PH.D. THESIS

Controllable Nonlinear Systems Driven by White Noise

by David L. Elliott
Advisor:

PhD 2004-1

UNIVERSITY OF CALIFORNIA

Los Angeles

Controllable Nonlinear Systems Driven by White Noise

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy in Engineering

by

David LeRoy Elliott

Committee in charge:

Professor A. V. Balakrishnan, Chairman

Professor Hector O. Fattorini

Professor Richard E. Mortensen

Professor Charles J. Stone

Professor Paul K.C. Wang

1969

The dissertation of David LeRoy Elliott is approved, and it is acceptable in quality and form for publication on microfilm:

Charles J. Stone

W. Fattom

P. K. C. Wang

Richard E. Mortensen

A. V. Balakrishnan

Committee Chairman

University of California, Los Angeles

1969

To Kiyoko

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

The author wishes to thank Professor A. V. Balakrishnan for suggesting this area of research, for many helpful suggestions during all phases of the research and writing, and for arranging financial support at a crucial time.

Thanks are also due to Professors Hector O. Fattorini, Richard E. Mortensen, Charles J. Stone, and Paul K.C. Wang for serving on my committee. Particularly appreciated were Professor Mortensen's useful questions and criticisms of the work-in-progress.

Early parts of this research were performed while the author was supported by a Navy Fellowship from the U.S. Naval Undersea Research and Development Center; the work was completed during the summer of 1969 at the Jet Propulsion Laboratory of the California Institute of Technology.

VITA

May 29, 1932-- Born, Cleveland, Ohio

1953-- B.A. cum laude, Pomona College

1953-1954-- Graduate Assistant, California Institute of Technology

1955-1969-- Mathematician, U. S. Naval Undersea Research and Development Center, Pasadena, California

1959-- M.A., University of Southern California

1969-- Acting Instructor, University of California at Los Angeles

PUBLICATIONS

Elliott, David LeRoy. "A note on systems of linear equations." SIAM Review, 3:66-69, January 1961.

Elliott, David LeRoy. "Behavior of controllable nonlinear systems driven by white noise," in Proceedings of the Second Hawaii International Conference on System Sciences, edited by B. S. M. Granborg, North Hollywood, California, Western Periodicals Co., 1969. p.859-862.

ABSTRACT OF THE DISSERTATION

Controllable Nonlinear Systems

Driven by White Noise

by

David LeRoy Elliott

Doctor of Philosophy in Engineering

University of California, Los Angeles, 1969

Professor A. V. Balakrishnan, Chairman

Given a time-invariant dynamic system S: $dx/dt = a_o(x) + a_1(x)u(t)$ ($t \geq 0$, x in euclidean n-space, u a square-integrable scalar control) with known properties, we wish to obtain analogous properties of the related Markov process Σ obtained (via an Itô equation) when u is replaced by a gaussian white noise process. The differential generator of Σ is $D = a_o(x)\cdot\nabla + (1/2)(a_1(x)\cdot\nabla)^2$, which is degenerate-elliptic in form; the classical existence, uniqueness, and smoothness theorems do not apply to the Cauchy problem $\partial g/\partial t = Dg$, $g(x,0) = f(x)$ (the backward Kolmogorov equation for Σ).

The chief property of S considered here is controllability; another is (zero-input) stability. S is said to be $\hat{T}$-controllable if there exists a control u connecting any given initial state y and terminal state x in any time greater than $\hat{T}(y,x) \geq 0$, some finite function of x and y.

If S is real-analytic and globally Lipschitzian, it is shown that the controls can be taken to be real-analytic in t without changing the reachable sets. Differential geometry techniques can then be used to show that T-controllability implies that the Lie algebra,generated by $-\partial/\partial t + a_o\cdot\nabla$ and $a_1\cdot\nabla$ and their commutators, has rank $n + 1$.

Therefore, by a theorem of Hormander, the operator $-\partial/\partial t + D$ is hypoelliptic; solutions of the Cauchy problem are smooth for $t > 0$. It can then be shown (using probabilistic arguments) that if f is a bounded Borel function, the unique bounded solution of the Cauchy problem is $g(x,t) = E_x f(\underline{x}_t)$, where $\underline{x}_t$, $t \geq 0$, is a sample path of Σ and E_x is the expectation conditioned on $\underline{x}_o = x$.

It is shown that if S is $\hat{T}$-controllable, P( ) is the transition function of Σ, and G is an open set, then $P(x, \Theta_i, G) > 0$ for a sequence of times $\Theta_i(x,G) \uparrow \infty$, $\inf[\hat{T}(y,x): x \text{ in } G] = \Theta_1$.

The analogy to stability in question is the existence of an invariant probability measure for Σ; we obtain this result if a_1 is bounded and V is a Lyapunov function of quadratic growth such that $a_o(x) \cdot \nabla V(x) < - cV(x)$.

CHAPTER 1. INTRODUCTION

(1.1) The problem.

Consider a dynamic system S given by the differential equation

a) $\frac{dx}{dt} = a_o(x) + a_1(x)u(t), \quad t \geq 0,$

where x belongs to the n-dimensional euclidean state space R^n and the input u belongs to a suitable class Ω of control functions. Under suitable hypotheses, Eq.(a) has a unique solution x_t called the response of S to u. A great deal is known about such systems; very little is known about their stochastic analogs when Ω is taken to be a stochastic process.

Our studies have had the goal of predicting from properties of S some analogous properties of the stochastic system Σ obtained when the input u is taken to be a "gaussian white noise" process (Chapter 3) and Eq.(a) is replaced with an appropriate Ito equation (Eq. 3.2a). Σ turns out to be a Markov process with differential generator

$D = a_o \cdot \nabla + (1/2)(a_1 \cdot \nabla)^2.$

If S is linear, $a_o(x) = Ax$ (A a constant matrix) and $a_1(x) = b$ is constant. It is well-known that S is controllable if and only if the set of vectors $(b, Ab, \ldots, A^{n-1}b)$ is linearly independent (see 2.12d). An equivalent criterion is that the matrix

$M(t) = \int_o^t e^{As}bb'e^{A's}ds$

is positive definite. If u is replaced by a gaussian white noise, the response $\underline{x}_t$ is a gaussian Markov process with mean $e^{At}x_o$ and variance matrix $M(t)$. Controllability of S then implies that Σ is non-degenerate and has a smooth transition density. If A is stable, as $t \to \infty$ the mean approaches zero and $M(t) \to M(\infty)$ (unique and finite), the variance of an invariant probability measure for Σ (which characterizes a stationary ergodic process on R^n).

We have been able to extend portions of that theory to real-analytic, globally Lipschizian, nonlinear systems, obtaining a necessary criterion for controllability and applying it to stochastic systems. Since the differential generator D is not elliptic, classical methods of studying diffusion processes do not apply; our chief tool in applying controllability is a remarkable new theorem of Hörmander, analogous to Weyl's Lemma. Its hypothesis is our criterion. Finally, we obtain a new sufficient condition for the existence of an invariant probability measure for Σ, as well as a proof that open sets have positive hit probability at a sequence of times approaching infinity.

(1.2) Comparison with previous work.

The linear gaussian system discussed above has a history going back to some work of Kolmogorov about 1935. Since n-th order time-invariant linear differential equations (with u on the RHS) are controllable systems of a rather general type, the thesis of [Dym] will serve to present the known results (references to works in our Bibliography are given in square brackets []).

In the case that Σ has n independent white noise inputs, the nonlinear problem has been attacked with much success by [Wonham], using the results of [Khasminskii] on invariant measures and the theory of elliptic operators. Our original intent was to give a parallel treatment [Elliott] for $r < n$ inputs, but the present work is independent of [Wonham]. For general material on diffusions, see [Dynkin] and [McKean].

If there are $r < n$ inputs, the operator D is formally degenerate, Such systems arise in the study of communication [Viterbi] and control [Kushner 1967] problems. [Wonham] suggested that controllability, rather than ellipticity, might be enough to yield the strong Feller property; [Kushner 1969] is the first attack on the invariant-measure problem using this idea. Kushner's systems are of the form (r inputs)

$$dx/dt = Ax + Bg(x) + Bu(t), \quad g(\cdot) \text{ bounded}, \quad B \text{ an } n \times r \text{ matrix},$$

the system with $g = 0$ being assumed controllable. Our results on the Kolmogorov equations improve on his in the case that g is real-analytic. (Real-analytic, C^{ω}, functions have at every point a Taylor series convergent in a neighborhood of that point. C^k means "having k continuous derivatives" and C^{∞} means all derivatives are continuous, but neither of these implies that the function or its derivatives are bounded on R^n. C_o^k means C^k with compact support and does imply boundedness.)

We have stated our results for the case $r = 1$ for notational reasons -- but they carry over to $r \geq 2$ <u>except</u> for Theorem 3.6 ; see Note 2 of Section 3.6 , for an illuminating example due to Mortensen.

Our results on invariant probability measures in Chapter 4 are not as complete as Kushner's, but avoid using the results of [Khasminskii], and do not require that we first prove positivity of the process Σ .

We know of no precursor of Theorem 3.6, an application of our notion of $\hat{T}$-controllability to establish positive hit probabilities for open sets. (It also "embeds" S in Σ.) We have not yet been able to show the relation of $\hat{T}$-controllability to strong connectedness [Balakrishnan] but the examples of Section 2.12 may serve to motivate it. The fact that $\hat{T}$-controllability implies H-controllability and the proof (2.8 and 2.10) are new; H-controllability as a <u>definition</u> is due to [Hermann 1963].

H-controllability is a rank condition on a Lie algebra associated with S, and has also been discussed by [Haynes] and [Hermes] (it is the criterion mentioned at the end of Section 1.1).

CHAPTER 2. ACCESSIBILITY AND CONTROLLABILITY

(2.1) Definitions.

The dynamic system S is given on the n-dimensional euclidean state-space R^n by the differential equation

a) $\frac{dx}{dt} = a_o(x) + a_1(x)u(t)$

where the vector functions $a_o(\cdot)$ and $a_1(\cdot)$ are C^ω (real-analytic) on R^n, and for all x and y in R^n they are globally Lipschitzian, i.e.

b) $||a_i(x)||^2 < K^2(1+||x||^2)$, $||a_i(x) - a_i(y)|| < K||x-y||$, $i= 0,1$.

The scalar $u(\cdot)$ is a real control function to be chosen from the function-space $\underline{L}^2 \overset{\Delta}{=} \{u: \int_0^\infty |u(t)|^2 dt < \infty\}$.

Given any initial state y and control $u(\cdot)$, defined for $t \geq 0$, Eq.(a) has a unique absolutely continuous solution $x_t \overset{\Delta}{=} X(y,t;u)$, $t \geq 0$ such that $x_o = y$; it is called the response of S to u starting at y. Note that S is time-invariant.

In Sections 2.4 and 2.8 we shall be concerned with the notion of accessibility. For $T > 0$, let

$A^y(T) \overset{\Delta}{=} \{x$: there exists $t \in [0,T]$ and $u \in \underline{L}^2$:: $X(y,t;u) = x\}$;

$A^y \overset{\Delta}{=} \bigcup_{T>0} A^y(T)$.

A^y is called the set of states accessible from y. If for every $y \in R^n$ the set A^y contains an open set, we say S has the accessibility property. The next Proposition shows that in the definition of $A^y(T)$ we can replace the class $\underline{L}^2$ with the class of controls that are real-analytic on finite intervals (the Taylor series converge at the endpoints).

(2.2) Proposition.

If for $u \in \underline{L}^2$ we have $X(y,T; u) = z$, then there exists $v \in C^\omega[0,T]$ such that $X(y,T; v) = z$ and $X(y, \cdot; v)$ is real-analytic on $[0,T]$.

Proof: Let $J(u) \overset{\Delta}{=} \int_0^T |u(t)|^2 dt$, and denote $\rho \overset{\Delta}{=} J(\tilde{u}) + 1$. We pose the "Lagrange problem" of minimizing the functional $J(\cdot)$ under the constraints $J(u) \leq \rho$, $X(y,T;\, u) = z$. The condition (2.1b) guarantees (via the Gronwall-Bellman inequality) that if $J(u) \leq \rho$, then the response of S to u satisfies the a priori bound

$$\int_0^T \left|\left|\frac{dx}{dt}t\right|\right|^2 dt \leq 2||y||^2\left(e^{3K^2T(T+\rho)} - 1\right) + 2K^2T(T+\rho)e^{3K^2T(T+\rho)}.$$

Then we may apply Theorem 6 of [Cesari] to show the existence of an optimal control v and response $\hat{x}$ that minimize J under the given constraints. However, any such optimal control and response must satisfy the Pontryagin necessary conditions [Lee and Markus]: there exists an n-component vector q of "multiplier functions" defined on [0,T] such that ($\cdot$ indicates inner product)

$$v(t) = -(1/4)q(t)\cdot a_1(\hat{x}_t)$$

$$\frac{dq}{dt} = -\frac{\partial}{\partial x}\left[\, q\cdot a_0(\hat{x}) - (1/4)(q\cdot a_1(\hat{x}))^2\right]$$

$$\frac{d\hat{x}}{dt} = a_0(\hat{x}) - (1/2)(\, q\cdot a_1(\hat{x})\,)a_1(\hat{x}).$$

These differential equations are C^ω in $(\hat{x},q)$ and (Theorem 8.2 of Ch. 1, [Coddington and Levinson]) the known solution is real-analytic on [0,T]; from the first equation, $v \in C^\omega[0,T]$. Also, the response $\hat{x}_t$ lies in the set $A^y(T)$.

In the next sections we will examine accessibility from the standpoint of differential geometry, guided by the presentation in [Hermann], especially his Chapters 8 and 18.

(2.3) Vector fields.

$\alpha_0 = a_0(x)\cdot(\partial/\partial x)$, $\alpha_1 = a_1(x)\cdot(\partial/\partial x)$, a_0 and a_1 being the C^ω coefficients of Eq.(2.1a), are scalar first-order partial differential operators; but they have another useful interpretation, as examples of vector fields on R^n. The class $\underline{V}$ of C^ω vector fields is thus the set of all first-order partial differential operators $\beta = b(x)\cdot(\partial/\partial x)$ which map $C^\omega(R^n)$ into itself (the derivations of the ring $C^\omega(R^n)$); note

a) $\beta(\phi_1\phi_2) = \phi_1\beta\phi_2 + \phi_2\beta\phi_1$, for ϕ_1 and ϕ_2 in $C^\omega(R^n)$.

The evaluation $\beta x = b(x)$ is a column-vector, with C^ω components.

$\underline{V}$ is closed under the operations of addition and multiplication by scalars from $C^\omega(R^n)$, so $\underline{V}$ is a C^ω-module, of dimension n. $\underline{V}$ is also closed under the commutator or Jacobi bracket operation: if $\alpha,\beta \in \underline{V}$,

b) $[\alpha,\beta] = \alpha\beta - \beta\alpha \in \underline{V}$.

Therefore $\underline{V}$ is a real (infinite-dimensional) Lie algebra.

If $\alpha \in \underline{V}$, $\alpha y = a(y)$, and if for each $y \in R^n$ there exists a (necessarily unique) trajectory in R^n of the differential equation

$$\frac{dy_t}{dt} = a(y_t), \quad y_o = y,$$

then for any C^ω function ϕ it makes sense to define the function

$$\frac{d}{dt}(\phi(y_t))\Big|_{t=0} = \alpha\phi(y),$$

the Lie derivative of ϕ along the vector field α. If also $\beta \in \underline{V}$, its Lie derivative is the vector field $[\alpha,\beta]$ ([Hermann 1968]); that this is consistent with our other definitions can be seen from (cf. Eq.(a))

c) $\displaystyle \frac{d}{dt}(\beta\phi(y_t))\Big|_{t=0} = [\alpha,\beta]\phi(y) + \beta\alpha\phi(y).$

The derivative of a function or vector field along a response of S, given a state x_o and a C^ω control $u(\cdot)$, will be called the system derivative; the operation is denoted by SD/dt. Thus let $\psi(x,t)$ denote a function that is C^ω on $R^n\times[0,T]$, and remembering $\alpha_i = a_i(x)\cdot(\partial/\partial x)$,

d) $$\frac{SD\psi}{dt}(x_t,\ t) \triangleq \lim_{s\to 0}(1/s)(\psi(x_{t+s},t+s) - \psi(x_t,\ t)\ ) = \frac{\partial}{\partial t}\psi(x_t,t) + \alpha_o\psi(x_t,t) + u(t)\alpha_1\psi(x_t,t);$$

e) $\left.\frac{SD\beta}{dt}\right|_{t=0} = [\alpha_o,\beta] + u(0)[\alpha_1,\beta] \triangleq \beta'$.

The second derivative is

$$S\left(\frac{D}{dt}\right)^2 \beta\Big|_{t=0} = [\alpha_o+u(0)\alpha_1,\ \beta'] + \frac{du}{dt}(0)[\alpha_1,\beta],\quad \text{and so forth.}$$

(2.4) Vector field systems.

Given the two vector fields α_o, α_1, defined for system S, we define the vector field system H as their span in $\underline{V}$, that is,

$$H = \{\psi_o\alpha_o+\psi_1\alpha_1 : \psi_o \text{ and } \psi_1 \in C^{\omega}(R^n)\}.$$

The evaluation of H at x is the two-dimensional vector space

$$H_x = \{\beta x = b(x),\ \beta \in H\} = \{va_o(x) + ua_1(x),\ v \text{ and } u \text{ real}\},$$

attached to the point x. Other, similar, evaluations will occur later.

Given an interval $[0,T]$, an integral curve of H is defined as a C^{ω} map z_t from $[0,T]$ to R^n such that $\frac{dz}{dt}_t \in H_{z_t}$, $0 \le t \le T$, that is, there exist C^{ω} real functions $u(\cdot)$ and $v(\cdot)$ such that

$$\frac{dz}{dt}_t = v(t)a_o(z_t) + u(t)a_1(z_t).$$

An integral path of H is a continuous curve, piecewise C^{ω}, each of whose pieces is an integral curve. The set of points x lying on the integral paths of H that originate at point y is called the leaf L^y of H through y, in the terminology of [Hermann 1968]. By an argument parallel to Prop. 2.2, it can be shown that for system S, the leaf L^y is exactly the set of points z_t that lie on integral curves with $z_o = y$.

If we are given a control $u \in C^{\omega}[0,T]$, the response of S to u is an integral curve of H; however, an integral curve $z_{\cdot}$ is a response of S only if $v(t) = 1$ in Eq.(a) of this Section, or (re-parametrizing time) if v is always positive, at least. See the example in 2.12c, below. We can conclude that $A^y \subset L^y$, and the inclusion may be strict.

$\Xi(\alpha_o,\alpha_1)$ or Ξ for short will denote the smallest vector field system containing H and closed under the bracket operation; Ξ is a Lie sub-algebra of $\underline{V}$ (see 2.12 c, for an example). Ξ contains, and is spanned by, the commutators $[\alpha_o,\alpha_1]$, $[\alpha_o,[\alpha_o,\alpha_1]]$, $[\alpha_1,[\alpha_o,\alpha_1]]$, ... The <u>evaluation of Ξ at x</u> is the vector space $\Xi_x = \{\beta x: \beta \in \Xi\}$.

If the dimension of Ξ_x is r for <u>all</u> x, we say Rank(Ξ) = r; if Rank(Ξ) = n, then $\Xi = \underline{V}$.

(2.5) <u>Proposition</u>.

If Rank(Ξ) = r, then for each $y \in R^n$ the leaf L^y is an r-dimensional connected C^ω manifold; the space L^y_x of tangent vectors to L^y at a point x is just Ξ_x.

This paraphrase of the Frobenius "complete integrability" theorem is from Chapter 8 of [Hermann 1968] (he does the C^∞ case; see p.92 of [Chevalley] for a C^ω proof). The functions ϕ that are constant on L^y satisfy $\beta\phi = 0$ for all $\beta \in \Xi$; these p.d.e.'s have n-r independent local solutions, which define little r-patches (C^ω images of r-dimensional euclidean neighborhoods) which can be pieced together to form a smooth manifold. The important point for us is that L^y cannot contain any n-dimensional neighborhood if $r < n$.

(2.6) <u>Multi-vector fields</u>.

An <u>r-vector field</u> $\beta_1\Lambda\beta_2\Lambda\cdots\Lambda\beta_r$ takes r-tuples of C^ω functions into $C^\omega(R^n)$, and is given by $\beta_1\Lambda\cdots\Lambda\beta_r(\phi_1,\phi_2,\cdots,\phi_r) = \mathrm{Det}(\beta_i\phi_j)$. Thus for $r = 2$, $\beta_1\Lambda\beta_2(\phi_1,\phi_2) = (\beta_1\phi_1)(\beta_2\phi_2) - (\beta_1\phi_2)(\beta_2\phi_1)$. We will write $\beta_1\Lambda\cdots\Lambda\beta_r(x) = 0$ if at x the function $\beta_1\Lambda\cdots\Lambda\beta_r(\phi_1,\cdots,\phi_r) = 0$ for any r-tuple of C^ω functions. This is equivalent to the linear dependence of the vectors $\beta_1 x$, $\beta_2 x, \cdots, \beta_r x$.

(2.7) Lemmas for Theorem (2.8).

a) From (2.6), $\mathrm{Dim}(\Xi_x) = r$ if and only if there exist $\beta_1,\ldots,\beta_r \in \Xi$ such that $\beta_1\Lambda\cdots\Lambda\beta_r(x) \neq 0$ (r linearly independent vectors $b_i(x)$ in Ξ_x).

b) From the multilinearity of the Λ-product, along a response x_t of S

$$\frac{SD}{dt}(\beta_1\Lambda\beta_2) = \left(\frac{SD}{dt}\beta_1\right)\Lambda\beta_2 + \beta_1\Lambda\frac{SD}{dt}\beta_2;$$

$$\frac{SD}{dt}(\beta_1\Lambda\beta_2)\big|_{t=0} = [\alpha_o,\beta_1]\Lambda\beta_2 + \beta_1\Lambda[\alpha_o,\beta_2] + u(0)[\alpha_1,\beta_1]\Lambda\beta_2 + u(0)\beta_1\Lambda[\alpha_1,\beta_2]$$

from (2.3e). If β_1 and β_2 belong to Ξ, each term on the RHS is of the form $\beta_i\Lambda\beta_j$, β_i and $\beta_j \in \Xi$. The second and higher system derivatives inherit this property (with coefficients $\dot{u}(0)$, etc.) and similar statements are true for r-vectors of any order with factors $\beta_1,\ldots,\beta_r \in \Xi$.

c) If $\beta_1,\ldots,\beta_r \in \Xi$, if x_o is a state such that $\mathrm{Dim}(\Xi_{x_o}) < r$, and if $u(\cdot)$ is real-analytic, then from lemmas (a) and (b), the system derivatives of $\beta_1\Lambda\cdots\Lambda\beta_r$ of all orders vanish at t = 0: $\frac{SD^k}{dt^k}(\beta_1\Lambda\cdots\beta_r)\big|_{t=0} = 0$.

d) $F(r) \triangleq \{x: \mathrm{Dim}(\Xi_x) < r\}$ is an invariant set for the system S. That is, if $x_o \in F(r)$, then for any $u \in C^{\omega}[0,T]$, $x_t \in F(r)$, $0 \leq t \leq T$.

Proof: x_t is real-analytic in t. If $\phi_i \in C^{\omega}(R^n)$, $i = 1,\ldots,r$, then $\beta_1\Lambda\cdots\Lambda\beta_r(\phi_1,\cdots,\phi_r)(x_t) \triangleq \phi(x_t)$ is real-analytic in t, and its Taylor series vanishes at $t = 0$ by lemma (c), so $\phi(x_t) = 0$ for t in a neighborhood of 0, and the result holds by analytic continuation. Since any point in A^{x_o} is accessible by a C^{ω} control, $A^{x_o} \subset F(r)$.

Discussion. The above lemma (d) and Theorem(2.5) are all we need to prove the next Theorem (2.8) and, by an easy extension, Theorem(2.10). These results are similar in some respects to those of [Hermann 1963], extending the Frobenius theorem to "foliations with singularities" in the C^{ω} case, but the statements and proofs we give are new (Hermann was not concerned with necessary consequences of accessibility).

(2.8) Theorem.

If S has the accessibility property, then $\mathrm{Rank}(\Xi(\alpha_o,\alpha_1)) = n$.

Proof: $F(n) = \{x: \mathrm{Dim}(\Xi_x) < n\}$. If $F(n)$ is empty, $\mathrm{Rank}(\Xi) = n$ and we are through; so suppose there exists a point $y \in F(n)$. By lemma (2.7d), $A^y \subset F(n)$. By the accessibility property, A^y contains an open set G, so $\beta_1 \wedge \cdots \wedge \beta_n(x) = 0$ on G if $\beta_1,\ldots,\beta_n \in \Xi$. C^ω functions that vanish on an open set vanish on R^n, so $F(n) = R^n$.

From the definition of $F(n)$ we see that either $\mathrm{Rank}(\Xi) = n-1$ or $F(n-1)$ is non-empty. First suppose $\mathrm{Rank}(\Xi) = n-1$. Then by Prop.(2.5), L^y is, for each y, an n-1-dimensional manifold; but $A^y \subset L^y$, and the accessibility property gives us a contradiction. Thus we must suppose that $F(n-1)$ is non-empty.

We can now repeat the above argument, substituting for n the numbers n-1, n-2,..., 1, 0. F(0) is empty, but our argument would contradict that, completing the proof (only the possibility $\mathrm{Rank}(\Xi) = n$ was uncontradicted).

(2.9) Controllability definitions.

S is called $\hat{T}$-controllable if there exists a finite non-negative function $\hat{T}(y,x)$ defined on $R^n \times R^n$ that satisfies the conditions

1) $\hat{T}(x,x) = 0$; [$\hat{T}$ is read "T- hat"]

2) for any y, $\{(x,t): t > \hat{T}(y,x)\}$ is open in $R^n \times [0,\infty)$;

3) for every $t_1 > T(y,x)$ there exists $u \in \underline{L}^2$ such that $X(y,t_1;u)=x$.

(The special case $\hat{T} \equiv 0$ is called "differentially controllable" in the system theory literature.) From Condition (3), $\hat{T}$-controllability implies that S has "state-to-state controllability in finite time"; using (2) we can show (see 2.10) that S has an extended accessibility property (in

(x,t)-space), that is all we need for the remaining results of this chapter (in Section 3.6 the notion of controllability-- every point of R^n can be reached from every other point-- will be important). See Section 2.12 for examples (e - h) of $\hat{T}$-controllable systems. It should be noted that Condition(2) is implied by

2') for all y, the function $T(y,\cdot)$ is bounded on compact sets.

We now give an algebraic definition of controllability, essentially the same as those given by [Hermann 1963], [Haynes], and [Hermes] but in a form suitable for application to Hörmander's Theorem (Sec. 2.11). If we lift our vector fields from R^n to the (x,t)-space R^{n+1}, we can adjoin the vector field $\partial/\partial t$ to $\underline{V}$ and use the ring $C^{\omega}(R^{n+1})$. The Lie algebra generated by $\alpha_o - \partial/\partial t$ and α_1 will be denoted by $\Xi'(\alpha_o - \partial/\partial t, \alpha_1)$ or Ξ' for short. Since S is time-invariant, we can use $C^{\omega}(R^n)$ for our ring of coefficients without affecting considerations of rank or dimension. Thus we say

S is H-controllable if Ξ' has rank $n+1$.

(2.10) Proposition.

If S is $\hat{T}$-controllable it is H-controllable.

Proof: we copy the time-axis with an extra state-variable ζ, thus setting up a new system S' on R^{n+1}, given by the differential equation

$$\frac{d}{dt}\begin{pmatrix} x \\ \zeta \end{pmatrix} = \begin{pmatrix} a_o(x) \\ 1 \end{pmatrix} + \begin{pmatrix} a_1(x) \\ 0 \end{pmatrix} u(t)$$

and satisfying the hypotheses of Section 2.1. The open set $\{(x,\zeta): \zeta > \hat{T}(y,x)\}$ is accessible from y, so S' has the accessibility property. Corresponding to Ξ we have $\Xi'(\alpha_o + \partial/\partial\zeta, \alpha_1)$, with rank $n+1$ by Theorem 2.8. Since α_o and α_1 are independent of ζ, we can replace $\partial/\partial\zeta$ with $-\partial/\partial t$ to obtain the desired conclusion.

Discussion. It is easy to see that if $\mathrm{Rank}(\Xi') = n + 1$ then $\mathrm{Rank}(\Xi) = n$. It is *not* true (without controllability hypotheses) that $\mathrm{Rank}(\Xi) = n$ implies $\mathrm{Rank}(\Xi') = n + 1$; consider the example on R^2 $d\xi/dt = u(t)$, $d\eta/dt = 1$; Ξ is spanned by $\partial/\partial\eta$ and $\partial/\partial\xi$; Ξ' is spanned by $\partial/\partial\eta - \partial/\partial t$, $\partial/\partial\xi$, so each has rank 2. Another example:

$$d\xi/dt = -\eta + \xi u(t), \quad d\eta/dt = \xi + \eta u(t), \quad \text{on } R^2 - \{(0,0)\};$$

$\mathrm{Rank}(\Xi') = 2 = \mathrm{Rank}(\Xi)$ on the punctured plane; the system has an invariant manifold in (ξ,η, t)-space: $\frac{SD}{dt}[t - \arctan(\eta/\xi)] = 0$ for all u. Such examples suggest the

Conjecture: If for every x and y in R^n there exists a time t and a C^ω control $u(\cdot)$ such that $X(y,t; u) = x$, then S is H-controllable.

(2.11) *Theorem* (Hörmander):

If $\mathrm{Rank}(\Xi') = n + 1$, then the operators $Q = -\partial/\partial t + \alpha_o + (1/2)(\alpha_1)^2$ and $D = \alpha_o + (1/2)(\alpha_1)^2$ are *hypoelliptic* on R^{n+1}, R^n, respectively.

(That is, if f and g are Schwartz distributions on R^{n+1} and $Qg = f$, then g is C^∞ on any open set in R^{n+1} where f is C^∞, after a modification of g on a set of measure zero to remove any "removable discontinuities"; the same is true for D and R^n.)

This has been proved [Hormander] for C^∞ operators of the form $\beta_0 + \sum_1^\nu \beta_i^2$, where $\beta_0,\ldots,\beta_\nu$ generate a C^∞ Lie algebra. If the C^ω rank of Ξ' is n+1, so is the C^∞ rank, since a non-vanishing C^ω (n+1)-vector is a C^∞ (n+1)-vector and the rank of Ξ' can be no greater than n+1. Since Ξ must have rank n, that will hold for Ξ as well as Ξ', substituting n for n+1.

(2.12) Remarks and Examples.

(a) The preceding results will be used in Chapter 3 in the following form: given that S is $\mathbf{H}$-controllable, if the initial value problem

$$Qg = -\partial g/\partial t + Dg = f(x)\delta(t), \quad t \geq 0$$

has a weak solution g [with respect to $C_0^\infty(R^n \times (0,\infty))$] for some initial function $f \in B(R^n)$, it has a C^∞ classical solution $\tilde{g}$ on the $t > 0$ half-space, obtained by modifying g on a set of measure zero.

(b) A generalization: consider systems of the form

$$dx/dt = a_o(x) + \sum_{\nu=1}^m u_\nu(t)a_\nu(x).$$

If this system is C^ω and H-controllable, Theorem 2.8 and Hörmander's Theorem hold, so the operator $-\partial/\partial t + \alpha_o + (1/2)\sum_{\nu=1}^m \alpha_\nu^2$ is hypoelliptic. The α_ν do not have to be linearly independent.

(c) Chow's Theorem [Hermann 1]:

If H is a C^∞ vector field system on R^n, if L^y is the leaf of H through y, and $\Xi(H)$ is the Lie algebra generated by H, then if $\Xi(H)$ has rank n, L^y is all of R^n.

Theorem 2.8 is a C^ω converse of Chow's Theorem.

It is not generally true that $A^y = R^n$ for an H-controllable system. The following example on R^2 is from [Haynes]:

$$dx/dt = u, \quad dy/dt = x^2 \ ; \ \Xi' = \Xi'(-\partial/\partial t + x^2\partial/\partial y, \ \partial/\partial x);$$

$\alpha_o = x^2\partial/\partial y$, $\alpha_1 = \partial/\partial x$, $[\alpha_o,\alpha_1] = 2x\partial/\partial y$, $[\alpha_o,[\alpha_o,\alpha_1]] = 2\partial/\partial y$, so Ξ' has basis $\partial/\partial t$, $\partial/\partial x$, $\partial/\partial y$ and its rank is 3, but from any point (x,y) in the plane only the region above y is accessible.

(d) For the special case of a linear system with constant coefficients

$$S_o\colon \ dx/dt = Ax + bv(t),$$

H-controllability is equivalent to complete controllability in the sense of "state-to-state controllability on any time interval," due to Kalman. It is easily seen that the Lie algebra $\Xi'(x'A'\nabla'-\partial/\partial t,\ b'\nabla')$ has rank $n+1$ only if the vectors $b, Ab, \ldots, A^{n-1}b$ are linearly independent, in which case (A, b) is called a controllable pair. For any $T > 0$, let $v(t) = e^{A(T-t)}c$, $0 \leq t \leq T$. The matrix $M(T) = \int_0^T e^{As}bb'e^{A's}ds$ is then non-singular. For a given x_o and the given family of controls $v(t;c)$

$x_T = e^{At}x_o + M(T)c$, which can be solved for c given any desired endpoint x_T. That is complete controllability. On the other hand, given complete controllability we have $\hat{T}$-controllability with the function $\hat{T}(x,y)$ identically zero. H-controllability follows by our Theorem 2.8.

In this work we are postulating that S is C^{ω} and $\hat{T}$-controllable (and not necessarily linear). The two examples that follow are to show that there are non-trivial systems of this sort.

(e) Let (A, b) be a controllable pair; consider the nonlinear system

S_1: $dx/dt = Ax + b\phi(x) + b\psi(x)u(t)$

where ϕ, ψ are C^{ω} real functions and $\psi(x) > 0$. Then S_1 is $\hat{T}$-controllable with $\hat{T}(x,y) \equiv 0$.

Proof: Given states y and z, let $v(t) = e^{A(T-t)}c$, where $c = M^{-1}(T)(z - e^{At}y)$. Let $u(t) = [v(t) - \phi(\tilde{x}_t)]/\psi(x_t)$, where $\tilde{x}_t$ is the response of system S_o to $v(\cdot)$, starting at y. From (d), we see $\tilde{x}_T = z$. We then have, for the S_1- response x_t starting at y with control $u(\cdot)$, the differential equation

$$\frac{d}{dt}[x_t - \tilde{x}_t] = A[x_t - \tilde{x}_t] + b[\ \phi(x_t) - \phi(\tilde{x}_t)],\quad 0 \leq t \leq T;$$

but this has the unique solution $x_t - \tilde{x}_t = 0$ on $[0,T]$. Therefore $x_T= z$.

(f) Let (A, b) be a controllable pair, and consider the system

S_2: $dx/dt = Ax + bu(t) + g(x)$,

where g is a bounded Lipschitzian vector function on R^n. Then S_2 is $\hat{T}$-controllable with $\hat{T}(x,y) \equiv 0$. We can require g to be C^ω.

The proof is easily obtained from that of [Hermes], Th. 1.2 by writing $x_T = e^{AT}x_o + M(T)c + h(c)$, where h turns out to be bounded and continuous in c; then one solves for c and applies the Brouwer fixed-point theorem, as shown by [Hermes], to obtain a control to any desired x_T, for any positive T.

(g) Theorem 2.8 does not remain true if C^ω is replaced by C^∞ in the hypothesis. Consider this example in R^2:

S_3: $dx/dt = u(t)$, $dy/dt = f(x)$,

where $f(x)$ vanishes for $|x| > 1$, $f(x) = x \exp[1/(x^2 - 1)]$ for $|x| \leq 1$. To obtain a trajectory connecting two given points, one makes use of the strips where y increases or decreases, as needed, and then moves in the x direction. Outside these strips the rank of Ξ' is **two**, yet the system is $\hat{T}$-controllable (with positive $\hat{T}(\cdot,\cdot)$).

(h) Finally we give a $\hat{T}$-controllable C^ω example with $\hat{T} > 0$. Again n=2.

S_4: $dx/dt = -x + u(t)$, $dy/dt = 2 \tanh(x) - \tanh(y)$.

It is easily seen that $\hat{T}$ is greater than 1/3 of the difference of the initial and final values of the y-coordinate.

CHAPTER 3. THE STOCHASTIC SYSTEM Σ

If the set of controls or inputs to the system S is given a probability measure P, S becomes a stochastic process. In some applications it is desirable to have an extension of S that makes sense when the input process is a white noise; the chief benefit is that S becomes a Markov process Σ on R^n. Such an extension is given by the Itô stochastic differential equation (see the following sections for precise definitions)

$$d\underline{x}_t = m(\underline{x}_t)dt + a_1(\underline{x}_t)dw_t, \quad \underline{x}_o = y$$

where $m(x) = a_o(x) + (1/2)\alpha_1 a_1(x)$, and w_t is a scalar Brownian motion.

This equation was first used by Stratonovich; if x is scalar (n=1) [Wong and Zakai] derives it from Eq.(2.1a) by requiring that $\int_0^t u(s)ds$ approach w_t in an appropriate topology. (The proof does not work for larger values of n.) As a mathematical model, it has engineering and physical justifications; see [Stratonovich] and [Mortensen 1968] for discussions of this point. We believe that the results of this chapter will also justify the use of this equation, and Chapter 4 gives a sample of its application to the problem of the existence of invariant measures.

In Theorem 3.5 we show that the transition probabilities of Σ, our Markov process, satisfy the backward Kolmogorov p.d.e., and have densities satisfying the forward Kolmogorov p.d.e.-- without *a priori* assumptions about the existence or differentiability of the density and without the use of classical existence-uniqueness theorems, which would require that the differential generator be an elliptic operator (here it is not). In Theorem 3.6 we show that T-controllability implies that for each initial y and each open set G in R^n there exists a sequence of times $\Theta_i(y,G)$ approaching infinity such that $P(y,\Theta_i, G) > 0$: and we "embed" S in Σ.

(3.1) Brownian motion and stochastic integrals.

By the Brownian motion, or Wiener process, we mean the trio $W \triangleq (C_b, F_\infty, P)$, where C_b is the space of continuous real (scalar) functions w_t, $t \geq 0$, with $w_o = 0$, equipped with the topology of uniform convergence on compact time-intervals; F_∞ is the smallest σ-algebra of subsets of C_b that includes all events of the form $\{\lambda \leq w_t < \mu\}$, $\lambda < \mu$, $t \geq 0$ (the open sets in the uniform topology are F_∞-measurable); and P is the completely additive probability measure induced on C_b by

$$P\{\lambda \leq w_t < \mu |\ w_r: r \leq s\} \triangleq \int_\lambda^\mu [2\pi(t-s)]^{-1/2} \exp[-(w_s - y)^2/2(t-s)]dy,$$

which gives positive measure to the open sets of C_b. If the Brownian motion is restricted to some interval $[0,T]$, the space $C_b[0,t]$ with the uniform ("sup") norm is a Banach space.

F_t will denote the smallest σ-subalgebra of F_∞ that includes all events of the form $\{w: \lambda \leq w_s < \mu\}$, $\lambda < \mu$, $0 \leq s \leq t$. Note that $w_{t+s} - w_s$, $t \geq 0$, is a Brownian motion independent of F_s [McKean].

A vector random function $\underline{h}_t = h_t(w)$ (we indicate dependence on w by the underline) is a function defined on $C_b\ [0,\infty)$ with range R^n, F_t-measurable (for each t) and t-measurable (for almost every w).

E denotes the expectation operator (integral with respect to P).

In order to discuss Itô integrals of random functions it is convenient to introduce the family of Hilbert spaces

$$\underline{H}_t = \{\underline{h}_\cdot : E\int_0^t ||\underline{h}_s||^2 ds < \infty\}, \text{ where } ||\cdot|| \text{ is the norm in } R^n,$$

adopting the convention that $\underline{h} = 0$ in q.m. (quadratic mean) if $E\int_0^t ||\underline{h}_s||^2 ds = 0$ and that elements of $\underline{H}_t$ are to be regarded as q.m. equivalence classes. Now the Itô stochastic integral can be defined [McKean] as a linear map from $\underline{H}_t$ to $\underline{H}_t$ (indefinite integral) with the properties

(i) $E\int_0^t \underline{h}_s dw_s = 0$

(ii) $\int_0^t h dw_s = hw_t$ if h is a constant

(iii) $\int_0^t \underline{h}_s \, dw_s$ is F_t- measurable and almost surely continuous in t.

(3.2) Definition of Σ

Let a_o and a_1 be C^∞ mappings (not necessarily C^ω) from R^n to R^n that satisfy the global Lipschitz conditions (2.1b). Let

$$m(x) = a_o(x) + (1/2)\alpha_1 a_1(x) \quad (\alpha_1 = a_1(x)\cdot(\partial/\partial x) \text{ as usual})$$

and note that since

$$\sup_{||y||=1} ||(y\cdot\partial/\partial x)a_1(x)|| \leq K,$$

$m(\cdot)$ is Lipschitzian with constant K^2 instead of K. If we are given a random function $\underline{x}_t$, $t \geq 0$, that satisfies the equation

a) $\underline{x}_t = y + \int_0^t m(\underline{x}_s)ds + \int_0^t a_1(\underline{x}_s)dw_s, \quad t \geq 0,$

almost surely, and $\underline{x}_\cdot \in \underline{H}_t$, then we say that $\underline{x}_\cdot$ is an Itô solution of Eq.(a), which is also commonly written

a') $d\underline{x}_t = m(\underline{x}_t)dt + a_1(\underline{x}_t)dw_t.$

For proof of the following facts the reader is referred to Chapter 11 of [Dynkin] or Chapter 3 of [Skorokhod].

Eq.(a) has an Itô solution $\underline{x}_t = Y(y,t;w)$ which is almost surely unique and continuous on any interval [0,T]; from continuity of the sample paths we see that $\underline{x}_t$ is bounded on [0,T] almost surely. The collection of trajectories $Y(y,t;w)$, $y \in R^n$, $t \geq 0$, $w \in W$, the σ-algebras F_t, and the conditional probabilities

$$P_y: \; P_y\{\underline{x}_t \in G\} = P\{w: \underline{x}_t \in G \mid \underline{x}_o = y\}, \quad G \text{ a Borel set of } R^n$$

constitute a Markov process Σ which is our stochastic system.

(3.3) Properties of Σ.

Let $\underline{B}$ denote the class of R^n Borel sets. The transition function

$P(x,t,G) = P_x\{\underline{x}_t \in G\}$ is for each (x,t) a probability measure on the sets $G \in \underline{B}$, and satisfies the Chapman-Kolmogorov equation

a) $\quad P(x, t+s, G) = \int P(y,s,G)P(x,t,dy)$, $\quad s$ and $t \geq 0$,

where dy signifies the volume element in R^n, the domain of integration.

Let $F(R^n)$ denote the class of Borel functions on R^n, $B(R^n)$ the Banach space of bounded Borel functions on R^n with the uniform norm $||f||_\infty = \sup_x |f(x)|$. If $f \in F(R^n)$ and the right hand side (RHS) exists,

$$T_t f(x) \triangleq E_x f(\underline{x}_t) \triangleq \int f(y)P(x,t,dy)$$

is called the conditional expectation of $f(\underline{x}_t)$ given $\underline{x}_o = x$. The one parameter family T_t, $t \geq 0$, of linear bounded operators mapping $B(R^n)$ into $B(R^n)$ is a strongly continuous contraction semigroup [Dynkin].

Theorems 11.4 and 11.5 of [Dynkin] establish that T_t maps the Banach space C of bounded continuous functions on R^n into itself (the Feller property) and that if f is C^2 with compact support ($f \in C_o^2(R^n)$) then

b) $$\lim_{t\to 0} \left|\left| \frac{T_t f(x) - f(x)}{t} - Df(x) \right|\right|_\infty = 0, \quad \text{where}$$

$$Df = \alpha_o f + (1/2)\alpha_1^2 f \quad \in C_o(R^n) \ ;$$

D is called the differential generator of Σ, which is thus a diffusion process in the terminology of [Dynkin].

Ito's Lemma: If $V(t,x)$ is C^1 in t and C^2 in x and if $\underline{x}_\cdot$ is a trajectory of Σ, the random (scalar) function $\underline{\Phi}(t) = V(t, \underline{x}_t)$ satisfies

c) $$d\underline{\Phi}(t) = \left[\frac{\partial\Phi}{\partial t}(t, \underline{x}_t) + DV(t, \underline{x}_t)\right]dt + \alpha_1 V(t, \underline{x}_t)dw_t .$$

The proof can be found in [McKean] or [Dynkin].

By establishing the q.m. differentiability of $Y(y,t;w)$ with respect to t and (twice) w.r.t. y, [Skorokhod] establishes the following

Proposition: If $m(\cdot)$ and $a_1(\cdot)$ have continuous bounded second derivatives and $f \in C_o^2(R^n)$, then the function $g = T_t f$ satisfies the backward

Kolmogorov p.d.e. $\partial g/\partial t = Dg$, $g(x,0) = f(x)$.

The proof occupies the latter half of Chapter 3 of [Skorokhod]. It should be noted that this proposition cannot be obtained by semigroup methods because $T_t f$ does not necessarily have compact support.

(3.4) Proposition.

Let Σ satisfy the conditions

(A) $\Xi'(-\partial/\partial t + \alpha_o, \alpha_1)$ has rank $n + 1$

(B) $m(\cdot)$ and $a_1(\cdot)$ have continuous bounded second derivatives;

if $f \in B(R^n)$, then the function $g = T_t f$ is equal a.e. to a function $\tilde{g} \in C^\infty(R^n \times (0,\infty))$ that satisfies

a) $Q\tilde{g} = -\partial\tilde{g}/\partial t + D\tilde{g} = 0.$

Discussion. Condition (B) will be removed in Theorem 3.5. Note that if the system S is T-controllable and C^ω, (A) is satisfied.

Proof: We only need to show that g is a weak solution of Eq.(a), that is, (denoting the adjoint of Q by Q^*)

b) $\int_{R^n\times(0,\infty)} g(x,t)Q^*\phi(x,t)\,dxdt = 0$ for all $\phi \in C_o^\infty(R^n\times(0,\infty))$;

$$Q^* = D^* + \partial/\partial t = \frac{\partial}{\partial t} - \alpha_o + \frac{\partial}{\partial x}\cdot a_1(x)\alpha_1 + \frac{1}{2}\alpha_1^2 + \chi ,$$

$$\chi(x) = \frac{1}{2}\left(\frac{\partial}{\partial x}\cdot a_1(x)\right)^2 + \frac{1}{2}\alpha_1\left(\frac{\partial}{\partial x}\cdot a_1(x)\right) - \frac{\partial}{\partial x}\cdot a_o(x),$$

and our conclusion will follow from the hypoellipticity of Q (Th.2.11).

Definition: if the numerical sequence $\{f_\nu(x)\}$ converges for all x as $\nu\to\infty$ and if $||f_\nu||_\infty < M$ (some positive number independent of ν), we say $\{f_\nu\}$ is w-convergent and $f_\nu \xrightarrow{W} f$, where $f(x) = \lim_{\nu\to\infty} f_\nu(x)$.

Lemma: if a family Ψ of real functions on R^n is closed under the operations of addition, multiplication by real numbers, and w-convergence, and if $C_o^2(R^n) \subset \Psi$, then $B(R^n) \subset \Psi$. (Proved in [Dynkin] as Lemma 5.12)

To apply this, the family Ψ is defined by

$$\Psi = \{f \in B(R^n) \colon g = T_t f \text{ satisfies Eq.(b)}\}.$$

By the Skorokhod Proposition of the preceding Section, $C_o^2(R^n) \subset \Psi$. T_t is linear, so Ψ is closed under addition and multiplication by reals. Now consider any w-convergent sequence $\{f_\nu \in \Psi\}$, $|f_\nu(x)| < M$, say, for all ν. Then $\{f_\nu\}$ has a w-limit $\hat{f} \in B(R^n)$. By the dominated-convergence theorem of Lebesgue,

$$\lim_{\nu\to\infty} E_x f_\nu(\underline{x}_t) = E_x \hat{f}(\underline{x}_t) \quad \text{for all } x \text{ and all } t > 0,$$

and the sequence $g_\nu = T_t f_\nu$ is uniformly bounded by M; so the sequence g_ν is w-convergent on $R^n \times (0,\infty)$, with some w-limit $\hat{g}$.

If $\phi \in C_o^\infty(R^n \times (0,\infty))$, so is $Q^*\phi$. Since $g_\nu - \hat{g} \xrightarrow{W} 0$,

$$\lim_{\nu\to\infty} \int_{R^n \times (0,\infty)} (g_\nu - \hat{g}) Q^*\phi \, dxdt = 0.$$

By hypothesis each g_ν satisfies Eq.(b), so $\hat{g}$ satisfies it for all ϕ.

Therefore $\hat{f} \in \Psi$; applying the lemma, $\Psi = B(R^n)$.

(3.5) Theorem.

If Σ satisfies (A) and if $f \in B(R^n)$, then $g = T_t f$ is C^∞ and $\partial g/\partial t = Dg$. If also $f \in C$ then g is the unique bounded solution of $\partial g/\partial t = Dg$ satisfying $\lim_{t\downarrow 0} g(x,t) = f(x)$. Furthermore, Σ has a transition density $p(x,t,\cdot)$ such that $\partial p/\partial t = D^*p$, for each x.

Proof: Step 1.

Consider the sequence of spheres $U_\nu = \{x \colon ||x|| < \nu\}$, $\nu = 1,2,\ldots.$ Choose a sequence of C^∞ functions $\theta_\nu > 0$ on R^n such that $\theta_\nu(x) = 1$ on U_ν and, as $||x|| \to \infty$, $\theta_\nu(x) \to 0$ so fast that the coefficients

$$a_o^\nu = \theta_\nu a_o, \quad a_1^\nu = \theta_\nu a_1, \quad m^\nu = a_o^\nu + (1/2)(a_1^\nu \cdot \tfrac{\partial}{\partial x}) a_1$$

have bounded second derivatives as required by condition (B) of Sec.(3.4).

Let the sequence of Markov processes Σ^{ν} be given on R^n by

$$d\underline{x}^{\nu} = m^{\nu}(\underline{x}^{\nu})dt + a_1^{\nu}(\underline{x}^{\nu})dw_t,$$

with D^{ν}, P_x^{ν}, E_x^{ν}, T_t^{ν}, defined appropriately for $\nu = 1,2,\ldots$. Since $\theta_{\nu} > 0$, $\text{Rank}(\Xi'(\alpha_o^{\nu} - \partial/\partial t,\ \alpha_1^{\nu})\ ) = n+1$, so Prop. 3.4 can be applied to Σ^{ν}. If $f \in B(R^n)$, $g^{\nu} = T_t^{\nu} f$ satisfies Eq.(3.4b), $\nu = 1,2,\ldots$.

By the <u>part of a Markov process</u> on U_{ν} we mean (see [Dynkin], [Khasminskii]) the process obtained by cutting off the trajectories at the <u>stopping time</u> τ_{ν} of first exit from U_{ν} [τ_{ν} is a random functional such that the event $\{w:\tau_{\nu} < t\} \in F_t$ for each $t > 0$]; for $t \geq \tau_{\nu}$ we assign $\underline{x}_t = \Delta$, Δ a "cemetery" state outside R^n on which functions of $\underline{x}_t$ are all defined to be zero.

The parts of Σ and Σ_{ν} on U_{ν} have the same transition function $P^{\nu}(x,t,G)$, $G \in B$ restricted to U_{ν}, $x \in U_{\nu}$. $\lim_{\nu\to\infty} P_x\{\underline{x}_s \in U_{\nu},\ 0 \leq s \leq t\} = 1$ so $P(x,t,G) = \lim_{\nu\to\infty} P_x\{\underline{x}_t \in G \mid \underline{x}_s \in U_{\nu},\ 0 \leq s \leq t\} = \lim_{\nu\to\infty} P^{\nu}(x,t,G)$. The sequence of numbers on the right-hand side is monotone non-decreasing for fixed (x,t,G), so the limit exists. $||g_{\nu}||_{\infty} \leq ||f||_{\infty}$, so $g_{\nu} \xrightarrow{w} g$ on $R^n \times (0,\infty)$. By the same argument we used in Prop. 3.4, we conclude that g satisfies Eq.(3.4b) and is equal a.e. to a C function $\tilde{g}$ that satisfies Eq. (3.4a). In Step 3 we will show that $g = \tilde{g}$.

If G is any compact set in R^n and for any $\varepsilon > 0$ $N_{\varepsilon}(x)$ is an ε-neighborhood of x, then Σ has the property that [Dynkin] calls

<u>N'(G)</u>: $\quad \lim_{t\downarrow 0} (1/t) \sup_G [\ 1 - P(x,t,N_{\varepsilon}(x))\ ] = 0.$

To prove this construct a function $h \in C_o^2(R^n)$ which equals 1 on $N_{\varepsilon/2}(x)$, vanishes outside $N_{\varepsilon}(x)$, and nowhere exceeds 1. h is in the domain of D, and Theorem 3.9' of [Dynkin] completes the proof. From this property,

$$\lim_{t\downarrow 0} g(x,t) = \lim_{t\downarrow 0} \int f(y)P(x,t,dy) = f(x) \text{ uniformly on compact sets.}$$

Step 2. Now let us use test functions $\psi(x,s) \in C_0^\infty(R^n \times (0,\infty))$ as initial functions (considering s as a parameter for the time being). Since $T_t\psi(x,s)$ is bounded, corresponding to $P(x,t,dy)$ there is a formal density $p(x,t,y)dy$ in the sense of a Schwartz distribution on our test functions ψ. Let us use the notation D_y to mean "D acting in the y coordinates, x fixed"; since $T_t\psi(x,s)$ is differentiable in t,

$$\frac{\partial}{\partial t}T_{t+s}\psi(x,s)\Big|_{t=0} = \frac{\partial}{\partial t}\int T_t\psi(y,s)P(x,s,dy)\Big|_{t=0} = \int D_y\psi(y,s)P(x,s,dy);$$

but, integrating the left-hand side,

$$\int\frac{\partial}{\partial t}T_{t+s}\psi(x,s)\Big|_{t=0} ds = -\int_{R^n\times(0,\infty)}\left[\frac{\partial}{\partial s}\psi(y,s)\right]P(x,s,dy)ds;$$

a) $\int[D_y\psi(y,s) + \frac{\partial}{\partial s}\psi(y,s)]P(x,s,dy)ds = 0$ for all ψ.

We see from Eq.(a) that $p(x,t,y)$ is a Schwartz-distribution solution of the partial differential equation $\partial p/\partial t = D_y^*p$, for fixed x. (Since D^* is of the form $\beta_0 + (1/2)\beta_1^2 + \chi$, where β_0 and $\beta_1 \in \Xi$ and χ is a C^∞ function, $D^* - \partial/\partial t$ is hypoelliptic.) Since p is a derivative of a probability distribution on R^n, p can have no removable discontinuities, and therefore p is actually C^∞ without modification (compare the use of Weyl's Lemma, and a similar argument, in [McKean]). It is easy to see that for given x and t $p(x,t,\cdot)$ is a probability density, and that as $t\downarrow 0$ it converges in the Schwartz sense to the Dirac δ-function $\delta(x-y)$.

Step 3. If $f \in B(R^n)$, from Step 1 $g(x,t) = \tilde{g}(x,t)$ a.e., where $\tilde{g}$ is C^∞. From Step 2, for $s > 0$

$$g(x, t+s) = \int g(y,t)p(x,s,y)dy = \int \tilde{g}(y,t)p(x,s,y)dy \in C.$$

Treating g itself as an initial function, as $s\downarrow 0$ $g(x, t+s) \to g(x,t)$ uniformly on compact sets, so $g \in C$. Therefore $g = \tilde{g}$, and $\partial g/\partial t = Dg$ from Prop. 3.4. It is easy to see that $\partial p/\partial t = D_x p$.

Step 4. Let the initial function $f \in C$; then to show that $g = T_t f$ is the unique bounded solution of $\partial g/\partial t = Dg$ satisfying $g(x,0) = f(x)$, we need only show that the problem

c) $\partial u/\partial t = Du$, $u(x,0) = 0$, u bounded and C^∞

has the unique solution $u = 0$.

Let u be a non-zero solution of problem (c) with the bound M. For any $\lambda > 0$ there exists the Laplace transform $v(x,\lambda) = \int_0^\infty e^{-\lambda t} u(x,t)\,dt$, and the integral is uniformly convergent; so v is continuous in x and bounded by M/λ. By the uniqueness theorem for the Laplace transform, $v(x,\lambda)$ is non-zero for some λ, which we fix. We may then suppose that for x in some neighborhood, $v(x) \triangleq v(x,\lambda) > 0$. Integrating by parts and using (c), we have $\lambda v(x) = \int_0^\infty e^{-\lambda t} Du(x,t)\,dt$; from this we can show that v is a weak (Schwartz) solution of $Dv = \lambda v$. However, $D - \lambda$ is hypoelliptic; knowing $v \in C$ we can conclude that v is C^∞ without any modification.

Step 5. Definition: we say that a function h is in the domain of the weak infinitesimal generator $\tilde{A}$ if as $t \downarrow 0$ $(1/t)(T_t h - h) \xrightarrow{W} \tilde{A}h \in \tilde{B}_0$ where $\tilde{B}_0 = \{f \in B\colon \operatorname{w-lim}_{t \downarrow 0} T_t f = f\}$.

This definition ([Dynkin] v. 1, p. 55) obviously requires that h and $\tilde{A}h$ be bounded on R^n. For our system Σ, $C \subset \tilde{B}_0$.

Lemma: If $h \in C \cap C^2$ (bounded and twice differentiable) and $Dh(x)$ is bounded on R^n, then h is in the domain of $\tilde{A}$ and $\tilde{A}h = Dh$.

Proof: For any fixed $\hat{x}$ and any neighborhood $U = \{x\colon ||x-\hat{x}|| < \delta\}$

$(1/t)(T_t h(\hat{x}) - h(\hat{x})) = I_1(t) + I_2(t) + I_3(t)$, where

$I_1(t) = (1/t)\int_U [f(y) - f(\hat{x})]P(\hat{x},t,dy)$,

$I_2(t) = (1/t)\int_{R^n} f(y)P(\hat{x},t,dy)$, $I_3(t) = -f(\hat{x})(1/t)[1 - P(\hat{x},t,U)]$.

Let $k = \sup_x |h(x)|$. Given $\varepsilon > 0$, from the property $N'(G)$ there exists t_1 such that $|I_3(t)| < k\varepsilon$ and $|I_2(t)| < k\varepsilon$, $0 \leq t \leq t_1$; if δ is sufficiently small, from $h \in C^2$ we have $|I_1(t) - Df(x)| < \varepsilon$, $0 \leq t \leq t_1$. Df is continuous and bounded by hypothesis, so $I_1(t) \xrightarrow{w} Df(x) = \tilde{A}f(x)$, proving the lemma.

v is then in the domain of $\tilde{A}$ and $\tilde{A}v - \lambda v = 0$; by Theorem 1.7 of [Dynkin], the resolvent $(\lambda - \tilde{A})^{-1}$ exists and $v = 0$, contradicting the original supposition that $u \neq 0$. (End of Proof)

(3.6) Theorem.

If a_o and a_1 are C^ω, globally Lipschitzian (2.1b) and if the system S is $\hat{T}$-controllable, then

(I) Σ is strongly Feller (i.e., $T_t f \in C$ for all $f \in B(R^n)$);

(II) if G is an open set, for $x \in R^n$ there exists a sequence of times $\Theta_i(x,G) \uparrow \infty$ such that $P(x, \Theta_i, G) > 0$, $i = 1, 2, \ldots$.

Proof of (I): Theorem 2.10 and Theorem 3.5.

Proof of (II): Step 1. Suppose there exist a time $r \geq 0$, a point z_p and an open set G_p such that for all $t \geq r$, $P(z_p, t, G_p) = 0$. We shall obtain a contradiction as follows.

Step 2. By Property (I), for $t > 0$ $P(x,t,G)$ is continuous in (x,t). Let us define, for $G \in \underline{B}$ and $x \in R^n$,

$$\Gamma_r(x,G) \triangleq \int_r^\infty e^{-t} P(x,t,G)\,dt;$$

this function is continuous in x and non-negative. Let

$Z \triangleq \{z \in R^n : \Gamma_r(z, G_p)\} = 0$; note $z_p \in Z$;

$G_\delta \triangleq \{x : \Gamma_r(x, G_p) > \delta\}$, $\delta > 0$;

$G_o \triangleq \bigcup_{\delta > o} G_\delta$. The sets G_δ, G_o are open.

In Steps 4 and 5, below, we shall show there exist $z_o \in Z$ and a time

$T \geqq 0$ such that $P(z_o, T, G_o) > 0$; then there exists $\delta > 0$ such that $P(z_o, T, G_\delta) > 0$. From Eq.(3.2d), for all $s \geqq 0$

$$P(z_o, T+s, G_p) \geqq \int_{G_\delta} P(y,s,G_p)P(z_o,T, dy);$$

$$\int_r^\infty e^{-s}P(z_o, T+s, G_p)ds \geqq \int_{G_\delta} \Gamma_r(y,G_p)P(z_o,t,dy);$$

$$e^T \Gamma_{T+r}(z_o, G_p) \geqq \delta P(z_o, T, G_\delta) > 0.$$

Therefore there exists $t_z \geqq T + r$ such that $P(z_o, T_z, G_p) > 0$, contradicting the supposition of Step 1 and proving Property (II).

Step 3. If $z_p \in G_o$, $P(z_p, 0, G_o) = 1$ and there is nothing to prove. So take $z_p \in R^n - G_o$. [It may be that $z_p \in G_p$, since G_p need not be a subset of G_o.]

Choose any point $z_1 \in G_o$ and any $T_o > \hat{T}(z_p, z_1)$; construct a control $u_o \in \underline{L}^2$ for system S such that $X(z_p, T_o; u_o) = z_1$, and let us denote the corresponding trajectory in R^n, parametrized by σ, by

$$\Lambda = \{x = X(z_p,\sigma, u_o): 0 \leqq \sigma \leqq T_o\} \subset R^n.$$

If σ is sufficiently near T_o, $x_\sigma \in G_o$; there exists $\sigma_o \geqq 0$ such that $\Gamma_r(x_{\sigma_o}, G_p) = 0$ and $\Gamma_r(x_\sigma, G_p) > 0$, $\sigma_o < \sigma \leqq T_o$.

Now (without loss of generality) we shall take the point x_{σ_o} to be the origin 0 of the x-coordinates of R^n. It is necessary that we treat the cases $a_1(0) \neq 0$, $a_1(0) = 0$ separately.

Step 4. Suppose $a_1(0) \neq 0$; then $a_1(x) \neq 0$ on some neighborhood $U(0)$. We can construct a C^ω homeomorphism ξ taking $U(0)$ into a neighborhood $\xi U(0)$ in a copy of R^n, $(\hat{x} = \xi(x), x = \xi^{-1}(\hat{x}))$ in such a way that the image $\hat{a}_1$ of $a_1(\cdot)$ is a constant vector; $\hat{x}_t = \xi(x_t)$ satisfies (x_t a response of S)

$$\frac{d\hat{x}}{dt}t = \frac{\partial \xi}{\partial x} a_o(\xi^{-1}(\hat{x}_t)) + u(t)\frac{\partial \xi}{\partial x} a_1(\xi^{-1}(\hat{x}_t) = \hat{a}_o(\hat{x}_t) + u(t)\hat{a}_1$$

(see [Hermann 1], Ch.6, for details). The $\hat{x}_1$-coordinate curves $\hat{x}_t = \hat{x}_o + \hat{a}_1 t$ are ξ-images of solution curves of the differential equation $dx/dt = a_1(x)$ (with initial state on some suitable hypersurface through the origin); the n-1 other coordinate functions $\xi_i(x)$ are solutions of the partial differential equation

$$\alpha_1 \xi_i = 0 \ , \quad i = 2, \ldots, n$$

that are functionally independent on $U(0)$.

Using Itô's Lemma (Section 3.3) for the n functions $\hat{\underline{x}}_t^i = \xi_i(\underline{x}_t)$, we obtain a stochastic differential equation for trajectories on $U(0)$:

$d\hat{\underline{x}}_t = D\xi(\underline{x}_t)dt + \alpha_1 \xi(\underline{x}_t)dw_t$; on $\xi[U(0)]$ this is written

a) $\quad d\hat{\underline{x}}_t = \hat{a}_o(\hat{\underline{x}}_t)dt + \hat{a}_1 dw_t$

and we cut off the trajectories at the stopping time τ_u of first exit from $\xi[U(0)]$. Now extend the vector function $\hat{a}_o(\hat{x})$ to the rest of R^n in a C_o^∞ way; then we can construct a new Markov process, no longer cut off, which we denote by $\hat{\Sigma}$. Since $\hat{a}_1$ is constant in the defining Eq.(a), we can construct the trajectories of $\hat{\Sigma}$ in a new way, which makes it possible to interpret trajectories of S as trajectories of the stochastic process; the construction is due to [Lamperti] (cf. [Mortensen 1966]) and is given in Ch.3 of [McKean]:

For each $w \in C_b$, the (ordinary!) integral equation

a') $\quad \hat{\underline{x}}_t = \hat{x}_o + \int_o^t \hat{a}_o(\hat{\underline{x}}_s)ds + \hat{a}_1 w_t$

has a unique continuous solution (in the usual sense) $\hat{x}_t(w)$, $t \geq 0$ obtained by successive approximation in the uniform norm (not q.m.). Given two samples w', w'' from C_b, if $\underline{x}'$, $\underline{x}''$ are corresponding solutions,

b) $\quad ||x_t' - x_t''|| \leq e^{Kt}||a_1|| \sup_{[0,t]} |w_s' - w_s''| \ , \ t \geq 0.$

That is, the map $w \to \hat{x}_t(w)$ is continuous on the Banach space $C_b[0,t]$. $\hat{x}_t(w)$ is also an Ito solution in $\underline{H}$, but each equivalence-class of solutions now has one member (this subclass of $\underline{H}$ is not complete in q.m.).

Now, once again cut off the trajectories of $\hat{\Sigma}$ at the time τ_u of first exit from $\xi[U(0)]$. The random function $\xi^{-1}(\hat{x}_t(w))$, $0 \leqq t < \tau_u$, is a representative of the equivalence-class $\underline{x}_t$ (of Σ trajectories) and is also a continuous functional, for given t, on the space $C_b[0,t]$.

We will write $x_t^w = \xi^{-1}(\hat{x}_t(w))$ to denote these representatives; by uniqueness, for any given $x_o \in U(0)$, our new solution and the Itô solution are related by $\underline{x}_t = x_t^w$ in q.m., if the process is cut off at τ_u.

Now let $w_t^\circ = \int_0^t u_o(s)ds$, where u_o is the control constructed in Step 3. Then the arc of the trajectory Λ from 0 to z_1 is given by $x_t^{w^\circ}$, $0 \leqq t \leqq T_o - \sigma_o$, given that 0 is the initial state. Choose T in that interval. In C_b the set $\{w: x_T^w \in G_o \cap U(0)\}$ contains w° and is open and thus has positive probability. To put it another way, Eq.(b) shows that (while the trajectories are in U(0)) trajectories of Σ starting at 0 stay near the trajectory Λ with positive probability. Then we may choose the origin 0 to be the required point z_o; $P_0\{\underline{\mathbf{x}}_T \in G_o\} > 0$.

Step 5. Assume now that $a_1(0) = 0$. There are two possibilities. If there are points x on the boundary $\partial G_o \subset Z$ such that $a_1(x) \neq 0$, then we modify the control u so that Λ passes into G_o at such a point , and we are back in the Step 3 situation, taking this point x as the origin. On the other hand, if $a_1(x)$ vanishes everywhere on ∂G_o, then the boundary is part of a real-analytic variety of dimension less than n. That is, ∂G is a hypersurface on which $||a_1(x)||^2 = 0$, and is piecewise C^ω (implicit function theorem) and we can assume that the origin is a good point. Now construct a little truncated cone with vertex at 0, axis in the $a_o(0)$ direction, lying completely inside G_o.

That is, for small enough λ and ρ, $\{x: a_o(0)\cdot x > \lambda||x||, ||x|| < \rho\} \subset G_o$. As in Remark 3 to Th. 11.4 of [Dynkin], for the trajectory starting at the origin at time 0 we have, for small t,

$\underline{x}_t = t\, a_o(0) + \eta_t$, where $E_0||\eta_t||^2 \leq ct \sup_{[0,t]} E_0||\underline{x}_s||^2$. Since $(1/t)\sup_{[0,t]} (E_0||\underline{x}_t||^2 - ||x_o||^2) = D||x_o||^2 + o(t)$ at $x_o = 0$, the Chebyshchev inequality gives us $P_0\{||\eta_t||^2 > t^2/\lambda^2\} \leq c\lambda^2 o(t)$, where $o(t)$ means a quantity that vanishes with t. For any sufficiently small t the probability that $\underline{x}_t$ lies in the cone is therefore positive. Thus we again conclude that $P_0\{\underline{x}_t \in G_o\} > 0$.

That concludes the proof of property (II).

Note 1: Since any $T > \sigma_o$ will suffice, it is evident that $\Theta_1(0, G_p) \leq \hat{T}(0, z_p)$. Then $\Theta_1(x, G) = \inf_{y \in G} \hat{T}(x,y)$, as one might expect.

Note 2: The results in Sections (3.1-3.6) can all be generalized to

$$d\underline{x}_t = a_o(\underline{x}_t)dt + (1/2)\sum_{\nu=1}^{r}(a_\nu\cdot\partial/\partial x)a_\nu(\underline{x}_t)dt + \sum_{\nu=1}^{r} a_\nu(\underline{x}_t)dw_t^\nu$$

where $w^1, \ldots, w^r$ are independent Brownian motions, and $r \leq n$. (Compare Remark 2.12b.) In Theorem 3.6, this generalization must be restricted: either $r = 1$, as in the above proof, or if $r \geq 2$ there must exist a coordinate system in which $a_1, \ldots, a_r$ are all constant.

The reason for this restriction is that $\underline{x}_t$ may not be a continuous functional on the r-fold Cartesian product $C[0,t]\times\cdots\times C[0,t]$. R. E. Mortensen [private communication] has provided this example for n=3, r=2:

$$d\underline{x}_t = dw_t^1, \quad d\underline{y}_t = dw_t^2, \quad d\underline{z}_t = \underline{y}_t dw_t^1 - \underline{x}_t dw_t^2.$$

To see the failure of continuity at $w^1 = 0$, $w^2 = 0$, consider the (sure) null sequence of "sample functions" $w^1(k) = k^{-1/2}\sin(2\pi kt)$, $w^2(k) = k^{-1/2}\cos(2\pi kt)$, $k = 1, 2, \ldots$. At time $t = 1$, $z_1 = 2\pi + z_o$, for all k; but the appropriate norm $||w^1(k)||_\infty + ||w^2(k)||_\infty \to 0$.

CHAPTER 4. INVARIANT PROBABILITY MEASURES

(4.1) Remarks.

Suppose that the initial state x of Σ is given a probability distribution on R^n, denoted by the measure μ ; that is, $P\{x \in G\} = \mu(G)$. If at every time $t > 0$ we have $P\{x_t \in G\} = \mu(G)$, we say μ is an invariant probability measure for Σ. In the following sections we will give formal definitions and prove a new existence theorem.

It would be possible at this point to extend to our slow diffusions the work of [Khasminskii] on invariant measures, except that the criterion he gives (involving a Dirichlet problem for the differential generator D) depends on the assumption that there exist arbitrarily large compact sets with regular boundaries. (If an open set G has a smooth boundary G and τ is the time of first exit from G, a point $x \in G$ will be regular if the limit as $y \to x$, $y \in G$, of $E_y\tau$ is zero.)

It can be seen from the problem in [Franklin and Rodemich] that for slow diffusions (even if $\hat{T}(\cdot,\cdot) = 0$, when Khasminskii's conditions 1°- 3° are satisfied) this regularity assumption needs checking. Consider their problem: Σ is described on R^2 by $d\underline{x}_t = dw_t$, $d\underline{y}_t = \underline{x}_t dt$; G is some open set lying above the x-axis with the segment $(0,1)$ of the x-axis as part of G. Points on this segment are not regular. For the relevant Dirichlet problem $Du = (1/2)u_{xx} + xu_y = -1$, boundary values cannot be prescribed on the segment. (Compare [Fichera].)

A construction has been announced by [Bony] that gives small sets with regular boundaries (corners are regular, curiously!). Such a result will permit the construction of invariant measures by the method of [Khasminskii], which also requires a regularity condition.

(4.2) Set functions.

If the domain of an integral is not stated, it is R^n.

Let $\underline{M}$ denote the space of σ-finite countably additive set functions ("signed measures") defined on $\underline{B}$, the Borel sets, and taking values on $(-\infty,\infty]$. Given a sequence $\{\mu_k\}$, we say $\mu_k \xrightarrow{w} \mu$ if for all

$f \in C$ $\quad \int f(x)\mu_k(dx) \to \int f(x)\mu(dx)$ as $k \to \infty$;

this is called weak convergence. Giving $\underline{M}$ the topology of weak convergence, it is a linear topological vector space [Dunford and Schwartz].

We define a family of linear operators $\{L_t,\ t \geq 0\}$ on $\underline{M}$ into $\underline{M}$, by

a) $\quad L_t\mu(G) = \int P(x,t,\ G)\mu(dx),\ G \in \underline{B}$.

Lemma 1. Each operator L_t is continuous on $\underline{M}$; the family $\{L_t\}$ is a semigroup, weakly continuous at $t = 0$.

Proof: Suppose we are given $\mu_k \xrightarrow{w} 0$ as $k \to \infty$. Σ is a Feller process ($T_t C \subset C$), so for all $f \in C$

$$\int f(x)L_t\mu_k(dx) = \int T_t f(x)\mu_k(dx) \to 0.$$

From Eq. (a), $L_t L_s = L_{t+s} = L_s L_t$. For all $f \in C$ and all $\mu \in \underline{M}$

$$\int f(x)[L_t\mu(dx) - \mu(dx)] = \int [T_t f(x) - f(x)]\mu(dx) \to 0 \text{ as } t \downarrow 0.$$

Definition: Let $\underline{M}+ = \{\mu \in \underline{M} : \mu(\cdot) \geq 0\}$ denote the measures on $\underline{B}$; $\Pi = \{\mu \in \underline{M} : \mu(\cdot) \geq 0,\ \mu(R^n) = 1\}$ denote the probability measures. If there exists $\tilde{\mu} \in \Pi$ such that $L_t\tilde{\mu} = \tilde{\mu}$ for all $t > 0$, $\tilde{\mu}$ is called an invariant probability measure for Σ.

To establish the existence of an invariant probability measure (under additional "stability" hypotheses), we must now develop a suitable criterion for compactness of sets in Π.

(4.3) Compactness and gauge functions.

We will say that a sequence $\mu_k \in \Pi$ is u.s.b., or uniformly stochastically bounded, if for every $\varepsilon > 0$ there exists $\rho(\varepsilon) > 0$ such that for $k = 1, 2, \ldots$, $\mu_k\{x: ||x||^2 > \rho\} < \varepsilon$.

Lemma 1. (Helly's selection theorem): Given a u.s.b. sequence $\mu_k \in \Pi$ there exists $\mu \in \Pi$ and a sub-sequence μ_{k_i} such that $\mu_{k_i} \xrightarrow{w} \mu$.

The proof is in [Feller], pages 243, 247, 261. Also compare [7], Theorem 0.20B (there w-convergence is with respect to $\hat{C}$; if the measures are u.s.b. then C can be used).

Definition. A gauge function $V(\cdot)$ is a non-negative twice continuously differentiable function on R^n such that $V(x) \to \infty$ as $||x|| \to \infty$ and $V(x) < K_1 + K_2||x||^2$ for some positive constants K_1 and K_2.

From that last requirement we see that for any finite x and t, $E_x V(\underline{x}_t)$ is finite. This is important in what follows.

Given a gauge function $V(\cdot)$ and some $\mu \in \Pi$, suppose $\int V(x)\mu(dx) = \gamma$. From Chebyshchev's inequality, for any positive λ $\mu\{x: V(x) > \lambda\} < \gamma/\lambda$. Then $\mu\{x: ||x||^2 > (\lambda - K_1)/K_2\} < \gamma/\lambda$; but that is the condition used in defining u.s.b. .

Given $V(\cdot)$ and any number $\gamma > \inf_x V(x)$, the set of measures

$$\Phi(V; \gamma) = \{\mu \in \Pi: \int V(x)\mu(dx) \leq \gamma\}$$

is non-empty. Construct a sequence of bounded functions $\{v_i \in C\}$ which converges monotonely to V from below. $\Phi(v_i; \gamma)$ is the intersection of a closed convex cone, a hyperplane, and a half-space in $\underline{M}$, all defined by functionals continuous on $\underline{M}$. Therefore $\Phi(v_i; \gamma)$ is weakly closed. Then $\Phi(V; \gamma) = \bigcap_i \Phi(v_i; \gamma)$ is weakly closed. Using Lemma 1 we conclude that $\Phi(V; \gamma)$ is weakly compact.

Lemma 2. (The Markov-Kakutani fixed point theorem):
Let Φ be a compact convex subset of a linear topological space. Let $\{L_t\}$ be a commuting t-indexed family of continuous linear mappings which map Φ into itself. Then there exists $\mu \in \Phi$ such that $L_t\mu = \mu$ for all t. (This is given as Theorem V.10.6 in [Dunford and Schwartz].

It is now clear that if we can find a gauge function $V(\cdot)$ and a number γ such that $\Phi(V; \gamma)$ is mapped into itself by each L_t, then we will be assured of the existence of an invariant probability measure.

Following [Wonham] and [Kushner 1967,1969] we use gauge functions in a way analogous to the Lyapunov functions of control theory; $V(\underline{x}_t)$ will not decrease along trajectories, but we can require its expectation to decrease with time when x_o is sufficiently far from the origin. If $DV(x)$ is negative (D is the differential generator of Σ) for large $||x||$, then the following lemma will be useful.

Lemma 3: If V is a gauge function and $|DV(x)| < K_3(1 + ||x||^2)$,
$E_x V(\underline{x}_t) = V(x) + \int_o^t E_x DV(\underline{x}_s)ds.$

Proof: (Similar results are given by Kushner.)
$\{G_k\}$ is a sequence of nested open sets with compact closures, increasing to fill up R^n. Let τ_k be the time of first exit from the interior of G_k. Let $V_k(x) = V(x)$ when $x \in G_k$, but outside G_k let V_k be bounded and twice continuously differentiable. Then V_k is in the domain of $\tilde{A}$ (see Step 5 of Th. 3.5). Denote $t \cap \tau_k = \min(t, \tau_k)$; it is a stopping time with expectation less than t, so we can apply the Corollary to Theorem 5.1 of [Dynkin]:

$$E_x V_k(\underline{x}_{t\cap\tau_k}) = V_k(x) + E_x \int_o^{t\cap\tau_k} \tilde{A} V_k(\underline{x}_s)ds.$$

x_s is bounded on $[0, t]$ and $\tau_k \to \infty$ as $k \to \infty$, almost surely. Then $t\cap\tau_k \to t$

almost surely. Then $V(\underline{x}_{t \cap \tau_k}) \to V(\underline{x}_t)$ a.s.. For any fixed x, the LHS of Eq.(a) is bounded in magnitude by $\sup_{[0,t]} V(\underline{x}_s)$, which has a finite expectation; so by Lebesgue's dominated-convergence theorem, for all x the LHS converges to $E_x V(\underline{x}_t)$.

For all s in the interval $[0, t \cap \tau_k]$, $V_k(\underline{x}_s) = V(\underline{x}_s)$ and using the lemma of Step 5, Theorem 3.5, we see $AV(\underline{x}_s) = DV(\underline{x}_s)$. Therefore the RHS of Eq.(a) converges to the limit $V(x) + \int_0^t E_x DV(\underline{x}_s)ds$.

(4.4) <u>Theorem</u>. If there exist a gauge function $V(\cdot)$ and positive constants K_3, c_1, c_2 such that $|DV(x)| < K_3(||x||^2 + 1)$ and

$DV(x) < c_1 - c_2 V(x)$ for all x,

then there exists an invariant probability measure μ for Σ.

<u>Proof</u>: Let γ be any number greater than c_1/c_2. Denote $g(x,t) = E_x V(\underline{x}_t)$, then we have the integral inequality

$$g(x,t) < V(x) + \int_0^t c_1 - c_2 g(x,s)ds$$
$$< c_1/c_2 + (V(x) - c_1/c_2)\exp(-c_2 t).$$

With our choice of γ, if $\mu \in \Phi(V;\gamma)$

$$\int V(x) L_t \mu(dx) = \int g(x,t)\mu(dx) \leq \gamma .$$

The conclusion follows from the results of the preceding Section.

<u>Discussion</u>: An example in R^2 may be useful.

$dx/dt = y - x, \qquad dy/dt = -y - \tanh(x) + u(t)\exp(-x\ );$

let $V(x) = \log\cosh(x) + y^2/2$, then $DV(x) = -y^2 - x\tanh(x) + (1/2)\exp(-2x\ )$ and a simple calculation shows $DV(x) < 1/2 - V(x)$, $|DV(x)| < 1/2 + x^2 + y^2$ and the K-controllability follows from example (2.12e). C^{ω} is obvious, so the corresponding process Σ **satisfies the Theorem, as well as the**

<u>Proposition</u>: If Σ satisfies the hypotheses of Theorem 3.6, then the support of any invariant measure is all of R^n.

(4.5) Proposition. If μ is an invariant probability measure for Σ and $\mathrm{Rank}(\Xi') = n+1$, then μ has a density q satisfying $D^*q = 0$.

Proof: Given that $L_t\mu = \mu$, $\int\phi(x)L_t\mu(dx) = \int\phi(x)\mu(dx)$ for all $\phi \in C_0^\infty(R^n)$ -- that is, $\int T_t\phi(x)\mu(dx) = \int\phi(x)\mu(dx)$, and differentiating w.r.t. time, $\int D(T_t\phi(x))\mu(dx) = 0$. μ as a Schwartz distribution has a formal density $q(\cdot)$ (compare Th. 3.5) such that $\int T_t\phi(x)D^*q(x)dx = 0$, $t > 0$. Since the class of functions $T_t\phi$, $t > 0$ is dense in $C_0^\infty(R^n)$, $D^*q(x) = 0$ (the "steady-state Fokker-Planck equation"); by hypo-ellipticity, q is a C^∞ function.

(4.6) Remarks.

For the hypothesis of Theorem 4.4 to be satisfied, it is sufficient that the system S, with zero input, possess a quadratic Lyapunov function V such that $SDV(x) < -c_2V(x)$ and that $||a_1(x)|| < c_1$. This implies that S must be exponentially stable, with exponent $-c_2$, and have the BIBO (bounded input gives bounded output) property. We conjecture

(a) the BIBO property (with a few restrictions) is enough to establish the conclusion of Theorem 4.4 given that $\mathrm{Rank}(\Xi') = n+1$; and

(b) to establish the existence of a unique invariant measure with support = R^n, it should suffice to require $\hat{T}$-controllability, real-analyticity, and the existence of a gauge function V such that $DV < 0$ for sufficiently large $||x||$.

BIBLIOGRAPHY

Balakrishnan, A. V. "On the controllability of a nonlinear system." Proc. Nat. Acad. Sci., 55:465-468, March, 1966.

Bony, J.-M. "Probleme de Dirichlet et inegalite de Harnack pour une classe d'operateurs elliptiques degeneres du second ordre." Comptes Rendues Acad. Sci. Paris, 266: A763-A765, April 8, 1968.

Cesari, L. "Existence theorems for optimal solutions in Lagrange and Pontryagin problems." SIAM Journal on Control, 3:475-498, 1965.

Chevalley, Claude. Theory of Lie groups I. Princeton, New Jersey, Princeton University Press, 1946.

Dunford, Nelson and J. Schwartz. Linear Operators, Part I. New York, Wiley (Interscience), 1964.

Dym, H. "Stationary measures for the flow of a linear differential equation driven by white noise." Trans. Am. Math. Soc., 123:130-164, May, 1966.

Dynkin, E. B. Markov Processes. Berlin, Springer-Verlag, 1965.

Elliott, D. L. "Behavior of controllable nonlinear systems driven by white noise." Second Hawaii International Conference on System Sciences. Proceedings. University of Hawaii, January 22-24, 1969, Honolulu, Hawaii; North Hollywood, California, Western Periodicals Company, 1969. p. 859-862.

Feller, William. An introduction to probability theory and its applications, Vol. II. New York, Wiley, 1966.

Fichera, G. "On a unified theory of boundary value problems for parabolic equations of second order," in Boundary problems in differential equations, edited by R. E. Langer, Madison, Wisconsin, University of Wisconsin Press, 1960. p. 97-120.

Franklin, J. N. and E. R. Rodemich. "Numerical analysis of an elliptic-parabolic partial differential equation." SIAM Journal on Numerical Analysis, 5:680-716, December, 1968.

Haynes, G. W. "Controllability of nonlinear systems." Natl. Aero. and Space Admin., Contractor Report, NASA CR-456, 1965.

Hermann, Robert. Differential geometry and the calculus of variations. New York, Academic Press, 1968.

Hermann, R. "On the accessibility problem in control theory," in International Symposium on Nonlinear Differential Equations and Nonlinear Mechanics, edited by J. LaSalle and S. Lefschetz, New York, Academic Press, 1963.

Hermes, H. "Controllability and the singular problem." SIAM Journal on Control, 2:241-260, 1965.

Hormander, L. "Hypoelliptic second order differential equations." Acta Math., 119:147-171, 1967.

Khasminskii, R. Z. "Ergodic properties of recurrent diffusion processes and stabilization of the Cauchy problem for parabolic equations." Theory of Probability and Its Applications, 5:179-196, 1960.

Kushner, Harold J. Stochastic stability and control. New York, Academic Press, 1967.

Kushner, H. J. "The Cauchy problem for a class of degenerate parabolic equations and asymptotic properties of the related diffusion processes." To be published in Journal of Differential Equations, 1969.

Lamperti, J. "A simple construction of certain diffusion processes." J. Math. Kyoto, 4:161-170, 1964.

Lee, E. Bruce and L. Markus. Foundations of optimal control theory. New York, Wiley, 1967.

Maruyama, G. and H. Tanaka. "Ergodic property of N-dimensional recurrent Markov processes." Mem. Fac. Sci. Kyushu U., Series A, 13:157-172, 1959.

McKean, Henry P. Stochastic Integrals. New York, Academic Press, 1969.

Mortensen, R. E. "Mathematical problems of modelling stochastic systems." Natl. Aeron. and Space Admin., Contractor Report, NASA CR-1168, September, 1968.

Skorokhod, A.V. Studies in the theory of random processes. Reading, Massachusetts, Addison-Wesley, 1965.

Stratonovich, R. L. "A new representation for stochastic integrals and equations." SIAM Journal on Control, 4:362-371, 1966.

Viterbi, Andrew J. Principles of coherent communication. New York, McGraw-Hill, 1966.

Wong, E. and M. Zakai. "On the convergence of ordinary integrals to stochastic integrals." Ann. Math. Stat., 36:1560-1564, 1965.

Wonham, W. M. "Liapunov criteria for weak stochastic stability." Journal of Differential Equations, 2:195-207, 1966.