On the Gap Between Deterministic and Stochastic Ordinary Differential Equations

Hector J. Sussmann

The Annals of Probability, Vol. 6, No. 1. (Feb., 1978), pp. 19-41.

Stable URL:
http://links.jstor.org/sici?sici=0091-1798%28197802%296%3A1%3C19%3AOTGBDA%3E2.0.CO%3B2-0

The Annals of Probability is currently published by Institute of Mathematical Statistics.

The Annals of Probability
1978, Vol. 6, No. 1, 19–41

ON THE GAP BETWEEN DETERMINISTIC AND STOCHASTIC ORDINARY DIFFERENTIAL EQUATIONS

BY HÉCTOR J. SUSSMANN

Rutgers University

We consider stochastic differential equations $dx = f(x)\,dt + g(x)\,dw$, where x is a vector in n-dimensional space, and w is an arbitrary process with continuous sample paths. We show that the stochastic equation can be solved by simply solving, for each sample path of the process w, the corresponding nonstochastic ordinary differential equation. The precise requirements on the vector fields f and g are: (i) that g be continuously differentiable and (ii) that the entries of f and the partial derivatives of the entries of g be locally Lipschitzian. For the particular case of a Wiener process w, the solutions obtained this way turn out to be the solutions in the sense of Stratonovich.

1. Introduction. The purpose of this paper is to present some results which bridge the gap between the theory of deterministic ordinary differential equations driven by a scalar input, and that of stochastic equations driven by processes such as white noise.

We consider an equation

$$\frac{dx}{dt} = f(x) + u(t)g(x) \tag{1}$$

where the *state variable* x ranges over n-dimensional Euclidean space $\mathbb{R}^n$, and where f and g are vector fields in $\mathbb{R}^n$. The "input" u is a real-valued function defined on some interval $[0, T]$. If f and g satisfy some reasonable technical hypotheses (e.g., a Lipschitz condition), then for every "nice" input u and every initial state x_0 there exists an $\varepsilon > 0$ such that, on the interval $[0, \varepsilon]$, there is a unique solution $t \to x(t)$ of (1) for which the initial condition $x(0) = x_0$ holds. If f and g satisfy some extra conditions (e.g., linear growth) then the solution $x(t)$ is actually defined for all $t \in [0, T]$.

The most obvious stochastic version of equation (1) is obtained by regarding (1) as an "equation which depends on a random parameter." That is, we assume that the input is a stochastic process, and that the initial state is a random variable. The solution should then be a stochastic process. More precisely, we are given a probability space

$$\mathscr{P} = (\Omega, \Sigma, P)\,,$$

together with a process $U = \{U(t), t \in [0, T]\}$, where each $U(t)$ is a random variable on $\mathscr{P}$. If an initial condition (i.e., a random variable X_0) is specified, then

Received October 18, 1976.

AMS 1970 *subject classifications.* Primary 60H10; Secondary 34F05.

Key words and phrases. Stochastic differential equations, sample paths.

the "stochastic initial value problem"

$$\frac{dX}{dt} = f(X) + U(t)g(X)\,, \qquad X(0) = X_0 \tag{2}$$

can be regarded, heuristically, as an initial value problem which depends on the random parameter $\omega \in \Omega$. For each $\omega \in \Omega$, the problem

$$\frac{dX}{dt}(t, \omega) = f(X(t, \omega)) + U(t, \omega)g(X(t, \omega))\,, \qquad X(0, \omega) = X_0(\omega) \tag{2ω}$$

is an ordinary initial value problem which corresponds to an equation of the same type as (1). The solution is, therefore, a function $t \to X(t, \omega)$ which also depends on ω, i.e., a family of random variables $X = \{X(t) : t \in [0, T]\}$ parametrized by t. (More generally, there may be explosions, in which case the function $t \to X(t, \omega)$ will only be defined up to a measurable explosion time $T(\omega)$, and therefore the random variable $X(t)$ will be defined on a subset $E(t)$ of Ω.)

The preceding heuristics can be transformed into rigorous mathematics if the input process U is not too irregular. For instance, it suffices to assume that, for almost every $\omega \in \Omega$, the sample path $t \to U(t, \omega)$ is a bounded measurable function. Unfortunately, one wants to solve equation (2) for inputs that are much more singular. In fact, one wants to consider inputs that are so singular the "function" $t \to U(t, \omega)$ only exists in some generalized sense (e.g., as the distributional derivative of some nondifferentiable continuous function). In this case, (2ω) will only make sense formally, and our heuristics cannot be made rigorous in a direct way. If a theory is to be developed for such inputs, it is clear that we must either (A) develop a theory of "solutions" of (1) for a class $\mathscr{U}$ of "generalized inputs" u which is so large that (2ω) has a solution for almost all ω, in all cases of interest; *or* (B) develop a theory of solutions of (2) in which one can solve (2) without having to solve (2ω) for each ω.

The most important type of input U to which such a theory should apply is white noise. For such an input, there actually exist at least two nonequivalent theories, due to Itô and Stratonovich, respectively. Both theories follow procedure (B).

In this paper, we show that procedure (A) *can be pursued.* The class $\mathscr{U}$ of inputs u for which a satisfactory theory of solutions of (1) can be developed is, simply, the class of all derivatives of continuous functions. Therefore, we succeed in developing a theory of solutions of (2) when the input U is given by $U = dW/dt$, where W is an *arbitrary process with continuous sample paths.* In our opinion, the approach presented here is aesthetically superior to the traditional one and, in addition, it has the following advantages:

(i) *Extra generality*, since W can be any process with continuous sample paths, and
(ii) *Simplicity*, since most proofs are quite elementary.

Our definition of solution is "natural." Let us rewrite equation (1) in the form

(3) $$dx = f(x)\,dt + g(x)\,dw\,,$$

where $w = [0, T] \to \mathbb{R}$ is a primitive of u. Our definition satisfies: (a) if w is continuously differentiable, then our concept of solution is identical with the ordinary one; and (b) the solutions of (3) depend continuously on w, relative to the topology of uniform convergence.

Since the C^1 functions on $[0, T]$ are dense in the space of all continuous functions on $[0, T]$, it is clear that conditions (a) and (b) uniquely determine what the definition should be. What is surprising is that the ordinary existence and uniqueness theory is valid, i.e., that solutions exist for *arbitrary continuous* w and arbitrary initial states x_0. This allows us to develop a theory of solutions of stochastic equations

(4) $$dX = f(X)\,dt + g(X)\,dW$$

following the "most natural approach" sketched above. Moreover, our definition of solution is such that the ordinary rules of calculus are obeyed (since they are obeyed when w is C^1). Therefore, it is reasonable to expect (and easy to prove) that *the solutions in our sense of equation* (4) *coincide with the Stratonovich solutions, when W is a Wiener process.* As an illustration of the usefulness of our approach, we give a completely trivial proof of a result of Wong and Zakai [11] on the approximation of solutions of (4) (with W a Wiener process) by solutions of

$$dX = f(X)\,dt + g(X)\,dW^m\,,$$

where the W^m are more regular processes which converge to W as $m \to \infty$.

In the last section of this paper, we give a simple example which shows that the theory presented here does not carry over in a simple way to equations of the form

$$dX = f(X)\,dt + \textstyle\sum_{i=1}^k g_i(X)\,dW_i$$

in which several inputs are involved. The reason for this is related to the "anomalies" that are known to occur when attempting to extend the Wong–Zakai result to the case $k > 1$ (cf. McShane [5]). It seems to us that the search for an appropriate extension of our results to $k > 1$ will require the use of substantially new methods and that, if such an extension is found, it will lead to significant progress in our understanding of the above mentioned anomalies.

REMARK. For $n = 1$, a construction similar to ours is given in Lamperti [3] (cf. also Elliott [1]). An announcement of our results has appeared in [9]. Related results appear in [8].

2. Basic definitions. If $v \in \mathbb{R}^n$, we write $|v|$ to denote the Euclidean norm of v, i.e., $|v| = (\sum_{i=1}^n v_i^2)^{\frac{1}{2}}$. If M is a square matrix, then $|M|$ denotes the matrix norm of M, i.e.,

$$|M| = \sup\{|Mx| : x \in \mathbb{R}^n,\ |x| = 1\}\,.$$

If ϕ is a scalar-, or vector-, or matrix-valued function defined on an open subset of $\mathbb{R}^n$, we say that ϕ is *Lipschitzian* on a set $S \subseteqq \mathbb{R}^n$ if there is a constant C such that $|\phi(x) - \phi(y)| \leqq C|x - y|$ for all x, y in S. We call ϕ *uniformly Lipschitzian* if ϕ is Lipschitzian on $\mathbb{R}^n$, and *locally Lipschitzian* if it is Lipschitzian on every compact subset of the domain of ϕ. We say that ϕ satisfies a *linear growth condition* if there is a constant C such that $|\phi(x)| \leqq C(1 + |x|)$ for all x.

If f is a vector field in $\mathbb{R}^n$ which is of class C^1, then Df denotes the matrix

$$Df = \left(\frac{\partial f_i}{\partial x^j}\right)_{i,j}$$

of the partial derivatives of the components of f. It is clear that, if Df is uniformly bounded, then f satisfies a linear growth condition, but the converse is not true.

We may consider vector fields f and g on an open subset U of $\mathbb{R}^n$, and a closed interval $[a, b]$ of the real line. We use $C^0([a, b])$ to denote the space of all continuous real-valued functions on $[a, b]$, and $C^0([a, b], \mathbb{R}^n)$ to denote the space of $\mathbb{R}^n$-valued continuous functions. Both spaces are given the topology of uniform convergence.

Let $t_0 \in [a, b]$, $x_0 \in U$ be given. Let $\bar{w} \in C^0([a, b])$.

DEFINITION 1. A curve $\gamma: t \to x(t)$, $a \leqq t \leqq b$, $x(t) \in U$, is said to be a solution of the initial value problem

$$(\mathrm{ivp}_{\bar{w}}) \qquad \begin{aligned} dx &= f(x)\,dt + g(x)\,d\bar{w} \\ x(t_0) &= x_0 \end{aligned}$$

if there exists a neighborhood $\mathscr{N}$ of $\bar{w}$ in $C^0([a, b])$ and a continuous map Γ: $\mathscr{N} \to C^0([a, b], \mathbb{R}^n)$ such that (writing Γ_w instead of $\Gamma(w)$):

(i) for each $w \in \mathscr{N}$ which is of class C^1, the curve Γ_w is a solution in the ordinary sense of

$$(\mathrm{ivp}_w) \qquad \begin{aligned} \frac{dx}{dt} &= f(x) + \frac{dw}{dt}(t)g(x) \\ x(t_0) &= x_0 \end{aligned}$$

and

(ii) $\Gamma_{\bar{w}} = \gamma$.

Having defined what is meant by a solution of (ivp) on a closed bounded interval $I = [a, b]$, we define the concept of solution for arbitrary intervals I in an obvious way. We say that the curve $\gamma: I \to \Omega$ is a solution of (ivp_w) if, for every closed bounded $I' \subseteqq I$ such that $t_0 \in I'$, the restriction of γ to I' is a solution of (ivp_w).

3. The main theorem.

THEOREM 1. *Assume that*

(i) *U is an open subset of* $\mathbb{R}^n$,

(ii) *f, g are vector fields on U,*
(iii) *f is locally Lipschitzian,*
(iv) *g is of class C^1 and its partial derivatives are locally Lipschitzian.*

Let I be an interval of the real line, and let $t_0 \in I$, $x_0 \in U$. Let w be a real-valued continuous function on I. Then

(1) *There exists an interval I', containing t_0 in its interior, and a curve $t \to x(t)$, $t \in I' \subseteqq I$, which is a solution of* (ivp$_w$).

(2) *If I' is any such interval, then the solution of* (ivp$_w$) *which is defined on $I' \subseteqq I$ is unique.*

THEOREM 2. *With the same hypotheses as in Theorem* 1 *assume, in addition, that $U = \mathbb{R}^n$, that f satisfies a linear growth condition, and that Dg is uniformly bounded.*

Then, for every choice of I, t_0, x_0, w, there is a solution of (ivp$_w$) *defined on the whole interval I.*

Moreover, the solution depends continuously on t_0, x_0 and w.

REMARK. The assumption that Dg is uniformly bounded implies, in particular, that g satisfies a linear growth condition. Theorem 2 is not true if the requirement that Dg be bounded is eliminated, even if g is required to grow linearly (cf. Section 8).

PROOF OF THEOREMS 1 AND 2. First, it is clear that both the existence and uniqueness results follow for arbitrary I if they are true for all compact I. Moreover, it clearly suffices to assume that $t_0 = 0$. From now on, it is assumed that I is compact, and that $t_0 = 0$.

The uniqueness is trivial. Indeed, if $\gamma_1 : t \to x_1(t)$ and $\gamma_2 : t \to x_2(t)$ are solutions of (ivp$_w$) on $I = [a, b]$, for a continuous $w : [a, b] \to \mathbb{R}$, then there are neighborhoods $\mathcal{N}_1$, $\mathcal{N}_2$ of w in $C^0(I, \mathbb{R})$, and continuous maps Γ^1, Γ^2 from $\mathcal{N}_1$, $\mathcal{N}_2$ to $C^0(I, \mathbb{R}^n)$ such that, for $i = 1, 2$, $\Gamma_w{}^i = \gamma_i$ and that, whenever $\tilde{w} \in \mathcal{N}^i$ is of class C^1, then $\Gamma_{\tilde{w}}{}^i$ is a solution of (ivp$_{\tilde{w}}$). Since the C^1 functions are dense in $\mathcal{N}_1 \cap \mathcal{N}_2$, it follows that $\Gamma_w{}^1 = \Gamma_w{}^2$, i.e., $\gamma_1 = \gamma_2$.

We now prove existence. We shall first prove the existence result of Theorem 2, and then use it to deduce the result of Theorem 1.

We begin by proving the existence of solutions in a special case. We assume that

(I) g is a constant vector field, i.e., there exists a vector $v \in \mathbb{R}^n$, $v = (v_1, \cdots, v_n)$, such that $g(x) = v$ for all x,

(II) f is locally Lipschitzian, and

(III) there exists a smooth real-valued function ϕ on $\mathbb{R}^n$ such that

$$\text{(i)} \quad \textstyle\sum_{i=1}^n f_i \dfrac{\partial \phi}{\partial x^j} \equiv 0$$

$$\text{(ii)} \quad \textstyle\sum_{i=1}^n v_i \dfrac{\partial \phi}{\partial x^i} \equiv 1 \quad \text{and}$$

(iii) for each $T > 0$ there is a constant C_T such that
$$|f(x)| \leqq C_T(1 + |x|) \quad \text{whenever} \quad |\phi(x)| \leqq T .$$

Let $w : [a, b] \to \mathbb{R}$ be a continuous function, and assume that $0 \in [a, b]$. We construct the solution $t \to x(t, \bar{x}, w)$ of $dx = f(x)\,dt + g(x)\,dw$, $x(0) = \bar{x}$ by the method of successive approximations. We let, for $t \in [ab]$, $x_0(t) = \bar{x}$

$$(\sharp) \qquad x_{k+1}(t) = \bar{x} + \textstyle\int_0^t f(x_k(\tau))\,d\tau + [w(t) - w(0)]v .$$

It is understood that the x_k also depend on $\bar{x}$ and w. We now show that

(*) The limit $\lim_{k\to\infty} x_k(t)$ exists for all $t \in [a, b]$, and for all $x \in \mathbb{R}^n$, $w \in C^0[a, b]$. Moreover, the convergence is uniform in t, $\bar{x}$, w, as long as $\bar{x}$ remains within a compact set $K \subseteq \mathbb{R}^n$ and w within a bounded set B of $C^0[a, b]$.

To prove (*), let $A_1 > 0$ be such that $2|w(t)| \leqq A_1$ for $t \in [a, b]$, $w \in B$. Let
$$A_2 = \sup\{|\phi(x)| : x \in K\} .$$
We claim that
$$(**) \qquad |\phi(x_k(t))| \leqq A_1 + A_2$$
for all $t \in [a, b]$, all k, all $w \in B$ and all $\bar{x} \in K$.

To prove this, fix k, t with $k > 0$, $t \neq 0$ and put
$$\xi(\tau) = \bar{x} + \textstyle\int_0^\tau \left[f(x_{k-1}(s)) + \dfrac{1}{t}(w(t) - w(0))v \right] ds .$$
Then $\xi(0) = \bar{x}$, $\xi(t) = x_k(t)$. Moreover, ξ is a C^1 function, and
$$\frac{d\xi}{d\tau}(\tau) = f(x_{k-1}(\tau)) + \frac{1}{t}[w(t) - w(0)]v .$$
Let $h(\tau) = \phi(\xi(\tau))$. Then
$$\begin{aligned}
\frac{dh}{d\tau}(\tau) &= [(\operatorname{grad}\phi)(\xi(\tau))] \cdot \frac{d\xi}{d\tau}(\tau) \\
&= \textstyle\sum_{i=1}^n \dfrac{d\xi_i}{d\tau}(\tau)\dfrac{\partial\phi}{\partial x_i}(\xi(\tau)) \\
&= \textstyle\sum_{i=1}^n \left[f_i(\xi(\tau)) + \dfrac{w(t) - w(0)}{t} v_i \right] \dfrac{\partial\phi}{\partial x_i}(\xi(\tau)) \\
&= \frac{w(t) - w(0)}{t} ,
\end{aligned}$$
because of assumptions (III)(i), (ii). Now $\phi(x_k(t)) = \phi(\xi(t)) = h(t)$ so
$$\begin{aligned}
\phi(x_k(t)) &= h(0) + \textstyle\int_0^t \dfrac{dh}{d\tau}(\tau)\,d\tau \\
&= h(0) + w(t) - w(0) \\
&= \phi(\bar{x}) + w(t) - w(0) .
\end{aligned}$$

Therefore

$$|\phi(x_k(t))| \leqq |\phi(\bar{x})| + A_1$$
$$\leqq A_1 + A_2$$

if $\bar{x} \in K$. This completes the proof of (**) if $k \neq 0$, $t \neq 0$. If $k = 0$ or if $t = 0$, then $x_k(t) = \bar{x}$ so (**) holds as well. So (**) has been established in all cases.

Now let $T = A_1 + A_2$, and let $A_3 = C_T$. Then

$$(^{\#\#}) \qquad |f(x_k(t))| \leqq A_3(1 + |x_k(t)|)$$

for all $\bar{x} \in K$, all $w \in B$, and all t, k (because of (III)(iii)). Let $A_4 = \sup\{|\bar{x}| : \bar{x} \in K\}$. Then $(^{\#})$ and $(^{\#\#})$ imply that

$$(^{\#\#\#}) \qquad |x_{k+1}(t)| \leqq A_4 + A_1|v| + A_3 \textstyle\int_0^t (1 + |x_k(\tau)|)\, d\tau .$$

Let $c = \max(|a|, |b|)$

$$A_5 = \max(A_4 + A_1|v| + A_3 c, A_3) .$$

Then $(^{\#\#\#})$ implies that

$$|x_{k+1}(t)| \leqq A_5 + A_5 \textstyle\int_0^t |x_k(\tau)|\, d\tau .$$

From this it follows by induction on k that

$$|x_k(t)| \leqq A_5 \sum_{j=0}^{k} \frac{A_5{}^j |t|^j}{j!}$$

(in the first step, we use the fact that $x_0(t) = \bar{x} \in K$ and therefore $|x_0(t)| \leqq A_5$).

It follows that

$$|x_k(t)| \leqq A_5 e^{A_5|t|} \qquad t \in [a, b],\ \bar{x} \in K,\ w \in B .$$

Therefore, if we let $A_6 = A_5 e^{A_5 c}$, we find that, for all k

$$|x_k(t)| \leqq A_6 , \qquad t \in [a, b],\ \bar{x} \in K,\ w \in B .$$

Let J be the compact set $\{x : |x| \leqq A_6\}$. Then f satisfies a Lipschitz condition on J with constant A_7. So

$$|x_{k+1}(t) - x_k(t)| = |\textstyle\int_0^t [f(x_k(\tau)) - f(x_{k-1}(\tau))]\, d\tau|$$
$$\leqq A_7 \textstyle\int_0^t |x_k(\tau) - x_{k-1}(\tau)|\, d\tau$$

since $x_k(\tau)$, $x_{k-1}(\tau)$ belong to J.

This formula implies by induction that

$$|x_{k+1}(t) - x_k(t)| \leqq \frac{2A_6 A_7{}^k |t|^k}{k!}$$

(using, in the first step, the estimate $|x_1(t) - x_0(t)| \leqq 2A_6$).

Since the constants A_6, A_7 only depend on K and B, it follows that the sequence of numbers $\{x_k(t)\}$ is Cauchy, and that this happens uniformly in t, $\bar{x} \in K$ and $w \in B$. So (*) is proved.

For each k, let $L_k : \mathbb{R}^n \times C^0[a, b] \to C^0([a, b], \mathbb{R}^n)$ be the map which, to each $\bar{x} \in \mathbb{R}^n$ and each $w \in C^0[a, b]$, assigns the curve $t \to x_k(t)$. Then L_k is continuous. It follows from (*) that the L_k have a limit L as $k \to \infty$, and that L_k converges to L uniformly on bounded subsets of $\mathbb{R}^n \times C^0([a, b])$. Therefore L is continuous. Now let $w \in C^0([a, b])$, and let $\bar{x} \in \mathbb{R}^n$. We claim that $L(\bar{x}, w)$ is a solution of

$$dx = f(x)\,dt + g(x)\,dw$$

in the sense of our definition. Indeed, if $\tilde{w}$ is a C^1 function on $[a, b]$, then

$$[\tilde{w}(t) - \tilde{w}(0)]v = \left[\int_0^t \frac{d\tilde{w}}{d\tau}(\tau)\,d\tau\right] v\,.$$

Therefore, the successive approximations $t \to L_k(\bar{x}, \tilde{w})(t)$ satisfy

$$L_{k+1}(\bar{x}, \tilde{w})(t) = \bar{x} + \int_0^t \left[f(x_k(\tau)) + \frac{d\tilde{w}}{d\tau}(\tau) g(x_k(\tau))\right] d\tau\,.$$

So the $L_k(\bar{x}, \tilde{w})$ are the ordinary successive appoximations that are used to construct the solutions of

$$(\beta) \qquad \frac{dx}{dt} = f(x) + \frac{d\tilde{w}}{dt} g(x)\,.$$

It follows that, for every $\bar{x}$, the function $L(\bar{x}, \tilde{w})$ is the solution of (β) which has the value $\bar{x}$ when $t = 0$. Since the map $w \to L(\bar{x}, w)$ is continuous, it follows that, for each $\bar{x}$, w, the curve $L(\bar{x}, w)$ is a solution of $dx = f(x)\,dt + g(x)\,dw$ in the sense of our definition. Since L also depends continuously on $\bar{x}$, it is clear that the last assertion of Theorem 2 follows, if the special assumptions (I), (II) and (III) hold. We now prove Theorem 2 in the general case. The idea is to reduce it to the special situation considered above. In order to do this, it is convenient to introduce some notations.

If X is a locally Lipschitz vector field on $\mathbb{R}^m$, we use Φ^X to denote the flow of X. Precisely, let $x \in \mathbb{R}^m$, and let $t \to \gamma(t)$ denote the integral curve of X for which $\gamma(0) = x$. We write $\Phi^X(t)x$ for $\gamma(t)$. In general, Φ^X is a local one-parameter group, i.e.,

(a) the domain of definition $\Omega_X \subseteq \mathbb{R} \times \mathbb{R}^m$ of the map $(t, x) \to \Phi^X(t)x$ is an open set which contains $\{0\} \times \mathbb{R}^m$,

(b) $\Phi^X(t)[\Phi^X(\tau)x] = \Phi^X(t + \tau)x$ whenever both sides are defined, and

(c) $\Phi^X(0)x = x$.

By definition, X is *complete* if $\Omega_X = \mathbb{R} \times \mathbb{R}^m$, i.e., if the integral curves are defined for all t. If X satisfies a linear growth condition then it is complete. If X is C^1 then the map $F\colon (t, x) \to \Phi^X(t)x$ is C^1. The time derivative of F is given by

$$\frac{\partial F}{\partial t} = X \circ F\,.$$

The matrix $D_x F = (\partial F_i / \partial x^k)_{1 \le i, j \le n}$ is determined as follows: for a given x_0, the matrix function $t \to (D_x F)(t, x_0)$ is the solution $A(t)$ of the equation

$$\frac{dA}{dt}(t) = (D_x X)(\Phi^X(t) x_0) A(t)$$

for which $A(0) =$ identity.

In particular, let us apply this with $m = n$, $X = g$. Define

$$F(t, x) = \Phi^g(t)x .$$

Then F is a C^1 map from $\mathbb{R} \times \mathbb{R}^n$ onto $\mathbb{R}^n$ (because g is complete and C^1).

Now define a vector field $\bar{f}$ on $\mathbb{R} \times \mathbb{R}^n$ as follows:

$$\bar{f}(t, x_0) = (0, [(D_x F)(t, x_0)]^{-1} f(F(t, x_0)) .$$

Also, put

$$\bar{g} = (1, 0, \cdots, 0) .$$

For $(t, x) \in \mathbb{R}^{n+1}$, let $F_*(t, x)$ denote the Jacobian matrix of F at (t, x). Then

$$F_*(t, x)\bar{g}(t, x) = g(F(t, x)) \qquad \text{and}$$
$$F_*(t, x)\bar{f}(t, x) = f(F(t, x)) .$$

Therefore, the following holds:

(A) If $t \to w(t)$, $a \leqq t \leqq b$, is a C^1 real-valued function, and if

$$t \to \xi(t) \in \mathbb{R}^{n+1}$$

is a solution of

$$\frac{d\xi}{dt}(t) = \bar{f}(\xi(t)) + \frac{dw}{dt}(t)\bar{g}(\xi(t))$$

then the curve

$$t \to F(\xi(t)) \in \mathbb{R}^n$$

is a solution of $dx = f\,dt + g\,dw$.

Let us assume for the moment that the conclusions of the theorem we are trying to prove hold for $\bar{f}$, $\bar{g}$ on $\mathbb{R}^{n+1}$. We claim that it follows that they hold for f, g on $\mathbb{R}^n$. Since we assume that the conclusion holds for $\bar{f}$, $\bar{g}$, there is a continuous map

$$\bar{L}\colon \mathbb{R}^{n+1} \times C^0([a, b]) \to C^0([a, b], \mathbb{R}^{n+1})$$

which is such that, whenever w is C^1, then $\bar{L}(\xi_0, w)$ is a solution in the ordinary sense of $d\xi = \bar{f}(\xi)\,dt + \bar{g}(\xi)\,dw$, with initial condition $\xi(0) = \xi_0$. Define $L'\colon \mathbb{R}^{n+1} \times C^0([a, b]) \to C^0([a, b], \mathbb{R}^n)$ by $L'(\xi_0, w) = F \circ (\bar{L}(\xi_0, w))$ (recall that $\bar{L}(\xi_0, w)$ is a curve). Since F is continuous, it is clear that L' is continuous. Now define $L\colon \mathbb{R}^n \times C^0([a, b]) \to C^0([a, b], \mathbb{R}^n)$ by $L(x_0, w) = L'((0, x_0), w)$. It is clear that L is continuous. If $w \in C^1$, then (A) implies that $L'((0, x_0), w)$ is a solution of $dx = f(x)\,dt + g(x)\,dw$ in the ordinary sense. The initial condition satisfied by this solution is, obviously, $x(0) = x_0$. So L is a continuous extension to all of $\mathbb{R}^n \times C^0([a, b])$ of the map which, to each (x_0, w) in $\mathbb{R}^n \times C^1([a, b])$, assigns the solution of $dx = f(x)\,dt + g(x)\,dw$ with initial condition $x(0) = x_0$.

This shows that the theorem for f, g follows if we prove it for $\bar{f}$, $\bar{g}$. In order to prove it for $\bar{f}$, $\bar{g}$, we will show that $\bar{f}$ and $\bar{g}$ satisfy conditions (I), (II) and (III). First, it is clear that $\bar{g}$ is a constant vector field. To prove that $\bar{f}$ is locally Lipschitzian, it suffices to show that the entries of $(D_x F)^{-1}$ and those of $f \circ F$ are locally Lipschitzian (since a product of locally Lipschitzian functions is locally Lipschitzian). Since f is itself locally Lipschitzian, and F is C^1, then $f \circ F$ is locally Lipschitzian. So we must show that $(D_x F)^{-1}$ is locally Lipschitzian. Now, the inverse of a matrix-valued locally Lipschitzian function M on $\mathbb{R}^n$ is necessarily Lipschitzian, as long as $M(x)$ is invertible for all x (reason: $M(y)^{-1} - M(x)^{-1} = M(y)^{-1}[M(x) - M(y)]M(x)^{-1}$. If $M(x)$ is invertible for all x then on each compact set K, the function $x \rightarrow |M(x)^{-1}|$ is continuous and therefore bounded. So, if $|M(x)^{-1}| \leqq C_1$ for $x \in K$, and $|M(x) - M(y)| \leqq C_2|x - y|$ for $x, y \in K$, it follows that $|M(x)^{-1} - M(y)^{-1}| \leqq C_1^2 C_2|x - y|$). So it suffices to show that $D_x F$ is locally Lipschitzian. Now $(\partial/\partial t)[(D_x F)(t, x)] = (Dg)(F(t, x)) \cdot (D_x F)(t, x)$ and $(D_x F)(0, x) = 1_n$ (where 1_n is the $n \times n$ identity matrix). So

$$(') \qquad (D_x F)(t, x) = 1_n + \textstyle\int_0^t (Dg)(F(\tau, x)) \cdot (D_x F)(\tau, x)\, d\tau\, .$$

Now let $K \subseteqq \mathbb{R}^{n+1}$ be a compact set of the form $K_1 \times K_2$, where $K_1 \subseteqq \mathbb{R}$ is a compact interval which contains 0, and K_2 is a compact subset of $\mathbb{R}^n$. Since F is C^1, and Dg is locally Lipschitzian, it follows that $(Dg) \circ F$ is locally Lipschitzian, so that there exists a constant C_1 such that

$$|(Dg)(F(\tau, x)) - (Dg)(F(\tau, y))| \leqq C_1|x - y|$$

whenever $x, y \in K_2$, $\tau \in K_1$. Let $C_2 = \sup\{|\tau| : \tau \in K_1\}$. Then, if $t \in K_1$, and $x, y \in K_2$, we have:

$$\begin{aligned} |(D_x F)(t, x) &- (D_x F)(t, y)| \\ &= \textstyle\int_0^t [(Dg)(F(\tau, x)) - (Dg)(F(\tau, y))] \cdot (D_x F)(\tau, x)\, d\tau \\ &\quad + \textstyle\int_0^t (Dg)(F(\tau, y))[(D_x F)(\tau, x) - (D_x F)(\tau, y)]\, d\tau| \\ &\leqq C_1 C_2 C_3|x - y| + C_4 \textstyle\int_0^t |(D_x F)(\tau, x) - (D_x F)(x, y)|\, d\tau\, , \end{aligned}$$

where

$$C_3 = \sup\{|(D_x F(\tau, \xi))| : (\tau, \xi) \in K\} \qquad \text{and}$$
$$C_4 = \sup\{|(Dg)(F(\tau, \xi))| : (\tau, \xi) \in K\}\, .$$

Gronwall's inequality then implies that $|(D_x F)(t, x) - (D_x F)(t, y)| \leqq C_5|x - y|$, where $C_5 = e^{C_4 C_2} C_1 C_2 C_3$.

It follows from (′) that $|(D_x F)(t, x) - (D_x F)(\tau, x)| \leqq C_3 C_4|t - \tau|$. Therefore, if $(t, x) \in K$, $(\tau, y) \in K$, we have

$$|(D_x F)(t, x) - (D_x F)(\tau, y)| \leqq C_6[|x - y| + |t - \tau|]$$

where $C_6 = \max(C_3, C_4, C_5)$.

This shows that $D_x F$ is locally Lipschitz and therefore that $\tilde{f}$ is locally Lipschitz.

We must now prove that (III) is satisfied. We take $\phi(t, x) = t$. It is clear that (i) and (ii) hold. To prove (iii), we must show that for each $T > 0$ there is a constant C_T such that $|\tilde{f}(t, x)| \leqq C_T(1 + |x|)$ whenever $|t| \leqq T$. So let T be chosen. Since g satisfies a linear growth condition, there exists E_1 such that

$|g(x)| \leqq E_1(1 + |x|)$ for all x. Then we can use

$$F(t, x) = x + \int_0^t g(F(\tau, x))\, d\tau$$

to conclude that, if $|t| \leqq T$, then

$$|F(t, x)| \leqq |x| + E_1 T + E_1 \int_0^t |F(\tau, x)|\, d\tau .$$

By Gronwall's inequality,

$$\begin{aligned} |F(t, x)| &\leqq (|x| + E_1 T)e^{E_1 T} \\ &\leqq E_2(1 + |x|) \end{aligned}$$

if $E_2 = e^{E_1 T} \max(1, E_1 T)$.

Since f satisfies a linear growth condition, we have

$$|f(x)| \leqq E_3(1 + |x|)$$

and therefore, if $|t| \leqq T$,

$$|f(F(t, x))| \leqq E_4(1 + |x|) , \qquad \text{if} \quad E_4 = E_3(1 + E_2) .$$

We now show that $(D_x F)^{-1}$ is bounded on $[-T, T] \times \mathbb{R}^n$. We know that, for a given x, $(D_x F)(t, x)$ is the value at t of the solution $A(\tau)$ of

$$\frac{dA}{d\tau}(\tau) = (Dg)(F(\tau, x))A(\tau)$$

that satisfies $A(0) = 1_n$. But then $(D_x F)(t, x)^{-1}$ is the value at t of the solution $B(\tau)$ of

$$\frac{dB}{d\tau}(\tau) = -B(\tau) \cdot (Dg)(F(\tau, x))$$

for which $B(0) = 1_n$. So

$$[(D_x F)(t, x)]^{-1} = 1_n - \int_0^t [(D_x F)(\tau, x)]^{-1}(Dg)(F(\tau, x))\, d\tau .$$

Gronwall's inequality gives, for $|t| \leqq T$,

$$[(D_x F)(t, x)]^{-1} \leqq e^{E_5 T} ,$$

where

$$E_5 = \sup\{|(Dg)(x)| : x \in \mathbb{R}^n\} .$$

Since $\bar{f}(t, x) = (0, [(D_x F)(t, x)]^{-1} f(F(t, x)))$, it follows that

$$|\bar{f}(t, x)| \leqq E_4 e^{E_5 T}(1 + |x|)$$

whenever $|t| \leqq T$. So, if we let $C_T = E_4 e^{E_5 T}$, we find that (III) holds.

This completes the proof of Theorem 2. We now prove Theorem 1. Let $r: U \to \mathbb{R}$ be a C^∞ function which vanishes in the complement of a compact subset of U, and which is equal to one in a neighborhood V of x_0. Define $\hat{f}$, $\hat{g}$ by $\hat{f}(x) = r(x) f(x)$, $\hat{g}(x) = r(x)\, g(x)$ for $x \in U$, $\hat{f}(x) = \hat{g}(x) = 0$ for $x \notin U$. It is then clear that $\hat{f}$ and $\hat{g}$ satisfy all the assumptions of Theorem 2, so there is a solution $\hat{\gamma}: t \to x(t)$ of

$$dx = \hat{f}(x)\,dt + \hat{g}(x)\,dw$$
$$x(0) = x_0$$

defined on $[a, b]$.

Since $x_0 \in V$, there are a', b' with $a \leqq a' < 0 < b' \leqq b$ such that $\hat{x}(t) \in V$ for $a' \leqq t \leqq b'$. Moreover, $\hat{f} \equiv f$ and $\hat{g} \equiv g$ on V.

Then the restriction γ of $\hat{\gamma}$ to $[a', b']$ is a solution of $dx = f(x)\,dt + g(x)\,dw$, $x(0) = x_0$, because of the following trivial result, which we state explicitly for future reference.

LEMMA 3. *Let U_1, U_2 be open subsets of $\mathbb{R}^n$, and let f_i, g_i $(i = 1, 2)$ be vector fields on U_i. Assume that w is a continuous real function on an interval I, and that the $U_1 \cap U_2$-valued curve $t \to \gamma(t)$, $t \in I$ is a solution of*

$$dx = f_1(x)\,dt + g_1(x)\,dw\,, \qquad x(t_0) = x_0\,.$$

Moreover, assume that $f_2 \equiv f_1$ and $g_2 \equiv g_1$ on some open set $U \subseteqq U_1 \cap U_2$ that contains $\gamma(t)$ for all $t \in I$. Then γ is a solution of

$$dx = f_2(x)\,dt + g_2(x)\,dw\,, \qquad x(t_0) = x_0\,.$$

The proof of the lemma is a straightforward application of the definition of "solution."

COROLLARY 4. *With the same assumptions as in Theorem* 1, *assume that* $0 \in I = (a, b)$ *and that a solution* $\gamma: t \to x(t)$ *of* (ivp_w) *exists on an interval* $(c, d) \subseteqq (a, b)$, *with* $0 \in (c, d)$ *and* $d < b$. *If the image of* (c, d) *under* γ *is contained in a compact subset of* U, *then there exists* $\varepsilon > 0$ *such that* γ *extends to a solution of* (ivp_w) *defined on* $(c, d + \varepsilon)$.

PROOF. Let K be a compact subset of U which contains the image of γ. Let $r: U \to \mathbb{R}$ be a C^∞ function which equals one in an open set V that contains K, and which vanishes in the complement of a compact subset of U. As in the last part of the proof of Theorems 1 and 2, define $\hat{f}$, $\hat{g}$ by $\hat{f}(x) = r(x) f(x)$, $\hat{g}(x) = r(x)\,g(x)$ for $x \in U$, $\hat{f}(x) = \hat{g}(x) = 0$ for $x \notin U$. Then $\hat{f}$, $\hat{g}$ satisfy all the hypotheses of Theorem 2. Therefore, there exists a solution $\hat{\gamma}$ of

$$dx = \hat{f}(x)\,dt + \hat{g}(x)\,dw$$
$$x(0) = x_0$$

defined on (a, b).

The restriction γ' of $\hat{\gamma}$ to (c, d) is also a solution of $dx = \hat{f}\,dt + \hat{g}\,dw$, $x(0) = x_0$. Moreover, $\hat{f} \equiv f$ and $\hat{g} \equiv g$ on a neighborhood of the set $\{\gamma'(t): c < t < d\}$. Therefore, by Lemma 3, γ' is a solution of $dx = f\,dt + g\,dw$, $x(0) = x_0$. Since γ is also a solution of this initial value problem, it follows that $\gamma \equiv \gamma'$. Now $\gamma(t) \in K$ for $c < t < d$. Since $\hat{\gamma}$ is continuous, there exists an $\varepsilon > 0$ such that $\hat{\gamma}(t) \in V$ for $c < t < d + \varepsilon$. But $\hat{f} \equiv f$ and $\hat{g} \equiv g$ on V. Therefore the restriction of $\hat{\gamma}$ to $c < t < d + \varepsilon$ is also a solution of $dx = f\,dt + g\,dw$, $x(0) = x_0$, completing the proof.

Now let f, g be vector fields on the open set $U \subseteqq \mathbb{R}^n$. Let $x_0 \in \mathbb{R}^n$, let I be an interval, $t_0 \in I$, and let $w : I \to \mathbb{R}$ be continuous.

Assume that f, g satisfy the hypotheses of Theorem 1. Then there exists an interval $J \subseteqq I$ such that t_0 is in the interior of J relative to I, and that there is a unique solution x_J of $dx = f\,dt + g\,dw$, $x(t_0) = x_0$, defined on J. By the uniqueness of solutions, there is a *maximal* interval $J_{\max}$ with that property, and if J is any other interval, then the restriction of $x_{J_{\max}}$ to J is x_J.

DEFINITION 2. The function $x_{J_{\max}}$ is the maximal solution of $dx = f\,dt + g\,dw$, $x(t_0) = x_0$.

4. Stochastic differential equations. We now consider a stochastic differential equation

$$\text{(st)} \qquad dX = f(X)\,dt + g(X)\,dW\,, \qquad X_{t_0} = \bar{X}\,.$$

Precisely, we assume that:

(α) A probability space $\mathscr{P} = (\Omega, \Sigma, P)$ is given (i.e., Σ is a σ-algebra of subsets of Ω, and P a probability measure on Ω);

(β) An interval $I \subseteqq \mathbb{R}$, and a process $W = \{W_t : t \in I\}$ are specified;

(γ) W has continuous sample paths (i.e., for almost every $\omega \in \Omega$, the function $t \to W_t(\omega)$ is continuous);

(δ) $\bar{X}$ is a random variable on $\mathscr{P}$.

DEFINITION 3. A solution of (st) is a family $\{X_t : t \in I\}$ of random variables such that, for almost every $\omega \in \Omega$ for which $w_\omega : t \to W_t(\omega)$ is continuous, the curve $t \to X_t(\omega)$ is a solution of

$$(\mathrm{st}_\omega) \qquad dx = f(x)\,dt + g(x)\,dw_\omega\,, \qquad x(t_0) = \bar{X}(\omega)$$

in the sense of Definition 1.

THEOREM 5. *If* (α), (β), (γ), (δ) *above hold, and if f and g are vector fields that satisfy the hypotheses of Theorem* 2, *then there is one and only one solution of* (st).

REMARK. The assertion that only one solution exists means that, if $\{X_t, t \in I\}$, $\{Y_t, t \in I\}$, then there is a set A such that $P(A) = 1$ and that for all $\omega \in A$, and all $t \in I$, $X_t(\omega)$ equals $Y_t(\omega)$.

PROOF OF THEOREM 5. The uniqueness is trivial. Indeed, if $\{X_t\}$ and $\{Y_t\}$ are two solutions, then there is a set $A \subseteqq \Omega$ such that $P(A) = 1$ and that the following are true for $\omega \in A$:

(i) w_ω is continuous;

(ii) $t \to X_t(\omega)$ and $t \to Y_t(\omega)$ are solutions of (st_ω).

But then, by the uniqueness part of Theorem 2, the functions $t \to X_t(\omega)$ and $t \to Y_t(\omega)$ coincide for $\omega \in A$. This clearly establishes the uniqueness.

The proof of existence is almost trivial. Let A be such that $P(A) = 1$ and that w_ω is continuous for all $\omega \in A$. For each $\omega \in A$, define the function x_ω to be the solution of (st_ω) (which, by Theorem 2, exists and is unique). Now put $X_t(\omega) = x_\omega(t)$. It is then clear that $\{X_t \colon t \in I\}$ satisfies all the desired properties, except for the fact that it is not completely obvious that the X_t are random variables (i.e., Σ-measurable). So we must prove that the X_t are measurable. This is most easily done by tracing through the steps of the proof of Theorem 2. Suppose first that f and g satisfy the very special conditions (I), (II), (III) that are described in the proof of Theorem 2. Then, for each $\omega \in A$, the function x_ω is constructed by successive approximations:

$$x_\omega(t) = \lim_{k\to\infty} x_{\omega,k}(t)$$

where

$$x_{\omega,0}(t) = \bar{X}(\omega)$$
$$x_{\omega,k+1}(t) = \bar{X}(\omega) + \textstyle\int_0^t f(x_{\omega,k}(\tau))\,d\tau + [w_\omega(t) - w_\omega(0)]v\,.$$

It is clear from these formulas that all the functions $\omega \to x_{\omega,k}(t)$ are Σ-measurable. Therefore $\omega \to x_\omega(t)$ is Σ-measurable, and our proof is complete.

In the general case, we let $\bar{f}$, $\bar{g}$, F be as in the proof of Theorem 2. Let $\bar{X}'(\omega) = (0, \bar{X}(\omega))$. Let $\bar{x}_\omega$ be the solution of $dx = \bar{f}\,dt + \bar{g}\,dw_\omega$, with initial condition $\bar{x}_\omega(0) = \bar{X}'(\omega)$. Then $x_\omega(t) = F(\bar{x}_\omega(t))$. We already know that, for each t, $\omega \to \bar{x}_\omega(t)$ is measurable. Hence $\omega \to x_\omega(t)$ is also measurable, and our proof is complete.

Corollary 6. *Let the hypotheses of Theorem 5 be satisfied. For each random variable Y, let Σ_Y denote the smallest σ-algebra with respect to which Y is measurable. For each $t \in I$, let $\mathscr{A}_t$ be the σ-algebra generated by $\{\Sigma_{W_\tau - W_{t_0}} \colon \tau$ between t_0 and $t\}$. Let $\tilde{\mathscr{A}}_t$ be the σ-algebra generated by $\mathscr{A}_t$ and $\Sigma_{\bar{X}}$. Then X_t is $\tilde{\mathscr{A}}_t$-measurable for each t.*

Proof. The solutions of (st) do not change if the process $\{W_t\}$ is replaced by $\{W_t'\}$, where $W_t' = W_t - W_{t_0}$. Therefore we might as well assume that $W_{t_0} \equiv 0$. Now assume, for simplicity, that $t > t_0$ (the case $t < t_0$ is similar). Let $\tilde{I} = [t_0, t]$, $\tilde{\Sigma} = \tilde{\mathscr{A}}_t$, $\tilde{P} =$ the restriction of P to $\tilde{\Sigma}$ and $\tilde{W} = \{\tilde{W}_\tau \colon \tau \in \tilde{I}\}$ where, for each $\tau \in \tilde{I}$, $\tilde{W}_\tau = W_\tau$. Then all the conditions of Theorem 5 hold for the probability space $\tilde{\mathscr{P}} = (\Omega, \tilde{\Sigma}, \tilde{P})$, the input process $\tilde{W}$, and the initial condition $\bar{X}$. So there is a unique solution $\{\tilde{X}_\tau\}$ of $dX = f(X)\,dt + g(X)\,d\tilde{W}$, $\tilde{X}_{t_0} = \bar{X}$, defined for $t_0 \leq \tau \leq t$. It follows easily from the definition of the terms involved that $\tilde{X}_\tau = X_\tau$ for $\tau \in \tilde{I}$. But, by definition, $\tilde{X}_t$ is $\tilde{\Sigma}$-measurable. Hence X_t is $\tilde{\mathscr{A}}_t$-measurable, and our proof is complete.

We now consider the more general case in which no linear growth assumptions are made on f, g. In this case, the solution $X_t(\omega)$ is going to be defined, for each ω, on some interval $T_1(\omega) < t < T_2(\omega)$, which can depend on ω. For simplicity, we shall only consider the case when $t_0 \in \mathring{I}$, the interior of I. (The cases $t_0 = \min I$, $t_0 = \max I$ are similar).

Definition 4. Let $\mathscr{P} = (\Omega, \Sigma, P)$ be a probability space, I an interval (not necessarily bounded or closed), $t_0 \in I$, and let T_1, T_2 be measurable functions on Ω (the values $\pm\infty$ being allowed) such that, for almost all $\omega \in \Omega$, $T_1(\omega) \in \bar{I}$, $T_2(\omega) \in \bar{I}$, and $T_1(\omega) < t_0 < T_2(\omega)$ (here $\bar{I}$ is the closure of I). Let $\bar{X}$ be a random variable on $\mathscr{P}$. A solution of $dX = f(X)\,dt + g(X)\,dW$, $X_{t_0} = \bar{X}$, defined between T_1 and T_2, is a function $(t, \omega) \to X_t(\omega)$ defined on the set $E = \{(t, \omega): T_1(\omega) < t < T_2(\omega)\}$ such that

(a) for each t, the function $X_t: \omega \to X_t(\omega)$ (which is defined on the measurable set $E_t = \{\omega : T_1(\omega) < t < T_2(\omega)\}$) is measurable, and

(b) for each ω, the function $t \to X_t(\omega)$ (defined on the interval $I_\omega = \{t \in I, T_1(\omega) < t < T_2(\omega)\}$) is a solution of $dx = f(x)\,dt + g(x)\,dw_\omega$, $x(t_0) = \bar{X}(\omega)$ (where $w_\omega(t) = W_t(\omega)$).

A solution $\{X_t\}$ defined between T_1 and T_2 is said to be an extension of a solution $\{X_t'\}$ defined between T_1' and T_2' if $T_1(\omega) \leqq T_1'(\omega)$, $T_2'(\omega) \leqq T_2(\omega)$ for almost all ω, and if $X_t'(\omega) = X_t(\omega)$ for $T_1'(\omega) < t < T_2'(\omega)$.

We say that $\{X_t\}$ is a strict extension of $\{X_t'\}$ if either $P(T_1(\omega) < T_1'(\omega)) > 0$ or $P(T_2(\omega) > T_2'(\omega)) > 0$. We say that $\{X_t\}$ is a maximal solution if there is no solution which strictly extends it.

THEOREM 7. *Assume that*

(a) *f and g are vector fields on an open subset U of $\mathbb{R}^n$, which satisfy the conditions of Theorem* 1.

Assume that a probability space $\mathscr{P} = (\Omega, \Sigma, P)$ is given, as well as a random variable $\bar{X}$ on $\mathscr{P}$, a process $W = \{W_t : t \in I\}$ (where $I \subseteqq \mathbb{R}$ is an interval), and an "initial time" $t_0 \in I$. Finally, assume that W has continuous sample paths. Then there exist measurable functions $T_1, T_2 : \Omega \to \bar{I}$ such that $T_1(\omega) < t_0 < T_2(\omega)$ for almost all ω, with the property that

(1) *There is a unique maximal solution $\{X_t\}$ of $dX = f(X)\,dt + g(X)\,dW$, $X_{t_0} = \bar{X}$, defined between T_1 and T_2;*

(2) *If $\{X_t'\}$ is another solution defined between T_1' and T_2', then $T_1 \leqq T_1'$, $T_2' \leqq T_2$ and $\{X_t\}$ extends $\{X_t'\}$;*

(3) *For almost all ω, either*

(i) *$T_2(\omega) = \sup I$ or*

(ii) *$\{X_t(\omega) : t_0 \leqq t < T_2(\omega)\}$ is not contained in a compact subset of U.*

Also, either $T_1(\omega) = \inf I$ or $\{X_t(\omega) : T_1(\omega) < t < t_0\}$ is not contained in a compact $K \subseteqq U$.

PROOF. Let A be such that $P(A) = 1$ and that w_ω is continuous for $\omega \in A$. For each $\omega \in A$, let $J_\omega \subseteqq I$ be the interval on which the maximal solution x_ω of $dx = f(x)\,dt + g(x)\,dw_\omega$, $x(t_0) = \bar{X}(\omega)$ is defined. Put $T_1(\omega) = \inf J_\omega$, $T_2(\omega) = \sup J_\omega$, $X_t(\omega) = x_\omega(t)$ for $T_1(\omega) < t < T_2(\omega)$. We will show later that T_1 and T_2 are measurable. Assuming this, it is clear that $\{X_t\}$ is a solution of (st) defined between T_1 and T_2. Moreover, it is also clear that, if $\{X_t'\}$ is any other solution defined between T_1' and T_2', then $\{X_t\}$ is an extension of $\{X_t'\}$, so $\{X_t\}$ is maximal.

This applies in particular if $\{X_t'\}$ is maximal, so $\{X_t\}$ is unique. If $\omega \in A$, and if $T_2(\omega) < \sup I$, then the set $\{X_t(\omega): t_0 \leqq t < T_2(\omega)\}$ cannot be contained in a campact subset of Ω for, if it were, then the solution x_ω would be extendable to an interval $(T_1(\omega), T_2(\omega) + \varepsilon)$, contradicting the fact that x_ω is maximal (cf. Corollary 4). A similar reasoning shows that either $T_1(\omega) = \inf I$ or the set $\{X_t(\omega): T_1(\omega) < t < t_0\}$ is not contained in a compact set. So all the conclusions have been established, except for the measurability of T_1 and T_2. We prove that T_2 is measurable (the proof for T_1 is similar). Take a sequence $\{K_m : m = 1, 2, \cdots\}$ of compact sets such that $K_m \subseteqq$ interior of K_{m+1}, $K_m \subseteqq U$, and $\bigcup_{m=1}^\infty K_m = U$. For each m, choose a C^∞ function ϕ_m which equals 1 on a neighborhood of K_m, and which vanishes in the complement of a compact subset of U. Define $f_m : \mathbb{R}^n \to \mathbb{R}^n$ by $f_m(x) = \phi_m(x) f(x)$ for $x \in U$, $f_m(x) = 0$ for $x \notin U$. Define g_m similarly. Now f_m and g_m satisfy the hypotheses of Theorem 2, so that for each ω there is a solution $x_\omega{}^m$ of $dx = f_m\,dt + g_m\,dw_\omega$, $x(t_0) = \bar{X}(\omega)$, defined for all $t \in I$. Let $T_2{}^m(\omega)$ be the supremum of those $t \in I$, $t \geqq t_0$ such that $x_\omega{}^m(\tau) \in K_m$ for $t_0 \leqq \tau < t$. Since $f_m \equiv f \equiv f_{m+1}$, $g_m \equiv g \equiv g_{m+1}$ in a neighborhood of K_m, it follows from Lemma 3 that $x_\omega{}^m(t) = x_\omega(t) = x_\omega{}^{m+1}(t)$ for $t_0 \leqq t < T_2{}^m(\omega)$. It is clear that $T_2{}^m(\omega) \leqq T_2{}^{m+1}(\omega)$, and therefore $\tilde{T}_2(\omega) = \lim_{m\to\infty} T_2{}^m(\omega)$ exists. We show that $\tilde{T}_2(\omega) = T_2(\omega)$. First, we observe that the solution $t \to x_\omega(t)$ clearly exists for $t_0 \leqq t < \tilde{T}_2(\omega)$, so that $\tilde{T}_2(\omega) \leqq T_2(\omega)$. If $\tilde{T}_2(\omega) < T_2(\omega)$, then we would be able to find an m such that the compact $\{x_\omega(t): t_0 \leqq t \leqq \tilde{T}_2(\omega)\}$ is contained in K_m. But then there would be an $\varepsilon > 0$ such that $\{x_\omega(t): t_0 \leqq t \leqq \tilde{T}_2(\omega) + \varepsilon\}$ is contained in K_{m+1}. Since $f \equiv f_{m+1}$ and $g \equiv g_{m+1}$ near K_{m+1}, the functions x_ω and $x_\omega{}^{m+1}$ coincide on $\{t: t_0 \leqq t \leqq \tilde{T}_2(\omega) + \varepsilon\}$. Hence, in particular, $x_\omega{}^{m+1}(t) \in K_{m+1}$ for $t_0 \leqq t \leqq \tilde{T}_2(\omega) + \varepsilon$, and therefore $T_2{}^{m+1}(\omega) \geqq \tilde{T}_2(\omega) + \varepsilon$. This is a contradiction (since $\tilde{T}_2(\omega) = \sup_m \tilde{T}_2{}^m(\omega)$). So $\tilde{T}_2 \equiv T_2$. Therefore $T_2 = \lim_{m\to\infty} T_2{}^m$. The measurability of T_2 will follow, if we prove that the $T_2{}^m$ are measurable. Now $T_2{}^m(\omega) \geqq t$ if and only if $x_\omega{}^m(\tau) \in K^m$ for all τ such that $t_0 \leqq \tau \leqq t$. Since $x_\omega{}^m$ is continuous, and K^m closed, we conclude that $T^m(\omega) \geqq t$ if and only if $x_\omega{}^m(\tau) \in K^m$ for every rational $\tau \in [t_0, t)$. For each τ, the set $\{\omega : x_\omega{}^m(\tau) \in K^m\}$ is measurable, since $\tau \to x_\omega{}^m(\tau)$ is a random variable. So $\{\omega : T_2{}^m(\omega) \geqq t\}$ is a countable intersection of measurable sets, and hence measurable. This shows that $T_2{}^m$ is measurable. As explained before, the measurability of T_2 follows, and the measurability of T_1 is proved in the same way. The proof is now complete.

Definition 5. With f, g, etc., as in the statement of Theorem 7, the family $\{X_t\}$ of (not necessarily everywhere defined) random variables described in the statement is called the maximal solution of $dX = f(X)\,dt + g(X)\,dW$, $X_{t_0} = \bar{X}$. The functions T_1, T_2 are the past and future explosion times of the solution (but observe that, if $T_2(\omega) = t$, this implies that an explosion occurs at time t if $t < \sup I$, but not necessarily if $t = \sup I$).

5. Equations driven by white noise. We now study the particular case when the input W is a Wiener process. That is, we assume that the joint distribution of $W_{t_1}, \cdots, W_{t_k}$ is Gaussian for all k and all $t_1, \cdots, t_k$, and that the covariance is $\operatorname{Cov}(W_t - W_s, W_\tau - W_\sigma) = |[\sigma, \tau] \cap [s, t]|$ for $s < t$, $\sigma < \tau$ (here "$|I|$" denotes the length of the interval I). In this case there are at least two nonequivalent ways of defining what is meant by a "solution" of

(st) $$dX = f(X)\,dt + g(X)\,dW\,.$$

Both ways proceed by replacing (st) by its equivalent integral form:

(ist) $$X(t) = X(t_0) + \int_0^t f(X(\tau))\,d\tau + \int_0^t g(X(\tau))\,dW(\tau)\,.$$

Then X is defined to be a solution of (st) iff X satisfies (ist). For this to make sense, one has to define what the integrals mean. Here the two ways differ. The first one, due to Itô, uses Itô's definition of stochastic integrals, which has several advantages, but leads to some strange formulae, since the usual rules of calculus are not obeyed (and, consequently, identities such as $\int W\,dW = W^2/2$ are not true).

The second way uses the Fisk–Stratonovich integral. Though this integral may be objectionable on the grounds that it fails to have some desirable properties, it has a fundamental advantage: it obeys the rules of calculus. For this reason, it is natural to use here the Fisk–Stratonovich integral. We shall not repeat here how this integral is defined (cf. Fisk [2], Stratonovich [6]) but we shall state those properties of the integral that will be needed. They are:

(1) $$\int_{t_0}^{t_1} dW(t) = W(t_1) - W(t_0)$$

and

(2) the fact that the rules of calculus hold. Precisely, suppose that

$$\begin{aligned} Y_j(t) = Y_j(t_0) &+ \int_{t_0}^t f_j(Y_1(\tau), \cdots, Y_k(\tau))\,d\tau \\ &+ \int_{t_0}^t g_j(Y_1(\tau), \cdots, Y_k(\tau))\,dW(\tau) \end{aligned}$$

(i.e., formally $dY_j = f_j\,dt + g_j\,dW$) and that

$$X(t) = h(Y_1(t), \cdots, Y_k(t))\,.$$

Then

$$X(t) = X(t_0) + \int_{t_0}^t k(Y_1(\tau), \cdots, Y_k(\tau))\,d\tau + \int_{t_0}^t l(Y_1(\tau), \cdots, Y_k(\tau))\,dW(\tau)$$

where

$$k(y_1, \cdots, y_k) = \sum_{j=1}^k \left(\frac{\partial h}{\partial y_j} f_j\right)(y_1, \cdots, y_k)$$

$$l(y_1, \cdots, y_k) = \sum_{j=1}^k \left(\frac{\partial h}{\partial y_j} g_j\right)(y_1, \cdots, y_k)$$

(formally: $dX = \sum_j (\partial h/\partial y_j)\,dY_j$).

THEOREM 8. *Assume that all the conditions of Theorem 5 are satisfied and that, in addition, W is a Wiener process. Then the solution in the sense of Definition* 3 *of* $dX = f(X)\,dt + g(X)\,dW$ *with initial condition* $X(t_0) = \bar{X}$ *is precisely the same as the Stratonovich solution.*

PROOF. If f and g satisfy the special assumptions (I), (II), (III) of the proof of Theorems 1, 2 (so that, in particular, $g(x) \equiv v$ is a constant), then the solution in our sense of $dX = f\,dt + g\,dW$ satisfies

$$X(t) = \bar{X} + \int_{t_0}^{t} f(X(\tau))d + [W(t) - W(t_0)]v\,.$$

Since $\int dW = W$, we conclude that our X satisfies the equation in the Stratonovich sense. For general f and g, we use the fact that one can find $\hat{f}$, $\hat{g}$ in $\mathbb{R}^{n+1}$ which satisfy the special assumptions (I), (II), (III), and a map

$$F\colon \mathbb{R}^{n+1} \to \mathbb{R}^n$$

such that $F * \hat{f} = f$, $F * \hat{g} = g$ and that $F(0, x) = x$. (This map is constructed in the proof of Theorems 1 and 2.)

For ω in the probability space, put $\hat{\bar{X}}(\omega) = (0, \bar{X}(\omega))$. Then $\bar{X} = F \circ \hat{\bar{X}}$. Let X_s be the Stratonovich solution of $dX = f(X)\,dt + g(X)\,dW$, $X(0) = \bar{X}$, and let $\hat{X}_s$ be the Stratonovich solution of $dX = \hat{f}(X)\,dt + \hat{g}(X)\,dW$, $\hat{X}_s(0) = \hat{\bar{X}}_s$. Since the rules of calculus are obeyed, $X_s = F \circ \hat{X}_s$. Our solution X of $dX = f(X)\,dt + g(X)\,dW$ is constructed precisely as $X = F \circ \hat{X}$, where $\hat{X}$ is the solution, in our sense, of $dX = \hat{f}(X)\,dt + \hat{g}(X)\,dW$ (cf. the proof of Theorems 1, 2). As we have already pointed out $\hat{X} = \hat{X}_s$. So $X = X_s$, and our proof is complete.

We leave it to the reader to state and prove an analogue of Theorem 8 for the case when f and g do not satisfy a linear growth condition, and therefore X may have explosions.

6. Two applications. It has been shown by Wong and Zakai that the solution X of an equation

$$\text{(st)} \qquad dX = f(X)\,dt + g(X)\,dW\,, \qquad X(t_0) = \bar{X}$$

driven by a Wiener process W is the limit, in an appropriate sense, of the solutions X^m of

$$(\text{st}_m) \qquad dX = f(X)\,dt + g(X)\,dW^m\,, \qquad X(t_0) = \bar{X}\,,$$

if the W^m are more regular processes which converge to W of $m \to \infty$. Here we shall show how this result follows trivially in our framework. Actually, there is no need to limit ourselves to a Wiener process W.

Let $W = \{W_t\}_{t \in I}$ be a process with continuous sample paths, defined on the probability space $\mathscr{P} = (\Omega, \Sigma, P)$. Let $\{W^m\}_{m=1}^{\infty}$ be a sequence of processes defined on $\mathscr{P} = (\Omega, \Sigma, P)$, with $W^m = \{W_t^{\,m}\}_{t \in I}$. Assume that W^m converges to W in the sense that, for almost every $\omega \in \Omega$, the function $t \to W_t^{\,m}(\omega)$ converges to $t \to W_t(\omega)$ as $m \to \infty$, uniformly for t in any compact subinterval of I. Let f, g be vector fields on $\mathbb{R}^n$, which satisfy the hypotheses of Theorem 2. Let $\bar{X}$ be a random variable on $\mathscr{P}$. For each m, let X^m be the solution of (st_m). Let X be the solution of (st). Then

THEOREM 9. *$X^m \rightarrow X$ as $m \rightarrow \infty$, in the sense that, for almost every $\omega \in \Omega$, the functions $t \rightarrow X_t^m(\omega)$ converge to $t \rightarrow X_t(\omega)$ as $m \rightarrow \infty$, uniformly on compact subintervals of I.*

PROOF. Just apply our definition of solution.

REMARK. If W is a Wiener process, then we know that the solution in our sense coincides with Stratonovich solution (Theorem 8). Let the W^m be "regularized versions of W" (e.g., let $\{\Pi^m\}$ be a sequence of partitions of I, such that mesh $(\Pi^m) \rightarrow 0$ as $m \rightarrow \infty$, and let the function $t \rightarrow W_t^m(\omega)$ coincide with $t \rightarrow W_t(\omega)$ at the partition points, and be linear on each partition interval). Then (st_m) can be solved, for each m, in the ordinary sense (i.e., for each $\omega \in \Omega$). Theorem 9 then shows that the Stratonovich solution of (st) is one that arises when W (which is a mathematical idealization) is regarded as the limit of the physically more realistic processes W^m. As indicated by Wong and Zakai in [11], by Wong in [10], and by McShane in [5], this shows that the Stratonovich concept of a solution of (st) is the most adequate one for the purpose of modelling "systems driven by white noise."

As a second application, we improve upon a result of Stroock and Varadhan (cf. [7], Theorem 5.1 and Remark 5.1). Let f, g satisfy the hypotheses of Theorem 2, let $\Omega = C([a, b])$ and let P be Wiener measure on Ω. Let $x_0 \in \mathbb{R}^n$ be fixed, and let $X = \{X_t \colon t \in [a, b]\}$ be the Stratonovich solution of $dX = f(X)\, dt + g(X)\, dW$, $X_a \equiv x_0$, where $W = \{W_t \colon t \in [a, b]\}$, $W_t(w) = w(t)$. Consider the transformation $\hat{T}\colon C^0([a, b], \mathbb{R}) \rightarrow C^0([a, b], \mathbb{R}^n)$ defined by $\hat{T}(w)(t) = X_t(w)$. Let T be the restriction of $\hat{T}$ to $C^2([a, b], \mathbb{R})$. Stroock and Varadhan show that $\hat{T}$ is an extension of T with the property that all the elements of $C^2([a, b], \mathbb{R})$ are continuity points of $\hat{T}$. (Their definition of "continuity point" need not be given here. However, this definition is so chosen that any point where $\hat{T}$ is continuous in the ordinary sense is a continuity point.)

Our results imply that *$\hat{T}$ is actually continuous* and that it *is the only continuous extension* of T to $C^0([a, b], \mathbb{R})$.

7. Several inputs. For equations

$$dx = f(x)\, dt + \textstyle\sum_{i=1}^{k} g_i(x)\, dw_i(t)$$

with $k > 1$, the results of the preceding sections are *not* valid. The implications of this fact are quite interesting, and will be pursued elsewhere. Here we will limit ourselves to some heuristic remarks, and to a simple example which will show what may happen when $k > 1$.

The main obstruction to the validity of our results for $k > 1$ is the *noncommutativity* of the vector fields g_i. If the g_i commute (i.e., if all the Lie brackets $[g_i, g_j] = (Dg_i) \cdot g_j - (Dg_j)g_i$ vanish), then a straightforward modification of our proofs for $k = 1$ will work in general. If the g_i do not commute, new phenomena appear. Heuristically, one should think of each input $u_i = dw_i/dt$ as representing an "infinite sequence" of "kicks" separated by infinitesimal time intervals. The effect of these kicks on the state x is determined by the vector fields g_i. The reason why our results fail is that the order in which the kicks are applied within each infinitesimal interval matters. The functions w_i only measure the cumulative effect of the i-kicks. It is possible, for instance, for two "sequences u_i^1, u_i^2 of kicks" to be such that their cumulative effect on any finite interval is the same, while the ordering of the kicks within infinitesimal intervals is different, in such a way that a visible "macroscopic effect" appears. The following example illustrates this. Let A, B, be two $n \times n$ matrices such that $AB - BA = C \neq 0$.

For $x \in \mathbb{R}^n$ (written as a column), put $g_1(x) = Ax$, $g_2(x) = Bx$.

Let the interval $[0, 1]$ be partitioned into n equal intervals $I_j^n = [(j-1)/n, j/n]$, $j = 1, \cdots, n$. Partition each I_j^n into four equal intervals $I_{j,1}^n$, $i = 1, 2, 3, 4$. Define $u_1(t)$ to be equal to $4n^{\frac{1}{2}}$ for $t \in I_{j,1}^n$, to $-4n^{\frac{1}{2}}$ for $t \in I_{j,3}^n$, and to zero for all other t. Similarly, let u_2 be equal to $4n^{\frac{1}{2}}$ for $t \in I_{j,2}^n$, to $-4n^{\frac{1}{2}}$ for $t \in I_{j,4}^n$, and to zero for all other t. Let w_1^n, w_2^n be the indefinite integrals of u_1, u_2 (chosen so that $w_1^n(0) = w_2^n(0) = 0$). It is easy to see that w_1^n and w_2^n converge to zero uniformly, as $n \to \infty$. On the other hand, the solution x^n of

$$dx = g_1(x)\,dw_1^n + g_2(x)\,dw_2^n\,, \qquad x(0) = x_0$$

has the limit

$$x(t) = e^{t[B,A]}x_0$$

as $n \to \infty$.

8. A counterexample. We show, by means of a counterexample, that the assumption that Dg is uniformly bounded cannot be replaced, in the statement of Theorem 2, by the weaker hypothesis that g grows linearly.

Let $a_1, \cdots, a_n, \cdots$ and $b_1, \cdots, b_n, \cdots$ be sequences of real numbers such that

$$0 < a_1 \tag{1}$$

$$a_n < b_n < a_{n+1} \quad \text{for all} \quad n \tag{2}$$

$$\lim_{n\to\infty} a_n = \infty \tag{3}$$

$$\textstyle\sum_{n=1}^{\infty} (b_n - a_n) < \infty\,, \qquad \text{and} \tag{4}$$

$$a_n - b_{n-1} > a_{n+1} - b_n \quad \text{for all} \quad n\,. \tag{5}$$

(For instance, let $a_1 = 1$, $b_n = a_n + 2^{-n}$, $a_{n+1} = b_n + 1/n$.)

Let $t_1, \cdots, t_n, \cdots$ be a sequence of strictly positive numbers, such that

$$\textstyle\sum_{n=1}^{\infty} t_n < \infty\,.$$

Let $\phi : \mathbb{R} \to \mathbb{R}$ be a C^∞ function such that $\phi(x) = -1$ for $x \leqq 0$, $\phi(x) = 1$ for $x \geqq 1$, $\phi(\frac{1}{2}) = 0$, and that ϕ is strictly increasing for $0 \leqq x \leqq 1$.

Define $g : \mathbb{R} \to \mathbb{R}$ as follows. For $x \leqq a_1$, we let $g(x) = 1$. For $b_1 \leqq x \leqq a_2$, we let $g(x) = -1$. In general, for $b_n \leqq x \leqq a_{n+1}$, we let $g(x) = (-1)^n$. To complete the construction of g, we define it on the intervals $a_n \leqq x \leqq b_n$ by "interpolation," using ϕ. Precisely, we put $g(x) = (-1)^n\phi((x - a_n)/(b_n - a_n))$ for $a_n \leqq x \leqq b_n$.

Notice that g is C^∞ and uniformly bounded (and hence, in particular, it satisfies a linear growth condition) but that dg/dx is unbounded.

Now define $f\colon \mathbb{R} \to \mathbb{R}$ by $f(x) = 1 - g(x)^2$. Then f is C^∞, and uniformly bounded. Moreover, f vanishes everywhere except on the intervals $a_n < x < b_n$.

It is clear that f and g satisfy all the hypotheses of Theorem 2 except for the fact that dg/dx is unbounded. We now find a $T > 0$ and a continuous function $w : [0, T] \to \mathbb{R}$ such that the solution $x(t)$ of $dx = f(x)\,dt + g(x)\,dw$, $x(0) = 0$, is defined for $0 \leqq x < T$, but that $\lim_{t\to T} x(t) = +\infty$.

Define, inductively,

$$T_0 = 0$$

$$T_n = T_{n-1} + t_n + (b_n - a_n)$$

so that

$$T_n = \textstyle\sum_{k=1}^n t_k + \sum_{k=1}^n (b_k - a_k)\,.$$

It is clear that $0 < T_1 < T_2 < \cdots < T_n < \cdots$ and that

$$T = \lim_{n\to\infty} T_n$$

exists and is finite.

Now define a function $u : [0, T) \to \mathbb{R}$ as follows. Let $0 \leqq t < T$. Then there is a unique n such that $T_{n-1} \leqq t < T_n$. We distinguish two cases:

(1) $$T_{n-1} \leqq t < T_{n-1} + t_n\,.$$

In this case we define

$$u(t) = (-1)^{n-1} \frac{a_n - b_{n-1}}{t_n}\,;$$

(2) $$T_{n-1} + t_n \leqq t \leqq T_{n-1} + t_n + (b_n - a_n) = T_n\,.$$

In this case, we let $u(t) = g(t - T_{n-1} - t_n + a_n)$.

We now study the solution $x(t)$ of

$$\frac{dx}{dt} = f(x) + u(t)\,g(x)\,.$$

We show by induction on n that this solution is defined on $[0, T_n]$ and that $x(T_n) = b_n$ (letting $b_0 = 0$). This conclusion is certainly true for $n = 0$. Assume that it is true for $n - 1$. For $T_{n-1} \leqq t < T_{n-1} + t_n$, let $\xi(t) = b_{n-1} + (a_n - b_{n-1})(t - T_{n-1})/t_n$.

Then $b_{n-1} \leqq \xi(t) \leqq a_n$, so that $f(\xi(t)) = 0$, $g(\xi(t)) = (-1)^{n-1}$. On the other hand, $u(t) = (-1)^{n-1}((a_n - b_{n-1})/t_n)$, so that $f(\xi(t)) + u(t)g(\xi(t)) = (a_n - b_{n-1})/t_n$. Therefore

$$\frac{d\xi}{dt}(t) = f(\xi(t)) + u(t)g(\xi(t))$$

for $T_{n-1} \leqq t < T_{n-1}$. Moreover, $\xi(T_{n-1}) = b_{n-1}$ and, by the inductive hypothesis, $x(T_{n-1})$ exists and equals b_{n-1}. Hence $x(t)$ exists for $T_{n-1} \leqq t \leqq T_{n-1} + t_n$ and it is equal to $\xi(t)$. In particular, $x(T_{n-1} + t_n)$ exists and equals a_n.

Now let $T_{n-1} + t_n \leqq t < T_n$. For such a t, put $\eta(t) = t - T_{n-1} - t_n + a_n$. Then $a_n \leqq \eta(t) \leqq b_n$. By the definition of u:

$$u(t) = g(\eta(t)) .$$

Since $a_n \leqq \eta(t) \leqq b_n$, we have $f(\eta(t)) = 1 - g(\eta(t))^2 = 1 - u(t)g(\eta(t))$. Hence $f(\eta(t)) + u(t)g(\eta(t)) = 1$. So $(d\eta/dt)(t) = f(\eta(t)) + u(t)g(\eta(t))$. Since $\eta(T_{n-1} + t_n) = a_n$, we conclude that $x(t)$ exists for all t such that $T_{n-1} + t_n \leqq t < T_n$, and that $x(t) = \eta(t)$ for such t. Hence $x(T_n)$ exists and equals b_n. The induction is complete.

Since $\lim_{n\to\infty} b_n = +\infty$, it follows easily that

$$\lim_{t\to T} x(t) = +\infty .$$

On the other hand, the function $t \to x(t)$ is the solution of

$$dx = f(x)\,dt + g(x)\,dw , \qquad x(0) = 0 ,$$

where

$$w(t) = \textstyle\int_0^t u(s)\,ds \qquad \text{for} \quad 0 \leqq t < T .$$

We now show that w is a continuous function on the closed interval $[0, T]$, i.e., that

$$\lim_{t\to T} w(t) = \lim_{t\to T} \textstyle\int_0^t u(s)\,ds$$

exists and is finite. To see this, observe that

$$w(T_n) = w(T_{n-1}) + (-1)^{n-1}(a_n - b_{n-1}) + (b_n - a_n) .$$

Hence

$$w(T_n) = \textstyle\sum_{k=0}^{n} (-1)^{k-1}(a_k - b_{k-1}) + \sum_{k=0}^{n} (b_k - a_k) .$$

Since both series $\sum (-1)^{k-1}(a_k - b_{k-1})$ and $\sum (b_k - a_k)$ converge—the former because it is an alternating series whose terms decrease in absolute value, by (5), and the latter because of (4)—it follows that $\lim_{n\to\infty} w(T_n)$ exists. The conclusion that $\lim_{t\to T} w(t)$ exists follows easily.

Acknowledgments. The author wishes to thank W. Xotus, S. Varadhan, H. McKean, D. W. Stroock, D. Elliott and A. Kennedy for their helpful comments.

REFERENCES

[1] Elliot, D. (1969). Controllable nonlinear systems driven by white noise. Unpublished doctoral dissertation, U.C.L.A.

[2] Fisk, D. L. (1963). Quasi-martingales and stochastic integrals. Michigan State Univ. Dept. of Statistics, Technical Report No. 1.

[3] Lamperti, J. (1964). A simple construction of certain diffusion processes. *J. Math. Kyoto* **4** 161–170.

[4] McKean, H. P. (1969). *Stochastic Integrals*. Academic Press, New York.

[5] McShane, E. J. (1972). On the use of stochastic differential equations in models of random processes. *Proc. Sixth Berkeley Symp. Math. Statist. Prob.* **3** 263–294.

[6] Stratonovich, R. L. (1966). A new form of representation of stochastic integrals and equations. *SIAM J. Control* 362–371.

[7] Stroock, D. W. and Varadhan, S. R. S. (1972). On the support of diffusion processes with applications to the strong maximum principle. *Proc. Sixth Berkeley Symp. Math. Statist. Prob.* **3** 333–359.

[8] SUSSMANN, H. J. (1976). Semigroup representations, bilinear approximation of input-output maps, and generalized inputs. In *Mathematical Systems Theory*, *Udine* 1975, 172–191. Springer-Verlag, New York.

[9] SUSSMANN, H. J. (1977). An interpretation of stochastic differential equations as ordinary differential equations which depend on the sample point. *Bull. Amer. Math. Soc.* **83** 296–298.

[10] WONG, E. (1971). *Stochastic Processes in Information and Dynamical Systems*. McGraw-Hill, New York.

[11] WONG, E. and ZAKAI, M. (1966). On the relationship between ordinary and stochastic differential equations and applications to stochastic problems in control theory. *Proc. Third IFAC Congress*, paper 3B.

[12] WONG, E. and ZAKAI, M. (1969). Riemann-Stieltjes approximations of stochastic integrals. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **12** 87–97.

DEPARTMENT OF MATHEMATICS
HILL CENTER FOR THE MATHEMATICAL SCIENCES
RUTGERS UNIVERSITY
NEW BRUNSWICK, NEW JERSEY 08903