PROBABILITY LIMIT THEOREMS AND THE CONVERGENCE OF FINITE DIFFERENCE APPROXIMATIONS OF PARTIAL DIFFERENTIAL EQUATIONS+

BY

HAROLD J. KUSHNER

Center for Dynamical Systems
Division of Applied Mathematics
Brown University
Providence, Rhode Island

+This research was supported in part by the National Science Foundation under Grant No. GK 2788, in part by the National Aeronautics and Space Administration under Grant No. NGL 40-002-015 and in part by the Air Force Office of Scientific Research under Grant No. AF-AFOSR 67-0693A.

PROBABILITY LIMIT THEOREMS AND THE CONVERGENCE OF FINITE DIFFERENCE APPROXIMATIONS OF PARTIAL DIFFERENTIAL EQUATIONS

Harold J. Kushner

Introduction

The Equations to be Considered

This paper is concerned with the application of certain convergence theorems (for probability measures on spaces of continuous functions) to a problem in the convergence of finite difference approximations to partial differential equations.

Let G be a bounded open set in R^r (Euclidean r-space) with a continuous boundary ∂G, and let $k(\cdot)$ and $\varphi(\cdot)$ be non-negative continuous functions on R^r (and, occasionally, when the argument t appears, on R^{r+1}). Consider the possibly degenerate elliptic or parabolic equations of either of the forms (1) - (5).

$$\mathcal{L} = \sum_{i,j} a_{ij}(x)\frac{\partial^2}{\partial x_i \partial x_j} + \sum_i f_i(x)\frac{\partial}{\partial x_i}$$

$$\mathcal{L}V(x) = k(x),\ V(\partial G) = \varphi(\partial G) \tag{1}$$

$$\mathcal{L}V(x) - \beta V(x) = k(x),\ V(\partial G) = \varphi(\partial G) \tag{2}$$

$$V_t(x,t) + \mathcal{L}V(x,t) = k(x,t),\ V(x,t) = \varphi(x,T) \tag{3}$$
$$V(\partial G,t) = \varphi(\partial G,t),\ t < T.$$

Note that 'time' flows backward in (3); a simple transformation converts it into the more standard problem. One of the main results of the paper concerns the convergence of finite difference approximations to (1) - (3), as the difference interval goes to zero.

Probabilistic Interpretation

(1) - (3) can be given a probabilistic but physical interpretation. In fact, this 'physical' probabilistic interpretation will be used very heavily in the interpretation of the finite difference equations, in the motivation of the development, and in the convergence proofs. Let z_t be a vector of independent Wiener processes (thus $Ez_t z_t' = It$), and let x_t be the solution to the Ito stochastic differential equation (Doob [1], Chapter 6)

$$dx_t = f(x_t)dt + \sigma(x_t)dz_t, \tag{4}$$

where $f(\cdot)$ and $\sigma(\cdot)$ are bounded by a real number K and satisfy a uniform Lipschitz condition; e.g.,

$$|f(y) - f(x)| \le K|y-x|.$$

x_t can be defined to be continuous w.p.1. and satisfy the properties+ (Doob [1])

+$O(t)$ is of the order of t and $o(h)/h \to 0$ as $h \to 0$.

$$E \max_{t \geq s \geq 0} E|x_s - x_0|^2 = O(t)$$

$$E(x_h - x_0 | x_0) = f(x_0)h + o(h)$$

$$\text{cov}\,(x_h - x_0 | x_0) = \sigma(x_0)\sigma'(x_0)h + o(h).$$

Define the matrix $a(x)$ by $2a(x) = \sigma(x)\sigma'(x) = (\sigma_{ij}(x))$ and let τ be the random time at which the diffusion x_t first reaches the boundary ∂G, for $x_0 = x \in G$, and suppose that $E_x\tau < \infty$. Then, with E_x denoting the expectation given the initial condition $x_0 = x$, under certain conditions (1) - (3) have the unique solutions (1a) - (3a), resp. (Dynkin [2], Chapter 13).

$$V(x) = E_x \int_0^\tau k(x_s)ds + E_x\varphi(x_\tau) \tag{1a}$$

$$V(x) = E_x \int_0^\tau e^{-\beta s}k(x_s)ds + E_x e^{-\beta\tau}\varphi(x_\tau) \tag{2a}$$

$$V(x,t) = E_{x,t} \int_t^{T\cap\tau} k(x_s,s)ds + E_{x,t}\varphi(x_{T\cap\tau}, T\cap\tau) \tag{3a}$$

where we define $t \cap s = \min(t,s)$, and in (3a), $E_{x,t}$ implies that $x_t = x$.

Since we allow $\{a_{ij}(x)\}$ to be degenerate, by letting t be the $r + 1$ st coordinate of x, (3) becomes a special case of (1). Then the cylinder $G \times [0,T] = \tilde{G}$ replaces G in (1) and $T \cap \tau$, the finite escape time from $\tilde{G}$ replaces τ in (1). Thus, we will not treat (3) separately.

The conditions under which (1) - (3) are known to have solutions which are smooth enough to satisfy (1) - (3) (strong solutions) are quite restricted; in particular, full ellipticity of $\mathcal{L}$ is generally required. Yet, in rather typical situations, this condition is violated. This occurs almost all the time in stochastic control theory, where, in fact, one uses (1) - (3) to represent the cost functions (1a) - (3a), and hopes to solve (1) - (3) in order to obtain (1a) - (3a). For a particular case, consider the formal differential equation

$$y^{(r)} + c_{r-1}y^{(r-1)} + \dots + c_0 y = \sigma\xi \tag{5}$$

where ξ is 'white Gaussian noise'. Putting (5) into the form (4) yields

$$\begin{pmatrix} dx_1 \\ \vdots \\ dx_r \end{pmatrix} = dx = \begin{bmatrix} 0 & 0 & 1 & . & . & 0 \\ 0 & 0 & 0 & . & . & 1 \\ -c_0 & & . & . & . & -c_{n-1} \end{bmatrix} xdt + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ \sigma \end{bmatrix} dz$$

and $a_{ij} = 0$, unless $i = j = r$.

Of course (1) - (3) may be derived from other than stochastic considerations. Yet, still, unless we know that there is a solution to (1) - (3) with suitably smooth derivatives, and have an appropriate discrete maximum principle available, the usual methods (see, e.g. [3])

for proving convergence of the finite difference approximations, as the difference interval converges, do not apply. Nevertheless, for each difference interval, the finite difference equations approximating (1) - (3) may still have unique solutions, and it is meaningful to ask whether the solutions converge to (1a) - (3a), as the interval decreases to zero.

This problem will be treated by a probabilistic method. We also note that even if $\mathcal{L}$ is the Laplacian, classical proofs of convergence cannot be used if G has corners, since then the second derivatives may not be uniformly continuous in G.

Section 2 describes the finite difference equations to be used, and gives a useful probabilistic interpretation of them. The method of proof is described, and main theorems stated in Section 3.

The theorems use some general conditions which are quite common in applications. This is illustrated in the typical (degenerate elliptic) example of Section 4. Proofs of the theorems appear in the appendix. The treatment of the discounted problem (2) - (2a) is similar to that of (1) - (1a) and will not be given.

2. Finite Difference Equations and Markov Chains

Terminology

For equation (1) let the difference interval be h (in any coordinate direction⁺) and let e_i be the unit vector in the i^{th}

⁺This is for convience in development. The difference interval can certainly depend on the direction.

coordinate direction. Suppose that G is strictly contained in a hypercube H with sides $[-A,A]$. Define the set of nodes R_h^r in R^r by $R_h^r = \{(n_1h,\ldots,n_rh),\ n_i$ ranging over $0, \pm 1, \pm 2,\ldots\}$. Define $G_h = G \cap R_h^r$.

In order to expose the method, and not get involved with the rather long finite difference equations arising when mixed second derivatives occur, we let $a_{ij} = 0$ for $i \neq j$. There is no trouble in extending the method to the more general case.

<u>Form of the Finite Difference Equations</u>

The following finite difference approximations will be used.

$$V_{x_i} \sim \frac{1}{h}\left\{\begin{matrix} V(x+e_ih) - V(x) \\ \\ V(x) - V(x-e_ih) \end{matrix}\right\}, \tag{6a}$$

where the upper term of (6a) is used if $f_i(x) \geq 0$, and the lower otherwise. (This usage will be carried throughout, upper entries in $\{\ \}$ always used if $f_i \geq 0$, etc.)

$$V_{x_ix_i}(x) \sim [V(x+e_ih) - 2V(x) + V(x-e_ih)]/h^2. \tag{6b}$$

The reason for the choice (6a) will appear shortly.

If $V_h(x)$ denotes the solution to the finite difference equations, then using (6) for $x \in G_h$, (1) yields

$$0 = \sum_i \frac{a_{ii}}{h^2} [V_h(x+e_i h) - 2V_h(x) + V_h(x-e_i h)] + \sum_i \frac{f_i}{h} \begin{Bmatrix} V_h(x+e_i h) - V_h(x) \\ V_h(x) - V_h(x-e_i h) \end{Bmatrix} + k(x)$$

or, by collecting terms

$$V_h(x) = \sum_i \frac{V_h(x+e_i h)}{Q_h(x)} \begin{Bmatrix} h|f_i| + a_{ii} \\ a_{ii} \end{Bmatrix} + \sum_i \frac{V_h(x-e_i h)}{Q_h(x)} \begin{Bmatrix} a_{ii} \\ h|f_i| + a_{ii} \end{Bmatrix} \quad (1b)$$

where $Q_h(x) = 2\sum_i a_{ii} + \sum_i h|f_i|$. Define $V_h(x) = \varphi(x)$ for $x \in R_h^r - G_h$. Rewrite (1b) as (with the obvious identification of terms)

$$V_h(x) = \sum_i V_h(x+e_i h)p_h(x,x+e_i h) + \sum_i V_h(x-e_i h)p_h(x,x-e_i h) + p_h(x)k(x)$$

$$V(x) = \varphi(x), \quad \text{for } x \in R_h^r - G_h. \quad (1c)$$

Now the reason for the choice (6a) will become clear. Note that since the $p_h(x,y) \geq 0$ and sum to at most unity, and can be defined for all $x,y \in R_h^r$, they can be considered to be transition probabilities for a Markov chain on the grid R_h^r. This is the setup used in (Kushner, Kleinman [4]), where problems concerning the computation of solutions of <u>non-linear</u> versions of (1c) were considered.

Denote the sequence of random variables of this Markov chain by (ξ_k^h). Thus $P(\xi_{k+1}^h = \xi_k^h + he_i) = p_h(x,x+e_i h)$, etc. Define

$$N_h = \inf\{k: \xi_k^h \notin G_h\}.$$

Now we proceed to investigate the behavior as $h \to 0$. Suppose+ $E_\xi N_h < K_h < \infty$. The solution to (1c) can be written as [4]

$$V_h(x) = E_x \sum_{k=0}^{N_h-1} \rho_h(\xi_k^h)k(\xi_k^h) + E_x\varphi(\xi_{N_h}^h). \qquad \text{(1d)}$$

3. The Method

The probabilistic interpretation (1a) of (1), and the probabilistic interpretation (1d) of the finite difference system (1c), as well as the similarity of the form (1d) to a Reimann sum approximation to (1a), suggest that one could treat the convergence problem as a problem in the convergence (in a suitable sense) of the measures associated with $\{\xi_k^h\}$ to that of $\{x_t\}$. In fact this procedure is quite fruitful, and much of the sequel is devoted to setting the problem up so as to use the following theorem of Gikhman and Skorokhod [5], Chapter 9. (Actually, Theorem A is a composite of several theorems of [5], Chapter 9, Sections 1,2.)

Theorem A. *Let* $C[0,T] \equiv \Omega$ *be the set of* R^r *valued continuous functions on the interval* $[0,T]$. *Let* $y^n(t)$, $y(t)$, $t \in [0,T]$ *be continuous processes with paths in the (topological) space* Ω. *Let*

+This is not restrictive in applications (see [4]). In fact the condition is implied by condition (11) of Theorem 3 which is also natural in applications (see Example).

μ_n and μ be the measures induced on Ω by the processes $y^n(\cdot)$ and $y(\cdot)$, resp. Let (for $0 \le t' \le t'' \le T$)

$$\lim_{\delta \to 0} \overline{\lim_{n}} P_x \Big(\sup_{|t'-t''| \le \delta} |y^n(t') - y^n(t'')| \ge \epsilon > 0 \Big) = 0 \qquad (*)$$

for any $\epsilon > 0$. Let the finite dimensional distributions of $\{y^n(t)\}$ converge to these of $y(t)$. Let $F(\cdot)$ be a bounded and continuous (w.p.1.) functional on the topological space Ω. Then

$$EF(y^n(\cdot)) \to EF(y(\cdot)).$$

In the Appendix and section on convergence, Theorem A is exploited and extended to yield a solution (Theorem 3) to our problem. The example illustrates that the conditions of Theorem 3 are quite natural for a very large class of problems.

In order to exploit Theorem A, the process $\{\xi_k^h\}$ must be related to a suitable continuous time process $\{\xi^h(t)\}$.

By a comparison of (1d) and (1a), we note that the 'discrete time' cost rate is $\rho_h(x)$ times the continuous time cost. In an intuitive sense, one step of the discrete process ξ_k^h should take $\rho_h(\xi_k^h)$ units of real time. Thus the following definition is natural.

Define the time sequence (t_k^h) by[+] (sometimes arguments of functions are deleted for simplicity)

$$\Delta t_k^h \equiv \Delta t^h(\xi_k^h) \equiv \rho_h(\xi_k^h)$$

$$t_0 = t_0^h = 0$$

$$t_k^h = \sum_{0 \le s < k} \Delta t_s^h .$$

Define a process $\xi^h(t)$ by

$$\xi^h(t_k^h) = \xi_k^h$$

at the times (t_k^h), and for $t_k \le t < t_{k+1}$, by the linear interpolation

$$\xi^h(t) = \xi_{k+1}^h \frac{(t - t_k^h)}{\Delta t(\xi_k^h)} + \frac{\xi_k^h(t_{k+1}^h - t_k^h)}{\Delta t(\xi_k^h)} .$$

Thus the continuous process $\xi^h(t)$ is piecewise linear and changes slope at the <u>random</u> break points (t_k) only.

The use of $\xi^h(t)$ is a natural way of relating (ξ_k^h) and x_t. This can be seen from the last part of the following remark and from the calculations (8), which indicate that the drift and diffusion coefficients of $\xi^h(t)$ converge to those of the x_t process as $t \to 0$.

[+] We use $\rho_h(\xi_k^h)$, Δt_k^h and $\Delta t^h(\xi_k^h)$ interchangeably. Also, sometimes the arguments of f_i and a_{ii} are omitted.

Remark. To illustrate the random time scaling, consider the scalar example where

$$dx = -xdt + \sigma dz$$

and

$$\frac{\sigma^2}{2} V_{xx} - xV_x + k(x) = 0, \qquad V(A) = V(-A) = 0.$$

For $x = nh$, $n > 0$, the discrete equations reduce to

$$V_h(x) = V_h(x-h)\frac{[\sigma^2/2+xh]}{\sigma^2+xh} + V_h(x+h)\frac{\sigma^2/2}{\sigma^2+xh} + k(x)\frac{h^2}{\sigma^2+xh}$$

$$= V_h(x-h)p_h(x,x-h) + V_h(x+h)p_h(x,x+h) + \rho_h(x)k(x).$$

A simple discrete time (continuous state space) approximation to x_t is given by

$$\tilde{X}_{n+1} = \tilde{X}_n - \tilde{X}_n\Delta + \sigma[z_{(n+1)\Delta} - z_{n\Delta}], \tag{7}$$

and $E[\tilde{X}_r - X_t]^2 \to 0$ as $n \to \infty$, if $n\Delta$ remains fixed at t. However, while the time step, Δ, is constant, the one step jumps are unbounded: as $\tilde{X}_n$ increases the average step size increases, etc. If we are to bound the step size at each n (as we do with the process $\{\xi_n^h\}$ approximating X_t), we must restrict the time Δ at each n in some way which depends on the only known variable $\tilde{X}_n$.

This is clearly seen in the degenerate case $\sigma = 0$. Then if $x \neq 0$,

$$\rho_h(x) = \frac{h}{|x|} = \frac{h}{|\text{velocity}|} ,$$

which is exactly the time which it takes a particle to move the standard distance h, if the velocity were fixed at x during the time of movement.

Equations (2) and (3) can be treated similarly to (1). For example, applying (6) to (2), collecting terms, and dividing by the coefficient of V(x) yields, for $x \in G_h$,

$$V_h(x) = \sum_i \frac{V_h(x+e_{ih})}{Q_{\beta h}(x)} \left\{ \begin{matrix} h|f_i| + a_{ii} \\ a_{ii} \end{matrix} \right\} + \sum_i \frac{V_h(x-e_{ih})}{Q_{\beta h}(x)} \left\{ \begin{matrix} a_{ii} \\ h|f_i| + a_{ii} \end{matrix} \right\} \qquad (2b)$$

$$+ k(x)h^2/Q_{\beta h}(x)$$

where

$$Q_{\beta h}(x) = 2 \sum_i a_{ii} + h \sum_i |f_i| + h^2\beta.$$

(2b) can be rewritten as

$$\begin{aligned} V_h(x) = r_{\beta h}(x)[\sum_i V_h(x+e_{ih})p_h(x,x+e_{ih}) \\ + \sum_i V_h(x-e_{ih})p_h(x,x-e_{ih}) + \rho_h(x)k(x)], \end{aligned} \qquad (2c)$$

with boundary values defined as

$$V(x) = \varphi(x), \; x \in R_h^r - G_h,$$

where

$$\gamma_{\beta h}(x) = (1 - \frac{\beta h^2}{Q_{\beta h}(x)}) = (1 - \beta\rho(x) + O(h^3)).$$

The solution to (2c) is the discounted cost

$$V_h(x) = E_x \sum_{n=0}^{N_h - 1} (\prod_{i=0}^{n} \gamma_{\beta h}(\xi_i^h)) k(\xi_n^h)\rho_h(\xi_n^h) + E_x(\prod_{n=0}^{N_h} \gamma_{\beta h}(\xi_n^h))\varphi(\xi_{N_h}^h). \quad (2d)$$

The convergence of (2d) to (2a) can be discussed along the same lines as for equation (1), but will not be developed here.

<u>A Canonical Form For</u> (ξ_k^h).

Write the j^{th} component of ξ_k^h as $\xi_{k,r}^h$. With $\xi_k^h = \xi$, the transition probabilities $p(x, x \pm e_i h)$ given by (1b) yield

$$E[\xi_{k+1}^h - \xi \mid \xi_k^h = \xi] = \frac{h^2 f(\xi)}{Q_h(\xi)} = \rho_h(\xi) f(\xi). \quad (8a)$$

The average change in $\xi^h(t)$ in time $\Delta t^h(\xi)$ is merely the mean drift of the diffusion (4) times the time interval $\rho_h(\xi) = \Delta t^h(\xi)$, a further check of the naturalness of our time scaling.

Since the process ξ_k^h <u>moves in only one direction at a time,</u>

$[\xi^h_{k+1,i}-\xi^h_{k,i}]$ can be non-zero for only one i. Thus the off diagonal elements of the first matrix on the right of

$$\mathrm{Cov}\,[\xi^h_{k+1}-\xi|\xi^h_k=\xi]=\frac{E[(\xi^h_{k+1}-\xi)(\xi^h_{k+1}-\xi)'|\xi^h_k=\xi]}{\Delta t^h(\xi)}-$$

$$-\frac{E[(\xi^h_{k+1}-\xi)|\xi^h_k=\xi]E[(\xi^h_{k+1}-\xi)'|\xi^h_k=\xi]}{\Delta t^h(\xi)}=\Sigma_h(\xi)=(\Sigma_{h,ij}(\xi))\equiv$$

$$\equiv\sigma_h(\xi)\sigma'_h(\xi)$$

are zero. Thus

$$\Sigma_{h,ij}(\xi)=-\Delta t^h(\xi)f_i(\xi)f_j(\xi)=O(h^2)\quad\text{for } i\neq j \tag{8b}$$

while, for $i=j$ (and using $2a_{ii}=\sigma_i^2$)

$$\begin{aligned}\Sigma_{h,ij}&=2a_{ii}+[h|f_i|-\Delta t^h(\xi)f_i^2]\\&=\sigma_i^2+h[|f_i|-|f_i^2|\Delta t^h(\xi)]\\&=\sigma_i^2+O(h).\end{aligned}\tag{8c}$$

As a further check on the scaling of $\xi(t)$, observe the connection between the 'infinitesimal' properties of the (ξ^h_k) and (x_t) processes:

$$\lim_{h\to 0} \frac{E(\xi^h_{k+1}-\xi^h_k|\xi^h_k = x)}{\Delta t^h(x)} = \lim_{\delta\to 0} \frac{E_x(x_\delta - x)}{\delta}$$

$$\lim_{h\to 0} \frac{\mathrm{Cov}\ (\xi^h_{k+1}-\xi^h_k|\xi^h_k = x)}{\Delta t^h(x)} = \lim_{\delta\to 0} \frac{\mathrm{Cov}(x_\delta - x)}{\delta}.$$

Next, we may write ξ^h_{k+1} as (recall $\Delta t^h_k = \rho(\xi^h_k)$)

$$\xi^h_{k+1} = \xi^h_k + f(\xi^k_k)\Delta t^h_k + \beta^h_k$$

where

$$\beta_k = [\xi^h_{k+1}-\xi^h_k - f(\xi^h_k)\Delta t^h_k], \quad E\beta_k = 0.$$

Let $j < k$. Then $E[\beta'_k\beta_j|\xi_j,\xi_{j+1},\xi_k] = 0$. implies that (β_j) is an orthogonal sequence. We next give a convenient representation for the 'driving term' β_k.

There is an orthogonal sequence (ω^h_k) satisfying

$$E[\omega_k|\xi_0,\ldots,\xi_k] = 0 \quad \text{and}$$

$$E[\omega_k\omega'_k|\xi^h_0,\ldots,\xi^h_k] = I\Delta t^h_k$$

$$\beta_k = \Delta t^h_k \cdot \Sigma^{1/2}_h(\xi^h_k)\omega^h_k \qquad (*)$$

This is obvious if $\Sigma_h^{-1}(\xi_k)$ exists, since then define $\omega_k^h = \Delta t_k^h \Sigma_h^{-1/2} \beta_k$.

Next, letting $\Sigma_h^{-1}(\xi_k^h)$ not exist, we show (*) under (C1) - (C2) of Theorem 1, and for all small h. By (8b-c) and (C1) - (C2), $h|f_i(\xi)| \geq 5r\rho_h(\xi)f_i^2(\xi)$ for h small and all ξ. Then[+], from (8b-c) and (C1) - (C2), we conclude that semidefiniteness of $\Sigma_h(\xi_k^h)$ requires that some diagonal element be zero. Thus, $a_{ii}(\xi_k^h) = f_i(\xi_k^h)$ for some i. But then, the i^{th} row and column of Σ_h are zero, and so is $\beta_{k,i}$.

Then by reordering the states (at ξ_k^h) and repeating the argument, we may suppose that $\hat{\beta}_k \equiv (\beta_{k,1},\dots,\beta_{k,s})' = 0$, $\tilde{\beta}_k \equiv (\beta_{k,s+1},\dots,\beta_{k,r})$ has linearly independent components and that

$$\Sigma_h(\xi_k^h) = \begin{bmatrix} 0 & 0 \\ 0 & B_h(\xi_k^h) \end{bmatrix}, \quad \Delta t_k^h B_h(\xi_k^h) = E[\tilde{\beta}_k \tilde{\beta}_k' | \xi_0^h,\dots,\xi_k^h]$$

Finally, define $\omega_k^h = (\hat{\omega}_k^h, \tilde{\omega}_k^h)'$, where $\tilde{\omega}_k^h = B_h^{-1/2}(\xi_k^h)$. Let $\{\hat{\hat{\omega}}_k^h\}$ be an independent s-vector Gaussian sequence with mean zero and unit variance. Define $\hat{\omega}_k^h = \sqrt{\Delta t_k^h}\, \hat{\hat{\omega}}_k^h$. Thus the generality of (*) is proved.

Next, note that from $\xi_{k+1}^h = \xi_k^h + \Delta t_k^h f_i(\xi_k^h) + \Sigma_h^{1/2}(\xi_k^h)\omega_k^h$

[+] Σ_h is uniformly dominated by (say) 1/2 of its diagonal matrix, for small h.

we also have

$$E[\omega_k^h|\omega_0^h,\ldots,\omega_{k-1}^h,\xi_0] = 0$$

$$E[\omega_k^h(\omega_k^h)'|\omega_0^h,\ldots,\omega_{k-1}^h,\xi_0^h] = I\Delta t_k^h.$$

Convergence Theorems

Let measures μ_h and μ (on the topological space $\Omega = C[0,T]$ correspond to processes $\xi^h(t)$ and x_t, resp. $t \in [0,T]$. The conditions (C1) - (C4) used in the sequel are quite natural for a large class of problems, and are illustrated in the example.

Theorem 1. Assume

(C1) *$f_i(\xi)$ and $\sigma_{ij}(\xi)$ are uniformly bounded and satisfy a uniform Lipschitz condition.* (*Recall* $a = \sigma\sigma'$.)

(C2) *Let* δ_h *equal* h *or* h^2. *For real positive* K_i *let*

$$K_1\delta_h \geq \Delta t^h(\xi) \geq K_2\delta_h.$$

(C3) *Let* $a(\xi)$, *have the form*

$$a(\xi) = \begin{bmatrix} 0 & 0 \\ 0 & \Sigma_0(\xi) \end{bmatrix}, \ \sigma(\xi) = \begin{bmatrix} 0 & 0 \\ 0 & \sigma_0(\xi) \end{bmatrix}, \ \Sigma_0(\xi) = \sigma_0(\xi)\sigma'(\xi)$$

where $\Sigma_0^{-1}(\xi)$ has uniformly bounded terms.

Then the finite dimensional distributions of the process $\xi^h(t)$ converge to those of[+] the process x_t and, for $0 \le t' \le t'' \le T$, and $\epsilon > 0$,

$$\lim_{\delta \to 0} \overline{\lim_{h \to 0}} P_x\left(\sup_{|t'-t''| \le \delta} |\xi^h(t') - \xi^h(t'')| \ge \epsilon \right) = 0. \tag{9}$$

Remark. (C2) means the following. Either we allow $\sum_i a_{ii}(\xi) \ge \epsilon_0 > 0$ for some real ϵ_0, in which case $\delta_h = h^2$, or we allow $\sum_i a_{ii}(\xi) \equiv 0$ and $\sum_i |f_i| \ge \epsilon_0 > 0$, in which case $\delta_h = h$. Thus 2 cases are considered - one case in which there is always some diffusion somewhere, and one case in which there is no diffusion - but where the velocity of x_t is never 0. In the intermediate case the ratio

$$\frac{\max_{\xi} \Delta t^h(\xi)}{\min_{\xi} \Delta t^h(\xi)}$$

may be infinite, invalidating our proofs. The first case is one of great importance.

For future reference, we note that (C1) - (C2) and (8b-c)

[+] By such convergence, we always mean convergence at the points of continuity of the relevant distributions for the process x_t.

imply+ (for some real K)

$$\text{(C4)} \quad \sum_{k=0}^{TK_2/\delta_h} |\sigma_h(\xi_k^h) - \sigma(\xi_k^h)|^2 \delta_h \le KTh$$

and

$$\text{(C5)} \quad \sum_{k=0}^{TK_2/\delta_h} |\xi_{k+1,i}^h - \xi_{k,i}^h|^{2+\delta} \le KTh^\delta$$

for any $\delta > 0$ and $i = 1,\ldots,r$.

Corollary 1. Assume (C1) - (C3). Let $F(\cdot)$ be a bounded continuous function on $C[0,T]$ w.p.1. (relative to μ). Then++ (with $\xi_0^h = x = x_0$)

$$E_x F(\xi^h(\cdot)) \to E_x F(x(\cdot)).$$

Theorem 2 uses condition (C6). (C6). There is an $h_0 > 0$ so that for $h < h_0$, G satisfies: Let $\ell = (a,b)$ be a line connecting two adjacent (along coordinate directions) points (a,b) of the grid R_h^r. If a and b are both in G, then so is the line ℓ connecting them.

(C6) can be weakened in many ways - but there seems little point in complicating the condition here. It is certainly satisfied (for any

---

+ For a vector x, $|x|^2 = \sum_i x_i^2$. For a matrix σ, $|\sigma|^2 = \sum_i a_{ii}$, where $(a_{ij}) = a = \sigma'\sigma$. Recall that $\sigma_h(\xi)\sigma_h'(\xi) = \Sigma_h(\xi)$.

++ E_x is the expectation given $x_0 = x$.

h_0) for convex G. (C6) is used to assure that the first passage times from G, of both $\xi^h(t)$ and ξ^h_k are approximately the same time. I.e., if $\xi^h(t)$ leaves the G between the n^{th} and $n+1^{st}$ steps of ξ^h_k, then $N_h = n$. It is used to avoid the possibility illustrated by Figure 3 (for all small h), where if the discrete process ξ^h_k jumps from a to b at time n, it has not actually left G, but the interpolated process $\xi^h(t)$ leaves right after time n.

Theorem 2. Assume (C1) - (C3) and (C6). Let $k(\cdot)$ and $\varphi(\cdot)$ be uniformly continuous and bounded on some open set containing $\bar{G} = G + \partial G$. Let τ denote the first (random) time that the process x_t leaves G ($\tau = \inf\{t: x_t \notin G\}$), and suppose that $T \cap \tau \equiv \min(T,\tau)$ is continuous+ w.p.1. (The w.p.1. statement is relative to μ) on $C[0,T]$. Denote $\tau_h = \inf\{t: \xi^h(t) \notin G\}$. Then

$$E_x \int_0^{T\cap\tau} k(x_s)ds = \lim_{h\to 0} E_x \int_0^{T\cap\tau_h} k(\xi^h(s))ds$$

$$= \lim_{h\to 0} E_x \sum_{s=0}^{N_h-1} k(\xi^h_s)\rho_h(\xi^h_s) \qquad (10a)$$

$$E_x\varphi(x_{T\cap\tau}) = \lim_{h\to 0} E_x\varphi(\xi^h(T\cap\tau_h)) = \lim_{h\to 0} E_x\varphi(\xi^h_{N_h}). \qquad (10b)$$

+$T \cap \tau$ is a function of the path x.

Theorem 3. Assume the conditions of Theorem 2, and let, for some $t_0 < \infty$,

$$P_x(\xi_k^h \text{ leaves } G_h \text{ at least once by time } t_0) \geq M_0 > 0 \qquad (11)$$

where M_0 is independent of $x \in G_h$ and $h > 0$, for small h. Then the $V_h(x)$ given by (1d) converge to (1) as $h \to 0$, uniformly in x in G; i.e., the solutions of the finite difference equations (1b,c) converge to the weak solution (1a) of the equation (1), as $h \to 0$.

The arguments of the example yield Theorem 4, which generalizes Theorem 3 and does not contain (11) explicitly.

4. Example

The conditions imposed in Theorems 1-3 are rather natural for a large class of problems, and in order to illustrate this, their validity will be checked for a 2-dimensional problem. It should be clear that the example is typical of a large class. Although the basic problem arose in numerical analysis, the approach taken here as well as the conditions, are probabilistic. Hence, the checking of the conditions involves probabilistic calculations on the underlying processes. Let

$$dx_1 = f_1(x_2)dt \qquad a = \begin{bmatrix} 0 & 0 \\ 0 & v^2 \end{bmatrix} = \sigma\sigma'$$
$$dx_2 = f_2(x)dt + vdz \tag{12}$$

where v is a constant and the f_i satisfy (C1). Let $f_1(x_2) = x_2$ in $\bar{G}$. We seek to solve

$$\mathcal{L}V(x) + k(x) = 0 \quad \text{in} \quad G$$
$$V(x) = \varphi(x) \quad \text{on} \quad \partial G \tag{13}$$

where $k(\cdot)$ and $\varphi(\cdot)$ are continuous and bounded,

$$\mathcal{L} = \frac{v^2}{2} \frac{\partial^2}{\partial x_2^2} + f_1(x) \frac{\partial}{\partial x_1} + f_2(x) \frac{\partial}{\partial x_2},$$

and G is the box

$$G = \{x: |x_i| \le A\}.$$

Thus (C6) holds for all $h_0 > 0$.

Note that $\mathcal{L}$ is degenerate and G has corners; hence, classical theory cannot be used to solve the convergence problem for (13) as $h \to 0$.

Using (6) gives

$$V_h(x) = \frac{V_h(x+e_1h)}{Q_h(x)}\begin{vmatrix} h|f_1| \\ 0 \end{vmatrix} + \frac{V_h(x-e_1h)}{Q_h(x)}\begin{vmatrix} 0 \\ h|f_1| \end{vmatrix} +$$

$$+ \frac{V_h(x+e_2h)}{Q_h(x)}\begin{vmatrix} v^2/2 + h|f_2| \\ v^2/2 \end{vmatrix} + \frac{V_h(x-e_2h)}{Q_h(x)}\begin{vmatrix} v^2/2 \\ v^2/2 + h|f_2| \end{vmatrix} + \frac{k(x)h^2}{Q_h(x)}$$

for x on G_h, the grid in G; on the grid $R_h^r - G_h$ outside of G, define

$$V_h(x) = \varphi(x) \quad \text{for} \quad x \in R_h^r - G_h.$$

We need only show that $T \cap \tau$ is continuous w.p.1. (relative to μ) on $C[0,T]$, and that (11) holds. First, we prove the continuity condition. Let ω be a generic point of $C[0,T] = \Omega$. Thus we may write x_t, the value of the process at time t, more explicitly as $x_t(\omega)$.

In fact, $T \cap \tau$ is not continuous everywhere on Ω. To see why, let y_t be a <u>scalar</u> process and define $\tau(\omega) = \inf \{t: y_t(\omega) \geq \lambda\}$. Consider the path $y_t(\omega)$ of Figure 1. For any continuous sequence $\{g_n(\cdot)\}$ for which $y_t(\omega) \geq g_n(t) \uparrow y_t(\omega)$ uniformly on $[0,T]$, we have

$$\inf \{t: g_n(t) \geq \lambda\} \cap T = T.$$

Thus $(T \cap \tau)(\omega)$ is not continuous at the ω corresponding to the path $y_t(\omega)$ of Figure 1. However, it is continuous at ω'.

Returning to the problem (12), (13) refer to Figure 2. It is clear that if tangencies at the boundary occur only w.p. zero, then, by virtue of the continuity of $x_t(\omega)$ w.p.1., $(T \cap \tau)(\omega)$ will be continuous w.p.1. This will now be shown to be the case.

We observe that

(a) for $x_{2t} > 0$, x_{1t} must increase (as time increases) since $dx_1 = x_2 dt$ in G. Hence, w.p.1. points on the boundary section L_4 (Figure 2) are not accessible. Similarly for L_3.

(b) Also, since $x_{2t} > 0$ on L_1, the path cannot be tangent on L_1, and similarly for L_2.

(c) Owing to the dominant effects of the diffusion on movement in the vertical direction, the points on L_5 and L_6 are regular in the sense of Dynkin [2]; i.e.

$$\overline{\lim_{\delta \to 0}} P_x\{x_\tau \in \partial G, \; x_{\tau+\epsilon} \in \overline{G} \quad \text{for all} \quad \delta \geq \epsilon > 0\} = 0 \qquad (14)$$

(d) $P_x\{\tau = T\} = 0$

(e) $P_x\{x_\tau = q_1 \text{ or } q_2\} = 0.$

In fact, (a) - (e) imply the existence of $\epsilon_1(\omega) > 0$ w.p.1. so that, for $\tau(\omega) \leq T$, $\tau(\omega) + \epsilon_2(\omega) \leq T$ w.p.1. (relative to $\{\omega: \tau(\omega) \leq T\}$) and

$$\text{distance}\,(x_{\tau(\omega)+\epsilon_2(\omega)}, G) \geq \epsilon_1(\omega).$$

(f) Since x_t is continuous, w.p.1., there are $\epsilon_3(\omega) > 0$, w.p.1., $\epsilon_4(\omega) > 0$ w.p.1. so that

$$\text{distance}\,(x_{t-\epsilon_3(\omega)}, \text{exterior of } \bar{G}) \geq \epsilon_4(\omega),$$

for all $t \leq \tau(\omega)$ if $\tau(\omega) \leq T$.

Now, denote by N_0 the sum of the exceptional null sets in (a) - (f). Let $\omega \in \Omega - N_0$ and $\tau(\omega) \leq T$. Let $\{g_n(t)\}$ in $C[0,T]$ satisfy (as $n \to \infty$)

$$\sup_{0 \leq t \leq T} |g_n(t) - x_t(\omega)| \to 0.$$

Let $\epsilon > 0$ be arbitrary. Then by (a) - (f), for large n, the first time $g_n(t)$ leaves G must be within ϵ of $\tau(\omega)$. This proves the continuity w.p.1. of $\tau(\omega) \cap T = (\tau \cap T)(\omega)$.

Only (11) remains to be proved. Let $N = t_0/K_1h^2$, for any $t_0 > 0$, and define

$$M_+(\xi) = \text{number of positive steps of } \xi^h_{k,2},\ k \leq N$$

$$M_-(\xi) = \text{number of negative steps of } \xi^h_{k,2},\ k \leq N.$$

A sufficient condition for (11) is

$$q_h(x) \equiv P_x(M_+(\xi) - M_-(\xi) \geq \frac{2A}{h}) \geq M_0 > 0. \qquad (15)$$

For some real K, we have the bounds,

$$\frac{1}{2} - Kh \leq P_x(\xi^h_{k+1,2} - \xi^h_{k,2} = h) \leq \frac{1}{2} + Kh.$$

Let (u^h_k) be a Markov process on $(0, \pm 1, \pm 2, \ldots)$ with transition probability

$$P(u^h_{k+1} = u^h_k + 1) = \frac{1}{2} - Kh = 1 - P(u^h_{k+1} = u^h_k - 1).$$

Define $M_\pm(u)$ analogously to $M_\pm(\xi)$. Then

$$q_h(x) \geq q_{h,u}(x) \equiv P(M_+(u) - M_-(u) \geq 2A/h).$$

The mean value of $u^h_{k+1} - u^h_k$ is $-2Kh$, and its variance is $1 - (2kh)^2$. Now

$$q_{h,u}(x) = P\left\{\frac{M_+(u)-M_-(u)-N(-2kh)}{\sqrt{N}\sqrt{1-(2Kh)^2}} \geq \frac{2A/h+2Nkh}{\sqrt{N}\sqrt{1-(2Kh)^2}}\right\}.$$

The left term in brackets converges in distribution to the normal zero mean and unity variance random variable, and the right hand term in the brackets is strictly less than some $K_3 < \infty$ for small h. Thus, for all small h

$$q_h(x) \geq q_{h,u}(x) \geq \frac{1}{2\pi} \int_{K_5}^{\infty} \exp - \frac{1}{2} y^2 dy,$$

which proves (11).

The crucial step in the proof of (11), the bounding of the drift in one direction, can easily be generalized. In fact, (11) holds if

$$a = \begin{bmatrix} 0 & 0 \\ 0 & \Sigma_0 \end{bmatrix}$$

where Σ_0^{-1} has uniformly bounded terms in G.

In fact, the example can be generalized to yield (The proof is a combination of the arguments of the example and of Theorem 3 and is omitted.)

<u>Theorem</u> 4. <u>Assume</u> (C1) - (C3) <u>and</u> (C6). <u>Let</u> $\partial G = B_1 + B_2$, <u>where points on</u> B_1 <u>are regular for the process</u> x_t <u>in the sense</u> (14) <u>of Dynkin</u> [2], <u>and on</u> B_2 <u>are inaccessable</u> w.p.1. <u>for the process</u> x_t. <u>If</u> $k(\cdot)$ <u>and</u> $\varphi(\cdot)$ <u>are continuous on a neighborhood containing</u> $\bar{G}$, <u>then</u> $V_h(x) \to V(x)$.

APPENDIX

The proof of Theorem 1 will be developed via a series of lemmas. The symbols K or K_i are used for any constant; values may change from usage to usage.

Lemma 1. Assume (C1) - (C5). Define the process (η_k^h) with initial condition $\eta_0^h = \xi_0 = \xi_0^h$ and

$$\eta_{k+1}^h = \eta_k^h + f(\eta_k^h)\Delta t_k^h + \sigma(\eta_k^h)\omega_k^h$$

where $\Delta t_k^h = \Delta t^h(\xi_k^h)$. Let $N = T/K_2\delta_h$

$$\sup_{N \geq k \geq 0} E|\eta_k^h - \xi_k^h|^2 \to 0 \quad \text{(A1)}$$

as $h \to 0$. Define $\eta^h(t)$ as the linear interpolation of η_k^h, i.e., for $t_k^h \leq t < t_{k+1}^h$,

$$\eta^h(t) = \eta^h(t_{k+1})\frac{(t-t_k^h)}{\Delta t_k^h} + \eta^h(t_k)\frac{(t_{k+1}^h-t)}{\Delta t_k^h}.$$

Then if the multidimensional distributions of the process $\eta^h(t)$ have limits, so do those corresponding to the limit of the $\xi^h(t)$ process and they are the same. (Note that the η_k^h and ξ_k^h equations differ only in that $\Sigma_h^{1/2}$ is replaced by its limit σ.)

Proof. Let $\mathcal{F}_{h,k}$ be the minimal σ-algebra over which

$(\omega_i^h,\ i \le k)$ is measurable. For simplicity of writing, we drop the index h on η_i^h, ξ_i^h, Δt_i^h, ω_i^h, $\mathscr{F}_{h,k}$, etc. whenever no confusion will arise. Both ξ_s and η_s, $s \le k$, are $\mathscr{F}_k$ measurable, and we use the notation $\Delta t_k = \Delta t(\xi_k)$. From

$$\eta_{k+1} - \xi_{k+1} = \eta_k - \xi_k + [f(\eta_k) - f(\xi_k)]\Delta t_k + [\sigma(\eta_k) - \sigma_h(\xi_k)]\omega_k,$$

we can write

$$\begin{aligned} M_{k+1} \equiv E|\eta_{k+1} - \xi_{k+1}|^2 &= (M_k + 2E(\eta_k - \xi_k)'[f(\eta_k) - f(\xi_k)]\Delta t_k) \\ &+ E|f(\eta_k) - f(\xi_k)|^2 \Delta t_k^2 + E\omega_k'[\sigma(\eta_k) - \sigma(\xi_k)]'[\sigma(\eta_k) - \sigma(\xi_k)]\omega_k \\ &+ E([\omega_k'[\sigma(\eta_k) - \sigma(\xi_k)]'[\sigma(\xi_k) - \sigma_h(\xi_k)]\omega_k) \\ &+ E\omega_k'[\sigma(\xi_k) - \sigma_h(\xi_k)]'[\sigma(\xi_k) - \sigma_h(\xi_k)]\omega_k \\ &\equiv A + B + C + D + E. \end{aligned}$$

Using the Lipschitz condition (C1), and (C2), yields, for some real K,

$$\begin{aligned} A &\le M_k(1+2K\delta_h) \\ B &\le KM_k\delta_h^2 \\ C &\le KM_k\delta_h \\ |D| &\le K\delta_h M_k + K\delta_h|\sigma(\xi_k) - \sigma_h(\xi_k)|^2 \\ |E| &\le K\delta_h|\sigma(\xi_k) - \sigma_h(\xi_k)|^2. \end{aligned}$$

Thus

$$M_{k+1} \le M_k(1+K_3\delta_h+K_4\delta_h^2) + K_5E|\sigma(\xi_k) - \sigma_h(\xi_k)|^2\delta_h$$

and

$$M_k \le K_7(1+K_6\delta_h)^N|\sigma(\xi_k) - \sigma_h(\xi_k)|^2\delta_h \cdot \sum_{j=0}^{N} |\sigma(\xi_j) - \sigma_h(\xi_j)|^2\delta_h.$$

Then (A1) follows from (C4) and

$$(1+K_6\delta_h)^N = (1+K_6\delta_h)^{T/K_2\delta_h} \le \exp K_6T/K_2.$$

Finally, since the number of terms of the process $\{\xi_k^h\}$ which effect the $\xi^h(t)$ process on $[0,T]$ is at most $T/K_2\delta_h$ and at least $T/K_1\delta_h$, and since $M_k \to 0$ uniformly as $h \to 0$, for $k \le N$, we have

$$E|\xi^h(t) - \eta^h(t)|^2 \to 0$$

on $[0,T]$, as $h \to 0$. Q.E.D.

<u>Lemma</u> 2. <u>Assume</u> (C1) - (C3). <u>Let</u> $N = T/K_2\delta_h$, $n = t/K_2\delta_h$. <u>Then for</u> $n \le N$,

$$E \max_{n\geq k\geq 0} |\xi_k^h|^2 \leq K(1+|\xi_0^h|^2)e^{Kt}$$

$$E \max_{n\geq k\geq 0} |\xi_k^h - \xi_0^k|^2 \leq Kt(1+|\xi_0^h|^2),$$

where K is a real number. The same result holds for the (η_k^h) process.

Proof. Again drop the index h on ξ_i^h, η_i^h, Δt_i^h, ω_i^h, etc., where convenient. Then

$$\xi_{k+1} = \xi_0 + A_k + B_k$$
$$A_k = \sum_0^k f(\xi_i)\Delta t_i, \quad B_k = \sum_0^k \sigma_h(\xi_i)\omega_i.$$

By (8b-c), the σ_h also satisfy a uniform Lipschitz condition (also uniform in h for small h) and are bounded uniformly in h for small h). Write $Y_k = \max_{k\geq i} |\xi_i'\xi_i|$ and $M_k = EY_k$.

Now

$$|\xi_{k+1}|^2 \leq K(|\xi_0|^2 + |A_k|^2 + |B_k|^2) \quad \text{(A2)}$$

and

$$\max_{n\geq k>0} |A_k|^2 \leq \sum_0^k |f(\xi_i)\Delta t_i| \leq K \sum_0^k (1+|\xi_i|^2)\delta_h. \quad \text{(A3)}$$

Further, B_k is a martingale and (Doob [1], Chapter 7)

$$E \max_{n \ge k \ge 0} |B_k|^2 \le KE|B_n|^2 \le KE \sum_0^n |\sigma_h(\xi_i)|^2 \delta_h$$
$$\le KE \sum_0^n (1+|\xi_i|^2)\delta_h. \quad (A4)$$

Combining (A2) - (A4), taking expectations, and replacing $E|\xi_i|^2$ by the majorant M_i, yields

$$M_{n+1} \le KM_0 + K(n\delta_h + n\delta_h^2 + \sum_0^n M_k \delta_h), \quad (A5)$$

which is bounded above by the expression given in the lemma for $E \max_{n+1 \ge k \ge 0} |\xi_k^h|^2$.

The proof of the other statements of the lemma are similar and are omitted. Q.E.D.

We next compare $\eta^h(t)$ to a process whose distributions are easier to relate to those of x_t.

Divide $[0,T]$ into intervals with endpoints $0, \Delta, 2\Delta, \ldots, N_\Delta \Delta$ where $\Delta \gg K_1 \delta_h$. Recall the definition $t_i^h = \sum_{s=0}^{i-1} \Delta t_s^h$. Define $n_0^h = 0$ and the (random) integer n_i^h by $n_i^h = \max\,(n\colon t_n^h \le i\Delta)$. Define the sets of integers

$$I_{i+1}^h = (r\colon t_{n_i}^h \le t_r^h < t_{n_{i+1}}^h).$$

Again, we drop the index h where convenient. Then (using n_i for n_i^h, etc.)

$$\Delta - 2K_1\delta_h \le t_{n_{i+1}} - t_{n_i} \le \Delta + 2K_1\delta_h$$

and

$$\sum_{r\in I_i} \Delta t_r = t_{n_{i+1}} - t_{n_i}.$$

$$\text{Cov} \sum_{r\in I_i} \omega_s = I\!E(t_{n_{i+1}} - t_{n_i}). \tag{A6}$$

Let $\tilde{y}_0 = \eta_0$, and for each h define the process

$$\tilde{y}_{k+1} = \tilde{y}_k + f(\tilde{y}_k) \sum_{s\in I_{k+1}} \Delta t_s + \sigma(\tilde{y}_k) \sum_{s\in I_{k+1}} \omega_s. \tag{A7}$$

<u>Lemma</u> 3. (<u>Again, omit index h where convenient</u>). <u>Assume</u> (C1) - (C3). <u>Then</u>

$$\lim_{h\to 0} \sup_{\frac{T}{\Delta}\ge k\ge 0} E|\tilde{y}_k - \eta_k|^2 = 0. \tag{A8}$$

<u>Let</u> $\tilde{y}(t)$ <u>be the linearly interpolated process with</u> $\tilde{y}(t_{n_i}) = \tilde{y}_i$. <u>Then the multidimensional distributions of</u> $\eta^h(t)$ <u>tend</u> (<u>as</u> $h \to 0$) <u>to the limit</u> (<u>as</u> $h \to 0$, <u>then</u> $\Delta \to 0$) <u>of those for the process</u> $\tilde{y}(t)$.

<u>Proof</u>. Let $\epsilon_k = \tilde{y}_k - \eta_{n_k}$. Then

$$\epsilon_{k+1} = \epsilon_k + \sum_{s\in I_{k+1}} [f(\tilde{y}_k) - f(\eta_s)]\Delta t_s + \sum_{s\in I_{k+1}} [\sigma(\tilde{y}_k) - \sigma(\eta_s)]\omega_s .$$

$$E|\epsilon_{k+1}|^2 \equiv M_{k+1} = M_k + A + B + C$$

where (<u>unindexed sums are over</u> $s \in I_{k+1}$)

$$\begin{aligned} A &= E|\sum [f(\tilde{y}_k) - f(\eta_s)]\Delta t_s|^2 \\ &\le E(\sum \Delta t_s)E \sum_{s=0}^{\Delta/K_2\delta_h} |f(\tilde{y}_k) - f(\eta_{n_k+s})|\Delta t_s \\ &\le K\Delta\cdot \sum_{s=0}^{\Delta/K_2\delta_h} (E|y_k-\eta_{n_k}|^2 + E|\eta_{n_k+s}-\eta_{n_k}|^2)\delta_h \\ &\le K(\Delta)^2 M_k + K\Delta^3, \end{aligned}$$

where Lemma 2 is used in the last step.

$$\begin{aligned} B &= E|\sum [\sigma(\tilde{y}_k) - \sigma(\eta_s)]\omega_s|^2 \\ &= E \sum|\sigma(\tilde{y}_k) - \sigma(\eta_s)|^2\Delta t_s \\ &= K \sum_{0}^{\Delta/K_2\delta_h} (E|\tilde{y}_k-\eta_{n_k}|^2 + E|\eta_{n_k+s}-\eta_{n_k}|^2)\delta_h \\ &\le K\Delta M_k + K\Delta^2, \end{aligned}$$

where, again, Lemma 2 is used in the last step.

$$C = E\, 2\epsilon_k' \sum [f(\tilde{y}_k) - f(\eta_s)]\Delta t_s$$

$$|C| \le KE^{1/2}|\epsilon_k|^2 E^{1/2}|\sum (f(\tilde{y}_k) - f(\eta_s))\Delta t_s|$$

$$\le KM_k^{1/2}\Big[\sum_{s=0}^{\Delta/K_2\delta_h} \delta_h \sum_{s=0}^{\Delta/K_2\delta_h} E|\tilde{y}_k - \eta_{s+n_k}|^2 \delta_h\Big]^{1/2}$$

$$\le KM_k^{1/2}[M_k\Delta^2 + K\Delta^3]^{1/2}$$

$$\le K\Delta M_k + K\Delta^{3/2},$$

where the second step used the bound

$$E|\tilde{y}_k - \eta_{s+n_k}|^2 \le KE|\tilde{y}_k - \eta_{n_k}|^2 + KE|\eta_{s+n_k} - \eta_{n_k}|^2$$
$$\le KM_k + K\Delta$$

and where the last step used a bound $EM_k < K_1$ for $k \le T/\Delta$ which is derivable by the method of either Lemmas 1 or 2.

Thus

$$M_{k+1} \le M_k(1+K_2\Delta) + K_3\Delta^{3/2}, \quad M_0 = 0,$$

which implies

$$M_k \le K_4(T)\Delta^{1/2}, \quad k \le T/\Delta \tag{A9}$$

where K_4 depends only on T and the constants in (C1) - (C2).

(A8) and the last statement of the lemma follow from (A9) and Lemma 2. Q.E.D.

Lemma 4. *Assume* (C1) *to* (C3). *Define*

$$y_{k+1} = y_k + f(y_k)\Delta - \sigma(y_k)\delta z_k$$

where $\delta z_k = z_{k\Delta+\Delta} - z_{k\Delta}$, *where* z_t *is a vector Wiener process* ($Ez_t z_t' = It$). *Then the distributions of* $(\tilde{y}_k)$ *converge to those of* (y_k) *as* $h \to 0$ *(for fixed* Δ).

Proof. A proof can be easily modelled along the lines of the proof of part of Theorem 1, p. 595 [5], and we only sketch the outline. Write $\omega_k = \begin{pmatrix} \omega_k \\ \tilde{\omega}_k \end{pmatrix}$, where $\tilde{\omega}_k$ has the dimension of the $\sigma_0(x)$ in (C3). Since $\sigma(x) = \begin{bmatrix} 0 & 0 \\ 0 & \sigma_0(x) \end{bmatrix}$, only the sequence $(\tilde{\omega}_k)$ enters into the definition of the sequence $(\tilde{y}_k)$, and we can write

$$\tilde{y}_{k+1} = \tilde{y}_k + f(\tilde{y}_k)(t_{n_{k+1}} - t_{n_k}) + \sigma(\tilde{y}_k)u_k \qquad \text{(A10)}$$

where

$$\tilde{\sigma}(x) = \begin{bmatrix} 0 \\ \sigma_0(x) \end{bmatrix}$$

and

$$u_k = \sum_{s \in I_{k+1}} \tilde{\omega}_s .$$

Let σ_0 have dimension r_0. Suppose that, for all uniformly bounded (by, say, K) r_0-vectors (λ_k), we have

$$\lim_{h \to 0} E \exp \sum_k i\lambda_k' u_k = \exp - \frac{1}{2} \sum_k \lambda_k' \lambda_k \Delta. \qquad \text{(A11)}$$

Then the multidimensional distributions of (u_k) converge to those of $((z_{k\Delta+\Delta} - z_{k\Delta}))$.

Gikhman and Skorokhod used the property (A11) to prove (in a relatively straightforward way) that the multidimensional distributions of $(\tilde{y}_k)$ converge to those of (y_k). (See Lemmas 2.3, p. 599-601 [5], and the proof concerning convergence in distribution of $(\eta_n^*(\tau_k))$ to $(\eta_0^*(\tau_k))$ on p. 601-602.) Of course, in [5], the η_n^* is scalar process (the vector extension is straightforward) and the coefficient of $f(\tilde{y}_k)$ is Δ. But, since $|t_{n_{k+1}} - t_{n_k}| \to \Delta$ uniformly in all variables as $h \to 0$, the proof can easily be modified to account for this minor difference.

Thus we will only prove (A11). First, let us introduce some notation. Let $T/\Delta = m$, and divide each interval $[i\Delta, i\Delta+\Delta]$ $(i=0,\dots,m-1)$ into subintervals of length b_h, where $b_h \to 0$ as $h \to 0$ and $b_h/\delta_h \to \infty$. Define the sets of indices (subsets of I_ℓ) $I_{\ell r}$

by

$$I_{\ell r} = \{s: t_s \in [(\ell-1)\Delta + (r-1)b_h, \ (\ell-1)\Delta + rb_h)\}$$

$\ell = 1,\ldots,m$, $r = 1,\ldots,\Delta/b_h$. Define

$$u_{\ell r} = \sum_{s \in I_{\ell r}} \omega_s$$

Let $u_{\ell r,i}$ and $\omega_{s,i}$ be the i^{th} components, resp., where $i = r - r_0 + 1,\ldots,r$ (the last r_0 components). (Note that $u_\ell = \sum_r u_{\ell r}$.)

Next, we show $\lim_{h \to 0} S_{h,i} = 0$ where

$$S_{h,i} = \sum_{\ell=1}^{m} \sum_{r=1}^{\Delta/b_h} |u_{\ell r,i}|^{2+\delta} \tag{A12}$$

for some $1 > \delta > 0$. In fact,

$$S_{h,i} \le \frac{T}{b_h} \max_{\ell,r} |u_{\ell r,i}|^{2+\delta}$$

and

$$|u_{\ell r,i}|^{2+\delta} = |\sum_{s \in I_{\ell r}} \omega_{s,i}|^{2+\delta} \le \frac{b_h}{K_2 \delta_h} \max_{s \in I_{\ell r}} |\omega_{s,i}|^{2+\delta}$$

$$\le K b_h \cdot h^{2+\delta} / K_2 \delta_h,$$

where $|\omega_{s,i}| \le Kh$ is used, $i = r - r_0 + 1, \ldots, r$. The inequality $|\omega_{s,i}| \le Kh$, $i = r - r_0 + 1, \ldots, r$ follows from

$$\tilde{\omega}_k = [\sigma_0(\xi_k) + O(h^{1/2})]^{-1}[\tilde{\xi}_{k+1} - \tilde{\xi}_k - \tilde{f}(\xi_k)\Delta t_k + O(h)]$$

which, in turn, follows from (8b,c) and (C1) - (C3). The $O(\cdot)$ terms are uniform in ξ_k, and $\tilde{\xi}, \tilde{f}$ are the last r_0 components of ξ, f, resp. Then

$$S_{h,i} \le K \frac{T}{b_h} \cdot \frac{b_h \cdot h^{2+\delta}}{\delta_h} \to 0.$$

Let $\mathcal{F}(m-1)$ denote the least σ-algebra measuring all $u_1, \ldots, u_{m-1}$ and let $\mathcal{F}(m-1,t)$ be the least σ-algebra measuring, in addition, $u_{m1}, \ldots, u_{mt}$. Let $\frac{\Delta}{b_h} = v$ and recall $m = T/\Delta$. Thus $\mathcal{F}(m-1,v-1)$ measures all the $u_{\ell r}$ except for the last one u_{mv}.

Now,

$$E[\exp i\lambda_m' u_m | \mathcal{F}(m-1)]$$

$$= E[\exp i\lambda_m' \sum_1^{v-1} u_{ms} | \mathcal{F}(m-1)] E[\exp i\lambda_m' u_{mv} | \mathcal{F}(m-1,v-1)]$$

$$= A \cdot B.$$

$$B = E[(1 + i\lambda_m' u_{mv} - \frac{(\lambda_m' u_{mv})^2}{2} + \tilde{\epsilon}_{mv}) | \mathcal{F}(m-1,v-1)]$$

where

$$|\tilde{\epsilon}_{mv}| \le K|\lambda'_m u_{mv}|^{2+\delta}.$$

Since $E[u_{mv} | \mathscr{F}(m-1,v-1)] = 0$, and

$$b_h \lambda'_m \lambda_m - K\delta_h \le E[(\lambda'_m u_{mv})^2 | \mathscr{F}(m-1,v-1)] \le b_h \lambda'_m \lambda_m + K\delta_h$$

we have

$$(1 - \frac{\lambda'_m \lambda_m b_h}{2} - M_{mv}) \le B \le (1 - \frac{\lambda'_m \lambda_m b_h}{2} + M_{mv})$$

where M_{mv} is a real number

$$M_{mv} = K\delta_h + K \max_{m,v} |\lambda'_m u_{mv}|^{2+\delta}.$$

Thus

$$AB = A(1 - \frac{\lambda'_m \lambda_m b_h}{2}) + \tilde{M}_{mv}$$

where

$$|\tilde{M}_{mv}| \le M_{mv}.$$

Continuing the procedure gives

$$AB = \left(1 - \frac{\lambda_m'\lambda_m b_h}{2}\right)^v + \tilde{M}_m$$

$$|\tilde{M}_m| \le \sum_{s=1}^{v} M_{ms}.$$

Similarly,

$$\exp i \sum_{k=1}^{m} \lambda_k' u_k = \prod_{\ell=1}^{m} \left(1 - \frac{\lambda_\ell'\lambda_\ell b_h}{2}\right)^v + \tilde{M}$$

$$|\tilde{M}| \le \sum_{\ell=1}^{m} \sum_{r=1}^{v} M_{\ell r}.$$

But $|\tilde{M}| \le \sum_{i=r-r_0+1}^{r} KS_{h,i} \to 0$. Thus, we have proved (A11). Q.E.D.

Lemma 5. Assume (C1) - (C3). Let $\{y^{\Delta}(t)\}$ be the linear interpolation of $\{y_k\}$. The multidimensional distributions of $y^{\Delta}(t)$ converge to those of x_t as $\Delta \to 0$.

The proof is well-known and is omitted.

Proof of Theorem 1. By Lemmas 1 - 5, the finite dimensional distributions of $\xi^h(t)$ converge to those of x_t, at points of continuity of the latter (for $t \le T$). (9) follows from Lemma 2. Q.E.D.

Proof of Corollary 1. If $F(\cdot)$ were continuous on $C[0,T] = \Omega$, then the Corollary follows from Theorem 1, p. 581 [5], since our Theorem 1 assures that the conditions of the cited theorem hold. Corollary 2, p. 579 of [5] asserts that the distribution of $F(\xi^h(\cdot))$

converges to that of $F(x(\cdot))$ for all $F(\cdot)$ continuous w.p.1. on $C[0,T]$ if (a) the multidimensional distributions of $\xi^h(t)$ converge to those of x_t, and (b) the measures μ_n are weakly compact. (a) is implied by Theorem 1, and (b) is also - since, by the proof of Theorem 1, p. 581, [5], weak compactness of $\{\mu_n\}$ is implied by (9) (or, equivalently, by equicontinuity of x_t on $[0,T]$ with probability arbitrarily close to 1). The last statement of the corollary follows from the previous part of the corollary. Q.E.D.

Proof of Theorem 2. By (C6) the last two terms of (10a) are equal, and so are the last two terms of (10b). By hypothesis, $\varphi(x_{T\cap\tau})$ is uniformly bounded w.p.1., and $T \cap \tau$ is continuous w.p.1. (relative to μ). x_t is continuous in t w.p.1. Hence, $x_{T\cap\tau}$ is continuous w.p.1. on $C[0,T]$; hence, $\varphi(x_{T\cap\tau})$ is. Thus, by Corollary 1,

$$E_x\varphi(\xi^h(T\cap\tau_h)) \to E_x\varphi(x_{T\cap\tau})$$

as $h \to 0$.

A similar conclusion for the convergence of the integral term follows from the continuity of

$$\int_0^t k(x_s)ds$$

in t. Q.E.D.

Proof of Theorem 3. Define $P(n,h) = \inf_x P_x\{\tau_h \le nt\}$. Then by (11),

$$P(n+1,h) \ge (1-P(n,h))M_0 + P(n,h)$$
$$\ge M_0 + P(n,h)(1-M_0).$$

Thus

$$1 - P(n,h) \le (1-M_0)^n \tag{A14}$$

which implies

$$\int_T^\infty \tau_h dP_x(\tau_h) \to 0$$

as $T \to \infty$, uniformly in x,h. Furthermore,

$$E_x \int_{T\cap\tau_h}^T k(\xi^h(s))ds \to 0$$

$$E_x[\varphi(\xi^h(T\cap\tau_h)) - \varphi(\xi^h(\tau_h))] \to 0$$

as $T \to \infty$, uniformly in x,h. (A15) and Theorem 2 imply the Theorem. Q.E.D.

References

1. J. L. Doob, Stochastic Processes, Wiley, 1953.

2. E. B. Dynkin, Markov Processes, Springer, 1965 (Translation of 1962 Russian book).

3. E. Isaacson, H. B. Keller, Analysis of Numerical Methods, Wiley, 1966.

4. H. J. Kushner, A. J. Kleinman, "Numerical Methods for the Solution of the Degenerate Nonlinear Equations Arising in Optimal Stochastic Control Theory", IEEE Trans. on Automatic Control, AC-13, August 1968.

5. I. I. Gikhman, A. V. Skorokhod, Introduction to Random Processes, Izdatelctvo Nauka, Moscow, 1966 (in Russian).