in S|x^{ij}\ge 0,\ 1\le i,\ j\le r\}$, where $\mathbf{1}$ denotes the vector of all 1's. Let $\{E_{ij}\}$ be the standard basis of $\mathbb{M}(r,r)$, $e_{ij}=\frac12(E_{ij}+E_{ji})$ and $\theta_{ij}=2-\delta_{ij}$. Then $\{e_{ij}|1\le i\le j\le r\}$ is a basis for the space of symmetric $r\times r$ matrices and $\{\theta_{ij}x^{ij}|1\le i\le j\le r\}$ are the coordinates of x with respect to this basis.

We consider a population of individuals which have two chromosomes and consider the two genes at a single locus of the chromosomes. The locus is assumed to be multiallelic, each gene taking values in $1,\dots,r$ and we classify individuals according to their genotype (i,j), $1\le i\le j\le r$. We represent the population by the matrix $x\in S_0$ such that $\theta_{ij}x^{ij}$ is the proportion of individuals with genotype (i, j). The dynamics of the populations are defined by the following:

1. Each individual dies at rate 1.
2. When an individual dies, it is immediately replaced by an individual whose genotype is determined by selecting two individuals at random to act as parents, selecting a gene from each and allowing each gene to mutate randomly into a gene of another type.

From the dynamics, it is clear that the population size is constant. Consider such a population of size n. Let $X_n(t) \in S_0$ describe the genotypic proportions of the population at time t. For $x \in S$, let

$$q_n^i(x) = \frac{\sum_{j=1}^r \sigma_n^{ij} x^{ij}}{\sum_{j,k=1}^r \sigma_n^{jk} x^{jk}},$$

$$p_n^i(x) = \Big(1 - \sum_{j \neq i} \mu_n^{ij}\Big) q_n^i(x) + \sum_{j \neq i} \mu_n^{ji} q_n^j(x),$$

$$R_n^{ijkl}(x) = \theta_{kl} x^{kl} \theta_{ij} p_n^i(x) p_n^j(x),$$

where $\sigma_n = (\sigma_n^{ij})$ is the symmetric matrix of selection coefficients and $\mu_n = (\mu_n^{ij})$ is the matrix of mutation coefficients. We take $\sigma_n^{ij} > 0$ and $\mu_n^{ij} \geq 0$ with $\mu_n^{ii} = 0$. Then $q_n^i(x)$ is the probability of selecting a gene of type i from a population of type x; $p_n^i(x)$ is the probability that, after mutation, the selected gene is of type i and $nR_n^{ijkl}(x)$ is the rate at which type (k, l) individuals are replaced by type (i, j) individuals. Then we can write

$$\begin{aligned} X_n(t) &= X_n(0) + \sum_{\substack{i \leq j \\ k \leq l}} (e_{ij} - e_{kl}) \frac{1}{n} N^{ijkl}\Big(n \int_0^t R_n^{ijkl}(X_n)\, ds\Big) \\ &= X_n(0) + \sum_{\substack{i \leq j \\ k \leq l}} (e_{ij} - e_{kl}) \frac{1}{n} \tilde{N}^{ijkl}\Big(n \int_0^t R_n^{ijkl}(X_n)\, ds\Big) \\ &\quad + \int_0^t \big[p_n(X_n) p_n(X_n)^T - X_n\big]\, ds, \end{aligned} \tag{8.2}$$

where $\{N^{ijkl} | i \leq j,\ k \leq l\}$ is a collection of independent unit Poisson processes and $\tilde{N}^{ijkl}(t) = N^{ijkl}(t) - t$. Let $Y_n(t) = X_n(nt)$ and $W_n^{ijkl}(t) = (1/n)\tilde{N}^{ijkl}(n^2 t)$, so that $\{W_n^{ijkl} | i \leq j,\ k \leq l\}$ converge in distribution to independent Brownian motions. Then

$$\begin{aligned} Y_n(t) &= Y_n(0) + \sum_{\substack{i \leq j \\ k \leq l}} (e_{ij} - e_{kl}) W_n^{ijkl}\Big(\int_0^t R_n^{ijkl}(Y_n)\, ds\Big) \\ &\quad + n \int_0^t \big[p_n(Y_n) p_n(Y_n)^T - Y_n(s)\big]\, ds \\ &= Y_n(0) + U_n(t) + n \int_0^t F_n(Y_n(s))\, ds, \end{aligned}$$

where $F_n(x) = p_n(x)p_n(x)^T - x$ and $U_n(t)$ is the obvious sum. We seek a diffusion approximation for Y_n under the conditions $\mu_n \to 0$, $n\mu_n \to \mu$, $\sigma_n \to J$ (the matrix of all 1's) and $n(\sigma_n - J) \to \sigma$. Then $q_n \to p$, $p_n \to p$ and $F_n \to F$ uniformly on S_0, where $p(x) = x\mathbf{1}$ and $F(x) = xJx - x$. Moreover, $H_n \equiv n(F_n - F)$ converges uniformly on S_0 to some function H. Then

$$Y_n(t) = Y_n(0) + U_n(t) + \int_0^t H_n(X_n)\,ds + n\int_0^t F(X_n)\,ds.$$

Note that the argument of each W_n^{ijkl} in U_n is Lipschitz uniformly in n, so $\{U_n\}$ is relatively compact. It is straightforward to verify that $\{U_n\}$ satisfies Condition 4.2.

Note that S has dimension $d = r(r+1)/2 - 1$ and $\Gamma = \{x \in S | xJx = x\}$ has dimension $m = r - 1$. It is easy to see that $N = \{x \in \mathbb{M}(r,r) | x = x^T; x\mathbf{1} = 0\}$ has dimension $r(r-1)/2 = d - m$ and that for $x \in \Gamma$ and $y \in N$,

$$\partial F(x)(y) = xJy + yJx - y = -y.$$

Thus at least $d - m$ eigenvalues of $\partial F(x)$ are -1. Since m of its eigenvalues are 0, $\partial F(x)$ has $d - m$ eigenvalues in $D(0)$.

It is easy to verify that the continuous flow of F is $\psi(x,t) = (1 - e^{-t})xJx + e^{-t}x$ and $\Phi(x) = xJx = p(x)p(x)^T$ so Theorems 6.2 and 6.3 apply. Then limit points (Y, U) *of* $\{(Y_n + e^{-nt}(Y_n(0)JY_n(0) - Y_n(0)), U_n)\}$ must satisfy

$$Y(t) = Y(0) + \int_0^t (YJ\,dU + dU\,JY) + \int_0^t H(Y)\,ds + \int_0^t dU\,J\,dU,$$

$$U(t) = \sum_{\substack{i \le j\\ k \le l}} (e_{ij} - e_{kl})W^{ijkl}\left(\int_0^t \theta_{ij}Y^{ij}\theta_{kl}Y^{kl}\,ds\right),$$

where $\{W^{ijkl} | i \le j;\ k \le l\}$ is a collection of independent Brownian motions. Also, Y satisfies the so-called Hardy–Weinberg proportions, $Y(t)JY(t) = Y(t)$. At this point it is much easier to deal with $P(t) = Y(t)\mathbf{1}$, which satisfies

$$P(t) = P(0) + U(t)\mathbf{1} + \int_0^t h(P)\,ds,$$

where

$$h^i(p) = \sum_j \sigma^{ij}p^j p^i - (p^T\sigma p)p^i - \sum_j \mu^{ij}p^i + \sum_j \mu^{ji}p^j.$$

From this we can compute the generator of P to be

$$\frac{1}{2}\sum_{i,j=1}^{r} p^i\left(\delta_{ij} - p^j\right)\frac{\partial^2}{\partial p_i\,\partial p_j} + \sum_{i=1}^{r} h^i(p)\frac{\partial}{\partial p_i}. \tag{8.3}$$

Note that $Y = PP^T$, so this characterizes Y. Ethier (1976) proves uniqueness for a class of diffusions with generators of the form (8.3). Applying this result, we get convergence of Y_n to $Y = PP^T$, as outlined in the remarks following Theorem 6.1. Historically, this model was first proposed by Moran (1958). Watterson (1964) applied diffusion techniques to it and Ethier and Nagylaki

(1980) proved diffusion approximations for several population genetics models, including this one.

Wright–Fisher diploid model. To describe this model we use the same notation and the same assumptions on the selection and mutation coefficients as in the previous example. The difference between this model and the Moran diploid model is in the dynamics. The Wright–Fisher model is a discrete-time model in which the following hold:

1. Each individual lives for one unit of time, so the entire population is regenerated at each time step.
2. Individuals of a new generation have iid genotypes. The probability that a new individual has genotype (i, j) given that the previous generation had genotypic frequencies described by $x \in S_0$ is $\theta_{ij} p_n^i(x) p_n^j(x)$.

For each $n \ge 1$, let $\{\xi_n^{k,i} | k \ge 0,\ 1 \le i \le n\}$ be an iid set of random variables taking values in the space of measurable functions from S to $\{e_{ij} | 1 \le i \le j \le r\}$ such that

$$E\left[\xi_n^{k,i}(x)\right] = p_n(x) p_n(x)^T.$$

Note that this completely characterizes the distribution of $\xi_n^{k,i}(x)$ for each $x \in S$ (but not the distribution of $\xi_n^{k,i}$). Then X_n can be modelled as the solution of

$$X_n(\tau + 1) = \frac{1}{n} \sum_{i=1}^{n} \xi_n^{\tau,i}(X_n(\tau)).$$

Let

$$\zeta_n^k(x) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\xi_n^{k,i}(x) - p_n(x) p_n(x)^T\right)$$

and $F_n(x) = p_n(x) p_n(x)^T - x$. Then

$$X_n(\tau) = X_n(0) + \frac{1}{\sqrt{n}} \sum_{k=0}^{\tau-1} \zeta_n^k(X_n(k)) + \sum_{k=0}^{\tau-1} F_n(X_n(k)).$$

Letting $Y_n(t) = X_n(\lfloor nt \rfloor)$,

$$\begin{aligned} Y_n(t) &= Y_n(0) + U_n(t) + \int_0^t F_n(Y_n)\, dA_n \\ &= Y_n(0) + U_n(t) + \int_0^t H_n(Y_n)\, dB_n + \int_0^t F(Y_n)\, dA_n, \end{aligned}$$

where $A_n(t) = \lfloor nt \rfloor$, $B_n(t) = \lfloor nt \rfloor / n$ and

$$U_n(t) = \frac{1}{\sqrt{n}} \sum_{k=0}^{\lfloor nt \rfloor - 1} \zeta_n^k\left(Y_n\left(\frac{k}{n}\right)\right).$$

We need to show that $\{U_n | n \ge 1\}$ is relatively compact and satisfies Condition 4.2. To do this, it is convenient to view elements of S as vectors instead of matrices. We identify the space of symmetric $r \times r$ matrices with $\mathbb{R}^d$, $d = r(r+1)/2$, in such a way that the coordinates of x in $\mathbb{R}^d$ are $\{\theta_{ij} x^{ij} | 1 \le i \le j \le r\}$, the genotypic frequencies. Using the algebraic structure of $\mathbb{R}^d$, let

$$a_n(x) = E\left[\zeta_n(x)\zeta_n(x)^T\right],$$

so that $a_n(x) \in \mathbb{M}(d,d)$. Then $a_n(x)$ is symmetric and

$$a_n^{(ij),(kl)}(x) = \begin{cases} p_n^i(x)p_n^j(x)\left(1 - p_n^i(x)p_n^j(x)\right), & \text{for } (ij) = (kl), \\ -p_n^i(x)p_n^j(x)p_n^k(x)p_n^l(x), & \text{otherwise}. \end{cases}$$

Under the conditions set forth in the Moran model, $a_n \to a$ uniformly on S_0, where

$$a^{(ij),(kl)}(x) = \begin{cases} p^i(x)p^j(x)\left(1 - p^i(x)p^j(x)\right), & \text{for } (ij) = (kl), \\ -p^i(x)p^j(x)p^k(x)p^l(x), & \text{otherwise}. \end{cases}$$

Note that $\{\zeta_n^k | k \ge 0\}$ is an iid set of mean-0 random variables, so U_n is a martingale. Moreover,

$$\sup_{n \ge 1} \sup_{x \in S_0} E\left[|\zeta_n^k(x)|^4\right] < \infty,$$

hence $\{|\zeta_n^k(x)|^2 | n \ge 1, k \ge 0, x \in S_0\}$ is uniformly integrable. This implies that $\{[U_n](t) | n \ge 1\}$ is uniformly integrable for each $t \ge 0$, so $\{U_n\}$ satisfies Condition 4.2. Theorem 3.8.6(c) of Ethier and Kurtz (1986) provides relative compactness of $\{U_n\}$. As shown in the Moran model, F satisfies the eigenvalue condition of Theorem 7.2. Moreover, $x + F(x) \in \Gamma$ for $x \in S$, so $\Phi(x) = x + F(x) = xJx$. Thus, Theorems 7.2 and 7.3 apply.

Let (U, Y) be the limit point of (U_n, Y_n). We want to characterize U in terms of Y. Note that

$$M_n(t) = U_n(t)U_n(t)^T - \int_0^t a_n(Y_n)\, dB_n$$

is a martingale. Moreover, $\{M_n(t)\}$ is uniformly integrable for each $t \ge 0$. M_n converges (at least along a subsequence) to

$$M(t) = U(t)U(t)^T - \int_0^t a(Y)\, ds.$$

The uniform integrability implies that M is a martingale with respect to the filtration generated by U and Y. This provides us with the cross variations of U, which is a first step toward our goal. Next, we want to write U as the integral of a symmetric matrix $\beta(Y)$ against a Brownian motion. The fact that M is a martingale tells us that if this is possible, the matrix $\beta(Y)$ must be a square root of the matrix $a(Y)$. Note that $a(x)$ is symmetric, is continuous in x and has nonnegative eigenvalues. Then considering a diagonalization of $a(x)$ implies that there exist measurable symmetric $d \times d$ matrices $\beta(x)$, $b(x)$ and $c(x)$ such that $\beta(x)^2 = a(x)$, $\beta(x)$ has nonnegative eigenvalues, $\beta(x)b(x) + c(x) = I$, $\beta(x)c(x) = c(x)\beta(x) = 0$ and $b(x)c(x) = c(x)b(x) = 0$. Applying

Theorem 2 of Dunford and Schwartz [(1963), page 922] to the function $f(z) = \sqrt{(\mathrm{Re}\, z)^+}$ implies that β is continuous. Let Z be a d-dimensional Brownian motion which is independent of (Y, U), and let

$$W(t) = \int_0^t b(Y)\, dU + \int_0^t c(Y)\, dZ.$$

We write $[X, X](t)$ for the cross variation matrix of a martingale X. The fact that M is a martingale implies that $d[U, U] = a(Y)\, dt$. Then the definition of W and the independence of Z and (Y, U) imply

$$\begin{aligned} d[W, W] &= b(Y)d[U, U]b(Y)^T + c(Y)d[Z, Z]c(Y)^T \\ &= \big(b(Y)a(Y)b(Y) + c(Y)^2\big)\, dt. \end{aligned}$$

But $bab + c^2 = (b\beta + c)(b\beta + c)^T = I$. Thus, by Lévy's characterization [see Durrett (1984), page 78], W is a standard d-dimensional Brownian motion. Moreover,

$$\begin{aligned} \int_0^t \beta(Y)\, dW &= \int_0^t \beta(Y)b(Y)\, dU + \int_0^t \beta(Y)c(Y)\, dZ \\ &= \int_0^t (I - c(Y))\, dU = U(t) - \int_0^t c(Y)\, dU. \end{aligned}$$

But,

$$\begin{aligned} E\left[\left|\int_0^t c(Y)\, dU\right|^2\right] &= \mathrm{Tr}\, E\left[\int_0^t c(Y) d[U, U] c(Y)\right] \\ &= \mathrm{Tr} \int_0^t E[c(Y)a(Y)c(Y)]\, ds = 0. \end{aligned}$$

Thus $U = \int \beta(Y)\, dW$ and

$$\begin{aligned} Y(t) = Y(0) &+ \int_0^t \partial\Phi(Y)\beta(Y)\, dW + \int_0^t \partial\Phi(Y)H(Y)\, ds \\ &+ \tfrac{1}{2}\sum_{ij} \int_0^t \partial_{ij}\Phi(Y)a^{ij}(Y)\, ds. \end{aligned}$$

Returning to the original matrix notation and considering $P(t) = Y(t)\mathbf{1}$, we see that $Y(t) = P(t)P(t)^T$ and P has the same generator as in the Moran model. Thus we actually have convergence (in distribution) of (Y_n, U_n) to (Y, U), and the limit here is the same as for the Moran model.

Brownian motion on a manifold. Assume Γ is a C^2 submanifold of U of dimension $m < d$. Assume that for every x in some neighborhood $U_\Gamma \subset U$ of Γ there exists a unique $\Phi(x) \in \Gamma$ such that $|x - \Phi(x)| = d(x, \Gamma)$. Assume further that Φ is C^2 on U_Γ. Note that $|x - \Phi(x)|^2 = \min_y |x - \Phi(y)|^2$, so

$$\partial\Phi(x)^T(x - \Phi(x)) = 0. \tag{8.4}$$

Let

$$F(x) = -\tfrac{1}{2}\nabla|x - \Phi(x)|^2 = -\big(I - \partial\Phi(x)^T\big)(x - \Phi(x)) = \Phi(x) - x.$$

We can assume for $x \in U_\Gamma$ and $u \in [0,1]$ that $ux + (1-u)\Phi(x) \in U_\Gamma$ as well. Then

$$\Phi(ux + (1-u)\Phi(x)) = \Phi(x).$$

This implies that $F(ux + (1-u)\Phi(x)) = uF(x)$ and that

$$\psi(x,t) = e^{-t}x + (1 - e^{-t})\Phi(x)$$

solves

$$\psi(x,t) = x + \int_0^t F(\psi(x,s))\,ds.$$

Note that $\Phi(x) = \lim_{t\to\infty}\psi(x,t)$. Let X_n solve

$$X_n(t) = X_n(0) + W(t) + \alpha_n\int_0^t F(X_n)\,ds,$$

where W is d-dimensional Brownian motion and $\alpha_n \to \infty$. We need to verify the eigenvalue condition of Theorem 6.2. Note that $\Phi(\Phi(x)) = \Phi(x)$. Differentiating and restricting to Γ,

$$(\partial\Phi(y))^2 = \partial\Phi(y), \qquad y \in \Gamma.$$

Thus $\partial\Phi(y)$ is a projection matrix. Since the range of Φ is contained in Γ, the range of $\partial\Phi(y)$ is contained in $N(y)$, the tangent space of Γ at y. But Φ is the identity on Γ so the range of $\partial\Phi$ must be all of $N(y)$. Differentiating (8.4) and restricting to Γ,

$$\partial\Phi(y)^T(I - \partial\Phi(y)) = 0, \qquad y \in \Gamma.$$

This implies that $\partial\Phi(y)$ is symmetric, so $\partial\Phi(y)$ is orthogonal projection onto $N(y)$. Then $\partial F(y) = \partial\Phi(y) - I$ has $d - m$ of its eigenvalues equal to -1, the remaining m eigenvalues being 0. Assuming that $X_n(0) \Rightarrow X(0) \in \Gamma$, Theorem 6.2 applies and the limiting process satisfies

$$X(t) = X(0) + \int_0^t \partial\Phi(X)\,dW + \tfrac{1}{2}\int_0^t \Delta\Phi(X)\,ds,$$

where $\Delta\Phi$ is the Laplacian of Φ.

Let Y be Brownian motion on Γ, defined as the diffusion whose generator is the Laplace–Beltrami operator corresponding to the induced Riemannian structure. Since $\partial\Phi(y)$ is orthogonal projection onto $N(y)$ for $y \in \Gamma$, Y satisfies the Stratonovich equation [see Rogers and Williams (1987), pages 182–189],

$$Y(t) = Y(0) + \int_0^t \partial\Phi(Y)\circ dW.$$

The corresponding Itô equation is [see Rogers and Williams (1987), page 185]

$$\begin{aligned} Y(t) &= Y(0) + \int_0^t \partial\Phi(Y)\, dW + \tfrac{1}{2}[\partial\Phi(Y), W](t) \\ &= Y(0) + \int_0^t \partial\Phi(Y)\, dW + \tfrac{1}{2}\sum_{ij} \int_0^t \partial_{ij}\Phi(Y)\partial_j\Phi^i(Y)\, ds. \end{aligned}$$

The identity $\Phi(\Phi(x)) = \Phi(x)$ implies (by differentiating twice) that for $y \in \Gamma$,

$$\Delta\Phi(y) = \sum_{ijk} \partial_{ij}\Phi(y)\partial_k\Phi^i(y)\partial_k\Phi^j(y) + \partial\Phi(y)\,\Delta\Phi(y). \tag{8.5}$$

Using that $\partial\Phi(y)$ is symmetric and idempotent, this simplifies to (suppressing the argument)

$$\Delta\Phi = \sum_{ij} \partial_{ij}\Phi\,\partial_j\Phi^i + \partial\Phi\,\Delta\Phi.$$

Recall that $\Phi(ux + (1-u)\Phi(x)) = \Phi(x)$ for $u \in [0,1]$. Differentiating with respect to u and setting $u = 1$ yields

$$\partial\Phi(x)(x - \Phi(x)) = 0.$$

Differentiating this with respect to x produces

$$\partial_\alpha\Phi(x) + \sum_i \partial_{i\alpha}\Phi(x)(x^i - \Phi^i(x)) - \partial\Phi(x)\,\partial_\alpha\Phi(x) = 0.$$

Although Φ is not necessarily thrice differentiable, we can differentiate this at $y \in \Gamma$ to get (suppressing the argument)

$$2\,\partial_{\alpha\beta}\Phi - \sum_i \left[\partial_{i\alpha}\Phi\,\partial_\beta\Phi^i + \partial_{i\beta}\Phi\,\partial_\alpha\Phi^i\right] - \partial\Phi\,\partial_{\alpha\beta}\Phi = 0.$$

This implies that on Γ,

$$2\,\Delta\Phi = 2\sum_{ij} \partial_{ij}\Phi\,\partial_j\Phi^i + \partial\Phi\,\Delta\Phi.$$

Together with (8.5), this implies $\partial\Phi\,\Delta\Phi = 0$ and

$$\Delta\Phi = \sum_{ij} \partial_{ij}\Phi\,\partial_j\Phi^i \quad \text{on } \Gamma.$$

Thus X and Y satisfy the same Itô equation, implying that X is Brownian motion on Γ. In summary, when $F(x) = -\frac{1}{2}\nabla d(x,\Gamma)^2$, the limit is Brownian motion on Γ.

REFERENCES

DUNFORD, N. and SCHWARTZ, J. (1963). *Linear Operators, Part II: Spectral Theory*. Wiley, New York.

DURRETT, R. (1984). *Brownian Motion and Martingales in Analysis*. Wadsworth, Belmont, Calif.

ETHIER, S. N. (1976). A class of degenerate diffusion processes occurring in population genetics. *Comm. Pure Appl. Math.* **29** 483–493.

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

ETHIER, S. N. and NAGYLAKI, T. (1980). Diffusion approximations of Markov chains with two time scales and applications to population genetics. *Adv. in Appl. Probab.* **12** 14–49.

ETHIER, S. N. and NAGYLAKI, T. (1988). Diffusion approximations of Markov chains with two time scales and applications to population genetics, II. *Adv. in Appl. Probab.* **20** 525–545.

FALCONER, K. J. (1983). Differentiation of the limit mapping in a dynamical system. *J. London Math. Soc.* **27** 356–372.

JAKUBOWSKI, A., MÉMIN, J. and PAGÈS, G. (1989). Convergence en loi des suites d'intégrales stochastique sur l'espace D^1 de Skorokhod. *Probab. Theory Related Fields* **81** 111–137.

JOFFE, A. and MÉTIVIER, M. (1986). Weak convergence of sequences of semimartingales with applications to multitype branching processes. *Adv. in Appl. Probab.* **18** 20–65.

JOHNSON, D. P. (1983). Diffusion approximations for optimal filtering of jump processes and for queueing networks. Ph.D. dissertation, Univ. Wisconsin–Madison.

KATZENBERGER, G. S. (1990). Solutions of a stochastic differential equation forced onto a manifold by a large drift. Ph.D. dissertation, Univ. Wisconsin–Madison.

KURTZ, T. G. (1978). Diffusion approximations of branching processes. In *Advances in Probability* (A. Joffe and P. Ney, eds.) **5** 262–292. Dekker, New York.

KURTZ, T. G. and PROTTER, P. (1991). Weak limit theorems for stochastic integrals and stochastic differential equations. *Ann. Probab.* **19** 1035–1070.

MORAN. P. A. P. (1958). Random processes in genetics. *Proc. Cambridge Philos. Soc.* **54** 60–71.

PETERSON, W. P. (1985). Diffusion approximations for networks of queues with multiple customer types. Ph.D. dissertation, Stanford Univ.

PROTTER, P. (1990). *Stochastic Integration and Differential Equations*. Springer, New York.

REIMAN, M. I. (1983). Some diffusion approximations with state space collapse. In *Modelling and Performance Evaluation Methodology: Proceedings of the International Seminar* 209–240. Springer, New York.

ROGERS, L. C. G. and WILLIAMS, D. (1987). *Diffusions, Markov Processes, and Martingales* **2**: *Itô Calculus*. Wiley, New York.

WATTERSON, G. A. (1964). The application of diffusion theory to two population genetics models of Moran. *J. Appl. Probab.* **1** 233–246.

WENTZELL, A. D. and FREIDLIN, M. I. (1969). Small random perturbations of a dynamical system with a stable equilibrium position. *Soviet Math. Dokl.* **10** 886–890.

WENTZELL, A. D. and FREIDLIN, M. I. (1984). *Random Perturbations of Dynamical Systems*. Springer, New York.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF WISCONSIN
MADISON, WISCONSIN 53706