

CASE FILE COPY

HAROLD J. KUSHNER

STOCHASTIC STABILITY

FEBRUARY 1972

N72. 25559



CENTER FOR DYNAMICAL SYSTEMS

STOCHASTIC STABILITY

by

H. J. Kushner^{*}

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

^{*}This research was supported in part by the Air Force Office of Scientific Research, AF-AFOSR-71-2078, in part by the National Aeronautics and Space Administration, NGL 40-002-015, and in part by the National Science Foundation, GK 31073X.

Stochastic Stability

by

H. J. Kushner

Brown University

Abstract

This paper will survey the field of stochastic stability, with special emphasis on the "invariance" theorems and their potential application to systems with randomly varying coefficients. First, we will survey some of the basic ideas underlying the "stochastic Liapunov function" approach to stochastic stability, then the invariance theorems will be discussed in detail and, if time permits, an example given.

In stability analyses of all types of deterministic dynamical systems, the concepts of ω -limit set and invariant set play an important role. Let x_t , $t \geq 0$, denote a bounded continuous solution to the deterministic ordinary differential equation $\dot{x} = f(x)$, where $f(\cdot)$ is continuous. A point x is said to be in the ω -limit set if $x_{t_n} \rightarrow x$ for some sequence $t_n \rightarrow \infty$. A set B is said to be an invariant set if for each $x \in B$, there is a path y_t , $t \in (-\infty, \infty)$ for which $y_t \in B$ and $\dot{y}_t = f(y_t)$, $t \in (-\infty, \infty)$. Thus, the invariant set contains entire trajectories. It is well known that the ω -limit set is a closed non-empty invariant set. Also, $V(\cdot)$ be a Liapunov function with $\dot{V}(x) = -k(x) \leq 0$, where $k(\cdot)$ is continuous. Then by a very useful theorem of LaSalle, x_t tends to the largest invariant set contained in $\{x = k(x) = 0\} \equiv K$, as $t \rightarrow \infty$. The theorem is important and useful, since in numerous applications the derivative $k(\cdot)$ of the Liapunov function $V(\cdot)$ is semi-definite, and the theorem gives a nice characterization of the subset of K to which x_t tends. In fact, this characterization often enables us to determine the minimal set in K to which x_t tends.

There are useful stochastic analogs to all the deterministic results (the dynamical system is a suitable dynamical system of measures), and they will be developed and explained. The concepts are felt to be useful for random parameter systems for the following reason. Suppose $\dot{x} = f(x, y)$, where the "parameter process" y_t is a Markov process. Then (x_t, y_t) will be Markov. Often, we are only interested in the asymptotic properties of x_t . Indeed y_t may even be stationary. Any stochastic Liapunov function would have to take both x_t, y_t into account in some way, and the "stochastic derivative" of that Liapunov function will often be semi-definite (it can't be a negative definite function of (x_t, y_t) if y_t is stationary). The invariance theorems enter here to assist us in studying the asymptotic properties of the x_t process.

1. Introduction

In this paper, we will discuss and prove some of the basic results in the theory of stochastic stability for systems governed by continuous time Markov processes. Our concern will be mainly with the asymptotic behavior of the paths of the process. The development will be along the lines of [1], [2]. A detailed and introductory discussion of stochastic stability for discrete parameter systems appears in [3]. In fact, [3] contains introductory discussions of asymptotic stability, the invariance theorems, the existence of (and convergence of the measures of the process to) invariant measures, and a number of examples.

Next, we give some definitions, then mention some of the problems with which stochastic stability deals. Then we give a brief introduction to some deterministic results, and discuss some of the probabilistic structures to be used in the sequel. Then some results on asymptotic stability w.p.l. will be discussed and proved, and an invariance theorem proved. Finally we give two examples, one dealing with a non-linear diffusion, and the other with a problem arising in the identification of the parameters of a linear differential equation. The paper will be as self contained as space permits. For the most part, the discussion will concern the case where the transition functions are homogeneous - since statements and proofs are notationally simpler there, but some results for the non-homogeneous case will also be stated.

Stochastic stability is a long way from being a mature subject, even from the theoretical point of view. Let u_t denote a Markov process. Then under some suitable condition on $f(\cdot)$, the process $x_t = (y_t, u_t)$ where y_t

given by $\dot{y} = f(y, u)$ is a Markov process, and we may desire to investigate whether $y_t \rightarrow 0$ w.p.1. In many applications u_t is stationary, or at least its paths do not converge. Yet the Liapunov functions must take both u_t and y_t into account. It is not clear what the appropriate theorems are for such cases, nor is it understood how to find (even in relatively simple cases) useful Liapunov functions. It is hoped that a combination of the stability results of part I, and the invariant set results of part II will be helpful here, since the "derivatives" of the Liapunov functions will probably be semidefinite in such cases. A relatively simple example is given in Example 2 in part III. Investigations into further possibilities are continuing.

I. Stochastic Stability.

1. Markov and Strong Markov Processes.

For our purposes a Markov process can be defined in the following way. Let X be a topological space (called the state space) and $\Omega, \mathcal{B}$ and $P_x\{\cdot\}$, a sample space, a σ -algebra on Ω and family of probability measures on $(\Omega, \mathcal{B})$ (for each $x \in X$ there is a measure $P_x\{\cdot\}$). Let $x_t, t \geq 0$ be a family of random variables from $(\Omega, \mathcal{B})$ to X with $\mathcal{B}_t \subset \mathcal{B}$ being the least σ -algebra which measures $x_s, s \leq t$. Let there exist a real valued function $P(\cdot, \cdot, \cdot)$ (called the transition function) on $X \times [0, \infty) \times \mathcal{B}(X)$, where $\mathcal{B}(X)$ is the σ -algebra on X which is induced by the topology on X . Let $P(\cdot, t, \Gamma)$ be measurable on $\mathcal{B}(X)$ for each $t \geq 0$ and $\Gamma \in \mathcal{B}(X)$, and let $P(x, t, \cdot)$ be a probability measure on $\mathcal{B}(X)$.

Furthermore, let $P(x, t, \Gamma) = P_x\{x_t \in \Gamma\}$ and

$$(1) \quad P_x\{x_{t+s} \in \Gamma \mid \mathcal{B}_s\} = P(x_t, s, \Gamma) \quad \text{w.p.1.}$$

for each $s \geq 0$, $t \geq 0$, $x \in X$, $\Gamma \in \mathcal{B}(X)$. Then we say that $(\Omega, P_x, \mathcal{B}_t, \mathcal{B}, x_t)$ (or simply that the process x_t) is a homogeneous Markov process.

Note: By (1) we can write (1) as $P_{x_s}\{x_t \in \Gamma\}$. The subscript x_s denotes that the initial condition is a random variable with the distribution of x_s . The argument $x_t \in \Gamma$ indicates the event whose probability is being written, under the given initial condition.

The definition of a non-homogeneous Markov process is similar - but we must keep track of two time indices - rather than one. Then we define the transition function $P(\cdot, \cdot; \cdot, \cdot)$ by $P_{x,t}\{x_s \in \Gamma\} = P(x, t; t+s, \Gamma)$ (the probability that with initial condition x at time t , the process is in Γ s units of time later).

Eqn (1) implies that $P(\cdot, \cdot; \cdot, \cdot)$ satisfies the Chapman-Kolmogorov equation

$$(2) \quad P(x, t+s, \Gamma) = \int P(x, t, dy) P(y, s, \Gamma), \quad s \geq 0, t \geq 0.$$

For our purposes, the Markov process is slightly too broad a class of processes, for the following reason. Let x_t be a Markov process whose state space is the real line. Define $\tau(\omega)$ by

$$\tau(\omega) = \inf\{t: x_t = b > 0\}$$

and suppose that $\tau(\omega) < \infty$ w.p.1. Then it is not necessarily true (even

if the terms are well defined) that w.p.1.

$$(3) \quad P_{x_{\tau+t}} \{x_{\tau+t} \in \Gamma \mid x_s, s \leq \tau\} = P_{x_{\tau}} \{x_t \in \Gamma\}.$$

(Observe that $x_{\tau} = b$ in the example.) In words - the distribution of the process t units after first hitting b , conditional upon all the path data up to the first hitting time, may depend on how we arrived at b and not simply on the fact that $x_{\tau} = b$. (For an example see Loeve [4], p. 578). Indeed, (3) should be true for a process arising in a physical application, and we will restrict our attention (without apparant loss of generality) to processes where (3) is true for a large class of random variables τ .

Definition. A non-negative random variable τ (defined on a set $\Omega_{\tau} \subset \Omega$) is called a Markov time if

$$\{\tau \leq t\} \in \mathcal{D}_t;$$

i.e., τ is a Markov time if we can tell whether or not $\tau \leq t$ by watching the process x_s up to time t only, for each $t \geq 0$.

Definition. If (3) holds for all Markov times τ , then x_t is said to be a strong Markov process.

Definition. We will consider only strong Markov processes. If x_t is a Markov process for which the function of x given by $E_x g(x_t)$ is continuous for each $t > 0$ and real valued, continuous and bounded $g(\cdot)$, then x_t is a Feller process.

A Feller process whose paths are continuous from the right is a strong Markov process. ([5], Theorem 3.10).

Definition. Let B denote the Banach space of real valued bounded measurable functions on X , and B_0 the subset of B for which

$$E_x f(x_t) \rightarrow f(x), \quad f \in B_0$$

weakly as $t \rightarrow 0$. If the weak limit

$$\frac{E_x f(x_t) - f(x)}{t} \xrightarrow{w} g(x),$$

exists (i.e., there is pointwise convergence, and the left hand side is bounded as $h \rightarrow 0$) and is in B_0 , we say that $g(\cdot)$ is in the domain $\mathcal{D}(\tilde{A})$ of the weak infinitesimal operator $\tilde{A}$, and write $\tilde{A}f = g$.

Suppose τ is a Markov time and $E_x \tau < \infty$, and x_t is right continuous w.p.l., and $f \in \mathcal{D}(\tilde{A})$, (the continuity conditions can be weakened; see [5], p. 133) then we have the important relation (4), known as

Dynkins formula ([5], p. 133),

$$(4) \quad E_x f(x_\tau) - f(x) = E_x \int_0^\tau \tilde{A}f(x_s) ds.$$

The operator $\tilde{A}$ plays a role for Markov processes, similar to the role the differentiation operator plays for differentiable non-random real valued functions. Equation (4) is an analog of the deterministic integral - differential relationship and, as such, will play an important role in the sequel.

The non-homogeneous case. If x_t is non-homogeneous, or if we wish to apply (4) to functions $f(x, t)$ of both state and time, then we can proceed as follows. Define t to be a state of the process (replace X by $X \times [0, \infty)$). Redefine B, B_0 appropriately, let $f \in \mathcal{D}(\tilde{A})$ and $g = \tilde{A}f$ if $f(\cdot, \cdot) \in B$ and

$$\frac{E_{x,t} f(x_s, t+s) - f(x, t)}{s} \rightarrow g(x, t)$$

weakly as $s \rightarrow 0$, and $E_{x,t} g(x_s, t+s) \rightarrow g(x, t)$ weakly as $s \rightarrow 0$. Then under the conditions on τ in (4),

$$(4') \quad E_{x,t} f(x_\tau, t+\tau) - f(x, t) = E_{x,t} \int_0^\tau \tilde{A}f(x_s, t+s) ds.$$

In (4') we understand that time is measured from the origin t ; i.e., the value of x_s in (4') is the value of the state s units of time after the initial time t .

2. A Few Sources of Stochastic Stability Problems.

Stochastic stability problems occur in almost all phases of physics, control theory, numerical analysis and economics where dynamical models subject to random disturbances appear, and the process is of interest over a long period of time. Only a few simple problem types will be mentioned here.

Suppose that y'_t is a Markov process which drives the differential equation $\dot{y}_t = f(y'_t, y_t, \alpha)$, where α is a parameter. y'_t may represent an external driving term, or random variations in some parameter of the equation. We may be interested in the range of α for which $y_t \rightarrow 0$ w.p.l., or for which $|y_t|$ remains bounded in some statistical sense. For α fixed at α_0 , we may be interested in the range of initial conditions y, y' for which (for some $1 > \delta > 0$)

$$P_{y, y'} \{ \sup_{\infty > t \geq 0} |y_t| \geq \lambda \} \leq \delta.$$

E.g., y_t may represent a stress in a mechanical structure, and it may be of interest to keep the stress less than $\lambda > 0$. Also, it may be desirable to know whether

$$P_{y, y'} \{ \sup_{\infty > t \geq 0} |y_t| \geq \lambda \} \rightarrow 0$$

as y or y' or both tend to zero (a type of stability of the origin w.p.l.).

The above stability properties are all properties of the paths of the processes. There are many problems of interest concerning the asymptotic

behavior of the moments and of the measures of the process - and even in cases where the process is of interest for only a finite time [1], [3].

A large class of stability problems arise in tracking situations. For example, suppose that we are driving on a road and sample our instruments and errors (e.g., distance from the center of the lane and from other cars) somewhat irregularly (as is usually done), then can we track the center of the lane within a certain error, etc. Tracking problems arise in radar and machine tool systems.

Many types of stochastic convergence can be studied. Here we deal with w.p.l. convergence mainly. Other stability problems deal with (a) convergence w.p.l. to a set, (b) recurrence - the process always returns to a bounded set w.p.l., (c) no finite escape time w.p.l., (d) convergence or boundedness of certain moments of the process, (e) convergence of the distributions of the process to an invariant measure. Types (d-e) are usually more difficult to treat than w.p.l. convergence, but also are of considerable practical importance.

3. A Brief Review of Deterministic Stability.

Some results in deterministic stability are briefly reviewed because, in a certain abstract sense, the stochastic results are analogies of the deterministic results.

Let R^r denote Euclidean r -space, $f(\cdot)$ a continuous function from R^r to R^r and suppose that there is a continuous solution to the homogeneous differential equation $\dot{x} = f(x)$. Let $V(\cdot)$ denote a continuous, non-negative, real valued, continuously differentiable function on R^r (whose gradient is denoted by $V_x(\cdot)$) satisfying $V(0) = 0$, $V(x) > 0$, for $|x| \neq 0$. Let the set defined by $Q_\lambda \equiv \{x: V(x) < \lambda\}$ be bounded with the derivative of $V(x_t)$ non-positive along trajectories in Q_λ ; namely

$$(5) \quad \dot{V}(x_t) = V'_x(x_t)f(x_t) \equiv -k(x_t) \leq 0$$

for $x_t \in Q_\lambda$. Let $x_0 = x$ be in Q_λ . The following statements can be made

$V(x_t)$ is non increasing. Then $x_t \in Q_\lambda$ for all $t \geq 0$.

From

$$(6) \quad V(x) - V(x_t) = \int_0^t k(x_s) ds \geq 0$$

we have that $\int_0^\infty k(x_s) ds < \infty$. This, and the uniform continuity of $k(x_s)$ on $[0, \infty)$ imply that $k(x_s) \rightarrow 0$ as $s \rightarrow \infty$, and $x_s \rightarrow \{x: k(x) = 0\} \cap$

$Q_\lambda \equiv K_\lambda$.

Furthermore as $x \rightarrow 0$, the maximum excursions of $|x_t|$ decrease to zero.

Define an invariant set of points G in R^r as follows. Let $x \in G$. Then there is a function x_t , $t \in (-\infty, \infty)$ which satisfies the equation $\dot{x} = f(x)$ with $x_0 = x$, and furthermore $x_t \in G$ for all $t \in (-\infty, \infty)$. Thus G contains entire trajectories over the doubly infinite time interval $(-\infty, \infty)$.

Let the trajectory x_t be bounded. In particular, let $x_0 = x \in Q_\lambda$, and assume (5). Then the invariance theorem [7] states that the path tends to the largest invariant set contained in K_λ .

The theorem is important since it is often used to show that the x_t tend to a much smaller set than K_λ . It gives a very nice characterization of the sets to which x_t can tend.

Example. Define the differential equation on $\mathbb{R}^2$,

$$\begin{aligned}\dot{x}_1 &= x_2 \\ \dot{x}_2 &= -g(x_1) - ax_2\end{aligned}$$

where

$$\begin{aligned}\int_0^t g(s) ds &\rightarrow \infty \text{ as } t \rightarrow \infty, \quad sg(s) > 0 \text{ for } s \neq 0 \\ g(0) &= 0, \quad a > 0.\end{aligned}$$

Define the Liapunov function

$$V(x) = x_2^2 + 2 \int_0^{x_1} g(s) ds.$$

Then

$$V'_x(x)f(x) = -k(x) = -2ax_2^2.$$

We can conclude that $x_{2t} \rightarrow 0$. But what about x_{1t} ?

It is natural to expect that $x_{1t} \rightarrow 0$ also, and indeed (although the Liapunov function argument does not directly yield it) it can be proved using a limiting argument, using the facts that $V(x_t)$ is non-increasing and $x_{2t} \rightarrow 0$. Yet it would be much simpler to merely substitute $x_{2t} \equiv 0$ in the differential equation, and see what trajectories are possible; namely, put the limit of x_{2t} into the equation, directly. The invariance

theorem allows us to do this, and to conclude that $x_{1t} \rightarrow 0$ also. In examples involving functional differential or more complicated systems, the invariance theorems can save an enormous amount of work.

In the sequel, we will develop stochastic counterparts of all the concepts which we just used. While homogeneity is required for the invariance theorem, there are straightforward non-homogeneous extensions of the Liapunov function theorems.

4. Stopped Processes.

The weak infinitesimal operator $\tilde{A}$ and Dynkins formula (4) will be used to replace (5), (6) for the stochastic problem. The domain $\mathcal{D}(\tilde{A})$ was defined to be a subset of a set of bounded functions on X . However, the Liapunov functions $V(\cdot)$ which are most likely to be used, and to which $\tilde{A}$ is to be applied, are usually unbounded (as is usual in the deterministic case). Even if $V(\cdot)$ were bounded, the process may have a stability property only in a bounded or compact set Q in X . I.e., $\tilde{A}f(x)$ may be non-positive only in some neighborhood Q of the origin.

There is no loss of generality in studying the process only while the paths are in such a set Q . For we can often (always, if X is σ -compact) find a sequence of sets $Q_n \uparrow X$, and, if desired, study the behavior of X by studying the "limits" of the behavior of the process up to, say, τ_n , where $\tau_n = \inf\{t: x_t \notin Q_n\}$. Thus, we can bound $V(x)$ for x "sufficiently far" from Q , or we can define a new process by merely stopping x_t on first exit from Q . The latter approach is much more convenient.

Let Q be a set in X . Dynkin ([5], Chapter 4) gives various general conditions under which $\tau = \inf\{x_t \notin Q\}$ is a Markov time. We

mention only the following. Let x_t be right continuous w.p.l.

- (a) Q is open and has compact closure. (Lemma 4.1)
- (b) Q is open, X is a metric space (or metrizable) and $X - Q$ is compact. (Lemma 4.1)
- (c) Q is open, and x_t is continuous. (p. 111)

Define the stopped process $\tilde{x}_t = x_{t \cap \tau}$, where $t \cap \tau = \min(t, \tau)$. Let either (a)-(c) above hold, and let x_t be right continuous w.p.l. Then x_t is a strong Markov process ([5], Theorem 10.2). Unless otherwise mentioned, $\tilde{A}_Q$ will be used to denote the weak infinitesimal operator of the process $\tilde{x}_t$. Let x_t be continuous w.p.l. Then to apply Dynkin's formula to an unbounded function $V(\cdot)$, we only need check that the restriction of $V(\cdot)$ to Q is in $\mathcal{D}(\tilde{A}_Q)$. If x_t is right continuous w.p.l., we need to check whether the restriction of $V(\cdot)$ to the union over $x = x_0$ in Q of the almost sure range $x_s, s \leq \tau$, is in $\mathcal{D}(\tilde{A}_Q)$. Such verification usually seems to be straightforward in examples.

5. Stochastic Stability and Asymptotic Stability.

Unless otherwise mentioned, we will use the following assumptions in this section. After the theorems are proved, extensions to more general cases will be discussed.

- (A1) X is Euclidean r -space
- (A2) $V(\cdot)$ is a non-negative real valued and continuous function on R^r .
- (A3) Define $Q_\lambda = \{x: V(x) < \lambda\}$ and assume that Q_λ is not empty.

Let x_t denote a right continuous homogeneous strong Markov process on the state space X , defined until at least the first time of exit from Q_λ . Write $\tilde{A}_\lambda$ for the weak infinitesimal operator of $\tilde{x}_t$, where $\tau_\lambda = \inf\{t: x_t \notin Q_\lambda\}$, and $\tilde{x}_t = x_{t \cap \tau_\lambda}$.

(A4) $V(\cdot) \in \mathcal{D}(\tilde{A}_\lambda)$ (where the definition of $V(\cdot)$ is assumed restricted to the union over $x_0 = x$ in Q of the almost sure range of $\tilde{x}_t$).

$$(A5) \quad \sup_{x \in Q_\lambda} P_x \left\{ \sup_{t \geq s \geq 0} \|x_s - x\| > \epsilon \right\} \rightarrow 0 \quad \text{as } t \rightarrow 0 \quad \text{for any } \epsilon > 0.$$

Observe that, if $y \notin Q_\lambda$, but is in the almost sure range of x_s , $s \leq \tau_\lambda$ for some $x_0 = x \in Q_\lambda$, then $\tilde{A}_\lambda V(y) \equiv -k(y) = 0$. We will use this fact implicitly in the following theorems.

Theorem 1. Assume (A1)-(A4). Let $\tilde{A}_\lambda V(x) \leq 0$ (recall that the operation $\tilde{A}_\lambda V(x)$ is defined for the stopped process). Then $V(\tilde{x}_t)$ converges w.p.l., as $t \rightarrow \infty$. Hence $V(x_t)$ converges for almost all paths remaining in Q_λ . For $x \in Q_\lambda$,

$$(7) \quad P_x \left\{ \sup_{\infty > t \geq 0} V(x_t) \geq \lambda \right\} = P_x \left\{ \sup_{\infty > t \geq 0} V(\tilde{x}_t) \geq \lambda \right\} \leq V(x) / \lambda.$$

If $V(0) = 0$ and $V(x) \neq 0$ for $x \neq 0$, then as $|x| \rightarrow 0$, the probability in (7) goes to zero (a type of stability of the origin).

Proof. Applying (4) gives

$$(8) \quad E_x V(\tilde{x}_t) - V(x) = E_x \int_0^t \tilde{A}_\lambda V(\tilde{x}_s) ds = E_x \int_0^{t \cap \tau_\lambda} \tilde{A}_\lambda V(x_s) ds \leq 0.$$

Thus, w.p.1.,

$$E_{\tilde{x}_s} V(\tilde{x}_t) \leq V(\tilde{x}_s)$$

or, equivalently, since $\tilde{x}_s$ is Markov, ($\mathcal{D}_s$ is the smallest σ -algebra which measures x_r , $r \leq s$)

$$E[V(\tilde{x}_{t+s}) | \mathcal{D}_s] \leq V(\tilde{x}_s), \text{ w.p.1.}$$

Thus $\{V(\tilde{x}_t), \mathcal{D}_t\}$ is a non-negative super-martingale. This gives the convergence of $V(\tilde{x}_t)$. (7) is the super-martingale probability inequality. The rest of the statements are obvious. Q.E.D.

Non-homogeneous case. Suppose that the Liapunov function $V(\cdot, \cdot)$ depends on x and t or that x_t is non-homogeneous. We state the following Theorem 2, without proof.

Theorem 2. Let the real valued continuous functions (on $R^r, R^r \times [0, \infty), R^r$, resp) $V_1(\cdot), V(\cdot, \cdot), V_2(\cdot)$ satisfy, for some real $t_0 \geq 0$, and $\lambda > 0$,

$$V_1(x) \leq V(x, s) \leq V_2(x)$$

for $s \geq t_0$ and $x \in Q'_\lambda \equiv \{x: V_1(x) < \lambda\}$. Let x_t be a right continuous strong Markov process defined until at least the first exit time τ_λ from Q'_λ . Let $\tilde{A}_\lambda$ denote the weak infinitesimal operator of the process $(\tilde{x}_t, t \cap \tau_\lambda)$, which is (x_t, t) stopped on first exit from Q'_λ . Suppose $V(x, t) \in \mathcal{D}(\tilde{A}_\lambda)$ and $\tilde{A}_\lambda V(x, t) \leq 0$ for $t \geq t_0$. Then, for $t \geq t_0$,

$$(9) \quad P_{x, t_0} \left\{ \sup_{s \geq 0} V_1(x_s) \geq \lambda \right\} \leq P_{x, t_0} \left\{ \sup_{s \geq 0} V(x_s, s+t_0) \geq \lambda \right\} \\ \leq V(x, t_0) / \lambda.$$

Also $V(x_t, t+t_0)$ converges for almost all paths for which $V(x_t, t+t_0) \leq \lambda$ for all $t \geq 0$, where we use $x_0 = x$ for the initial condition at the initial time t_0 ; thus there is convergence with at least probability $1 - V(x, t_0) / \lambda$.

Let $V_2(x) \rightarrow 0$ as $|x| \rightarrow 0$; then the right hand side of (9) goes to zero as $|x| \rightarrow 0$. Let also $V_1(0) = 0$ and $V_1(x) > 0$ for $|x| \neq 0$; then for any $\epsilon > 0$ and any neighborhood of the origin A_1 , there is a neighborhood $A_2 \subset A_1$ so that, if $x \in A_2$, the probability of x_t ever leaving A_1 is no greater than ϵ .

Theorem 3. (Asymptotic Stability). Assume (A1)-(A5). Let $\tilde{A}_\lambda V(x) \leq -k(x) \leq 0$ in Q_λ . Then $k(\tilde{x}_t) \rightarrow 0$ in probability and $V(\tilde{x}_t)$ converges w.p.l. Thus $k(x_t) \xrightarrow{P} 0$ (and also $V(x_t)$ converges) for almost all paths which never leave Q_λ . (Equation (7) gives a lower bound $1-V(x)/\lambda$ to the probability of never leaving Q_λ .)

Let $k(\cdot)$ be uniformly continuous in Q_λ . Then $x_t \rightarrow \bigcap_{\epsilon > 0} \{x: k(x) < \epsilon\} \cap Q_\lambda \equiv P_\lambda$ for almost all paths which never leave Q_λ . If the hypotheses hold for all $\lambda < \infty$ and $V(x) \rightarrow \infty$ as $|x| \rightarrow \infty$, then $x_t \rightarrow \bigcap_{\epsilon > 0} \{x: k(x) < \epsilon\} = K$ w.p.l. The convergence in the last two sentences is in the topology for the compactified R^r , if the $Q_\lambda \cap \{x: k(x) < \epsilon\}$ are unbounded. If they are bounded, replace $\bigcap_{\epsilon} \{k(x) < \epsilon\}$ by $\{x: k(x) = 0\}$.

Proof. The key to the proof is the fact that the total time which the process $\tilde{x}_s$ can spend outside of the set $K_\epsilon = \{x: k(x) \geq \epsilon > 0\} \cap Q_\lambda$ is finite w.p.l. for any $\epsilon > 0$. This follows from the inequality

$$(10) \quad V(x) \geq -E_x V(\tilde{x}_t) + V(x) = E_x \int_0^t k(\tilde{x}_s) ds \geq \epsilon E_x T'(t, \epsilon),$$

where $T'_x(t, \epsilon)$ is the total time that $k(\tilde{x}_s) \geq \epsilon$ in $[0, t]$. That $k(\tilde{x}_t) \xrightarrow{P} 0$ follows from (10).

We next prove the first statement of the second paragraph of the theorem. Let $T(t, \epsilon)$ denote the total time that $k(\tilde{x}_s) \geq \epsilon$ in $[t, \infty)$. Then $T(t, \epsilon) \rightarrow 0$ w.p.1. for any $\epsilon > 0$ and $x \in Q_\lambda$. The rest of the proof combines this fact with (A5), the uniform stochastic stability assumption, to yield the w.p.1. convergence. Let C_ϵ denote the set

$$C_\epsilon = \{x: k(x) < \epsilon\} \cap Q_\lambda.$$

Assume that $k(x) < 0$ for some $x \in Q_\lambda$, for otherwise the theorem is trivial. Then, by uniform continuity of $k(\cdot)$ in Q_λ , there is some ϵ_0 so that the distance between $Q_\lambda \cap C_\epsilon^c$ and $C_{\epsilon/2}^c$ (C_ϵ^c is the complement of C_ϵ) is positive (say $\geq \delta(\epsilon)$) for $0 < \epsilon < \epsilon_0$, and $Q_\lambda \cap C_\epsilon^c$ is not empty.

Define the Markov times σ_n, σ'_n (finite on sets Ω_n, Ω'_n , resp.) as follows. (If σ_n or σ'_n is not defined at ω , set it equal to ∞ there.) $\sigma_0 = 0, \sigma'_0 = \inf\{t: \tilde{x}_t \in C_{\epsilon/2}\}, \sigma_1 = \inf\{t: \tilde{x}_t \in Q_\lambda \cap C_\epsilon^c, t \geq \sigma'_0\}, \sigma'_1 = \inf\{t: \tilde{x}_t \in C_{\epsilon/2}, t \geq \sigma_{n-1}\}, \sigma_n = \inf\{t: \tilde{x}_t \in Q_\lambda \cap C_\epsilon^c, t \geq \sigma'_n\}$, etc. There is some $\rho > 0$ so that

$$(11) \quad \sup_{x \in Q_\lambda} P_x \left\{ \sup_{\rho \geq s \geq 0} |x_s - x| \leq \delta(\epsilon)/2 \right\} \geq 1/2.$$

Define

$$A_n = \{\omega: \tilde{x}_{\sigma_n+s} \in C_{\epsilon/2}^c \cap Q_\lambda, 0 \leq s \leq \rho, \sigma_n < \infty\}.$$

If $\omega \in A_n$ infinitely often, then the total time out of $C_{\epsilon/2}^c \cup Q_\lambda^c$ is infinite for the corresponding path $\tilde{x}_t(\omega)$. Then $\omega \in A_n$ only finitely often w.p.1. But $\sum I_{A_n} \rightarrow \infty$ w.p.1. if and only if $(\mathcal{B}_{\sigma_n}$ measures $\tilde{x}_s, s \leq \sigma_n$, thus all $A_i, i = 0, \dots, n-1$, are in $\mathcal{B}_{\sigma_n}$)

$$\sum P_x\{A_n | \mathcal{B}_{\sigma_n}\} \rightarrow \infty \text{ w.p.1.}$$

([5], p. 398-399) and, by the strong Markov property ($P_{x_t}\{A\} = 0$ for $t = \infty$)

$$\begin{aligned} (12) \quad \sum P_x\{A_n | \mathcal{B}_{\sigma_n}\} &\geq \sum P_{x_{\sigma_n}} \left\{ \sup_{\rho \geq s \geq 0} |\tilde{x}_s - x| > \frac{\delta(\epsilon)}{2} \right\} I_{\{\sigma_n < \infty\}} \\ &\geq \frac{1}{2} \sum I_{\{\sigma_n < \infty\}}. \end{aligned}$$

Thus $\sigma_n < \infty$ only finitely often w.p.1. The remaining statements of the theorem follow easily from what we have already proved. Q.E.D.

Discussion and extensions.

(1) It is not necessary that $k(\cdot)$ be continuous in Q_λ , nor even that $k(x) = 0$ anywhere. See the hypothesis and proof of [1], Theorem 2, Chapter 2. There are examples arising in control theory where $X = \mathbb{R}^r$ has a hole in it, i.e., a target set S_1 is deleted, and $k(x) \geq 1$ for $x \notin S_1$. The set S_2 is absorbing, so there is a discontinuity of $k(\cdot)$ on the boundary of X .

(2) If the hypotheses of Theorem 3 hold for all λ and $V(x) \rightarrow \infty$ as $|x| \rightarrow \infty$, then $x_t \rightarrow \{x: k(x) = 0\}$ w.p.1.

(3) If $\tilde{A}_\lambda V(x) \leq -k(x) \leq 0$ in Q_λ , then $k(\tilde{x}_t) \xrightarrow{P} 0$ for any state space, provided that the Dynkins formula is valid for the $\tilde{x}_t$ process.

(4) If Q_λ is unbounded, (A5) may not hold. Furthermore $\{x: k(x) < \epsilon\} \cap Q_\lambda$ may be unbounded. Suppose that (A5) holds if Q_λ is replaced by any compact subset S of X . Then, we can obtain the following. For any compact S and $\epsilon > 0$, there is a random variable $\tau_{S,\epsilon} < \infty$ w.p.l., so that $x_t \notin S - C_\epsilon$ for $t \geq \tau_{S,\epsilon}$. Thus $x_t \rightarrow \{x: k(x) = 0\} \cup \{\infty\}$ w.p.l. in the one-point-compactification topology of R^r .

Sometimes subsidiary conditions can be used to eliminate the point $\{\infty\}$. Refer to the next section for the definition of the terms "weakly bounded" and "invariant set". Let the measures of the process $\tilde{x}_t$ be weakly bounded, and let $k(\tilde{x}_t) \xrightarrow{P} 0$. Then x_t tends in probability to the support of the largest invariant set whose support is contained in $[\{x: k(x) = 0\} \cap Q_\lambda] \cup Q_\lambda^c$. Thus, $\tilde{x}_t$ tends in probability to the union of Q_λ^c and a subset of $\{x: k(x) = 0\} \cap Q_\lambda$.

The remarks and results for unbounded Q_λ are motivated by the stability problem for a process of the type $\dot{y} = f(u, y)$, where u_t , and the pair $(u_t, y_t) = x_t$ are Markov processes. The process u_t may serve as a time varying parameter, and not converge in any sense. We may be concerned with the convergence of the component y_t only, but the Liapunov function may depend on both components.

(5) If X is a metric space, the proof still goes through under (A2) - (A5), if we replace R^r in (A2) by a metric space X . It may be difficult to verify (A5) and the uniform continuity of $k(\cdot)$ in this case, and the closure of Q_λ will not usually be bounded. But it sometimes happens that if $x = x_0 \in Q_\lambda$, then the path $\tilde{x}_s$, $\tau_\lambda > s \geq 0$, is contained in a bounded subset of Q_λ w.p.l. Then Q_λ is "effectively" contained in a bounded subset, and if $k(\cdot)$ is uniformly continuous and (A5) holds on this subset, then the proof goes through. See [6] for a specific example. (A5) plays a crucial role in the proof (since we need to guarantee that $\tilde{x}_t$ does not jump (w.p.l.) from $C_{\epsilon/2}^c$ to C_ϵ^c and back to $C_{\epsilon/2}^c$ infinitely often in a total integrated time which is finite), and some form of uniform stochastic continuity condition is probably essential.

II. Invariant Set Theorems and Applications to Stochastic Dynamical Systems.

In this Section we will develop a stochastic theory of invariance analogous to the deterministic theory in [7], [8]. The main conclusion is that, under given conditions, the measures of the process x_t tend to an invariant set of measures, and that x_t tends to the closure of the support set of this set of measures in probability as $t \rightarrow \infty$.

Note that we are using the terms "invariance" and "invariant" according to their usage in the general theory of dynamical systems. The term has nothing to do with the stochastic notion of invariant measure. In this Section x_t will be a homogeneous strong Markov process. We essentially follow the development in [2], with some changes and corrections.

1. Definitions. Let X , the state space of the process x_t , be a separable metric space. Let φ denote the initial measure of the process; i.e., $P\{x_0 \in A\} = \varphi(A)$. Let $m(t, \varphi, \cdot)$ denote the measure induced on the Borel sets of X by the process at time t , with initial measure φ . The semigroup property⁺

$$m(t+s, \varphi, \cdot) = m(t, m(s, \varphi), \cdot)$$

holds.

Let $\mathcal{M}$ denote the space of probability measures on X . A sequence $\{\psi_n\}$ in $\mathcal{M}$ is said to converge weakly to ψ if $\int f(x) \psi_n(dx) \rightarrow \int f(x) \psi(dx)$ for every $f(\cdot)$ in C_X , the space of continuous bounded functions on X . We may abbreviate the convergence relation as $f[\psi_n] \rightarrow f[\psi]$. A set $M = \{\psi_\alpha\}$ in $\mathcal{M}$ is weakly bounded if, for each $\epsilon > 0$, there is a compact set $K_\epsilon \subset X$

⁺Occasionally for simplicity φ is written for $\varphi(\cdot)$ and $m(s, \varphi)$ for $m(s, \varphi, \cdot)$ or $m(s, \varphi(\cdot), \cdot)$.

for which $\psi_\alpha(X - K_\epsilon) \leq \epsilon$ for all α . Define an ω -limit set⁺ as a set $W(\varphi)$ in $\mathcal{M}$ with the property: $\psi \in W(\varphi)$ if there is a sequence $t_n \rightarrow \infty$ so that $f[m(t_n, \varphi)] \rightarrow f[\psi]$ (ψ is a weak limit of a sequence of measures taken along the trajectory) for all $f(\cdot) \in C_X$. A set $M \subset \mathcal{M}$ is an invariant set if for each $\psi \in M$, there is a sequence of measures $m'(t, \cdot)$, for $t \in (-\infty, \infty)$ where $m'(0, \cdot) = \psi(\cdot)$, the initial measure, and $m(t, m'(s, \varphi), \cdot) = m'(t+s, \cdot)$ for any $t \geq 0$ and $s \in (-\infty, \infty)$. Thus for each $\psi \in M$, there is a trajectory of measures defined for all $t \in (-\infty, \infty)$ and satisfying the law of motion of the process x_t and initial condition ψ . Let ψ be in $\mathcal{M}$. $x \in X$ is in the support set $S(\psi)$ of ψ if $\psi(N) > 0$ for each neighborhood of N of x . Similarly $S(Q) = \bigcup_{\psi \in Q} S(\psi)$ is the support set of a set Q in $\mathcal{M}$. The set $S(\psi)$ is closed, but $S(Q)$ is not necessarily closed. The process x_t is a Feller process if $E_x f(x_t)$ is continuous in x for $t > 0$ and $f(\cdot) \in C_X$.

Next, the main theorem and a useful corollary will be given. Then the conditions of the theorem will be replaced by more easily verifiable conditions.

2. The Invariance Theorem.

Theorem 4. Assume (B1) - (B3).

(B1) The trajectory $\{m(t, \varphi), t \geq 0\}$ is weakly bounded.

(B2) For each $f(\cdot)$ in C_X , $f[m(t, \varphi)]$ is continuous in t on any finite t ($t \geq 0$) interval, uniformly in φ , for φ in any weakly bounded set.

(B3) $f[m(t, \varphi)]$ is weakly continuous in φ for each fixed $t \geq 0$.
[I.e., as $\varphi_n \xrightarrow{W} \varphi$, $f[m(t, \varphi_n)] \rightarrow f[m(t, \varphi)]$ for each $f(\cdot) \in C_X$ and each $t \geq 0$.]

⁺It is important to keep in mind that the ω -limit set is an ω -limit set of a trajectory of measures.

Then $W(\varphi)$ is a non-empty, weakly bounded, weakly compact invariant set and there is a sequence $\tilde{\psi}(t)$ in $W(\varphi)$, $t \geq 0$, so that

$$f[m(t, \varphi)] - f[\tilde{\psi}(t)] \rightarrow 0$$

for all $f(\cdot) \in C_X$, as $t \rightarrow \infty$.

Proof. According to Theorem 1, Section 1, Chapter 9 of [10], a sufficient condition for a sequence in $\mathcal{M}$ to have a weakly convergent subsequence is that it be weakly bounded. Thus $W(\varphi)$ is not empty.

Let $\{\epsilon_i\}$ denote a real sequence which tends to zero. By (A1), there are compact sets G_i so that $G_{i+1} \supset G_i$ and $m(t, \varphi, X - G_i) \leq \epsilon_i$, all $t \geq 0$. For each G_i , there is a countable family $\mathcal{F}'_i$ of continuous functions, defined on G_i , and dense in C_{G_i} . Each element of $\mathcal{F}'_i$ can be extended to a continuous function X without increasing its norm (using the normality of the metric space and [9], Theorem 1.5.3). Let $\mathcal{F}_i$ denote the countable family of such extensions and $\mathcal{F} = \bigcup_i \mathcal{F}_i$. Write $G = \bigcup_i G_i$. Observe that, for any $f(\cdot) \in C_X$,

$$(*) \quad \int_G f(x) m(t, \varphi, dx) = \int f(x) m(t, \varphi, dx).$$

(*) also holds for $m(t, \varphi, \cdot)$ replaced by an element in the weak closure of $\{m(t, \varphi, \cdot)\}$.

Let $m(t_n, \varphi, \cdot)$ converge weakly to $\psi(\cdot)$ in the ω -limit set $W(\varphi)$. Define the function $F_n(\cdot, \cdot)$ by

$$F_n(t, f) = \int f(x) m(t_n + t, \varphi, dx).$$

If $t_n - T \geq 0$, then $F_n(t, f) = \int f(x) m(t + T, m(t_n - T, \varphi), dx)$. Since $\{m(t_n - T, \varphi)\}$ is weakly bounded, (B2) implies that $F_n(t, f)$ is continuous in t on $[-T, T]$, uniformly in n , for each $f(\cdot)$. Thus Ascoli's Theorem implies that there is a uniformly convergent subsequence on $[-T, T]$. By successive applications of the diagonal procedure, we can extract a subsequence (of t_n) for which $F_n(t, f)$ converges to a continuous function of t , $F(t, f)$ for each $f(\cdot) \in \mathcal{F}$, and uniformly on any compact $[-T, T]$ interval. Since, for any $f(\cdot) \in C_X$ and $\epsilon > 0$, there is an $f_\epsilon(\cdot)$ in $\mathcal{F}$ for which $|F_n(t, f_\epsilon) - F_n(t, f)| < \epsilon$ for all n and $t \geq -t_n$, the asserted convergence is for all $f \in C_X$.

Define the set function $\psi(t, \cdot)$ by

$$\psi(t, A) = \inf_{f \geq I_A} F(t, f)$$

where $f(\cdot) \in C_X$, and I_A is the indicator function of the Borel set A in X .

The argument in [10], pp. 441-444, can be used to prove that, for each $t \in (-\infty, \infty)$, $\psi(t, \cdot)$ is a unique probability measure, $\psi(t, G) = 1$ and

$$F(t, f) = \int f(x) \psi(t, dx)$$

for each $f(\cdot) \in C_X$. Thus $m(t_n + t, \varphi, \cdot) \xrightarrow{w} \psi(t, \cdot)$ for each $t \in (-\infty, \infty)$, where $\psi(0, \cdot) = \psi(\cdot)$. The weak closure of $\{m(t, \varphi, \cdot)\}$ is also weakly bounded and is supported in G . Thus, by (B3), we can write, for any $t \in (-\infty, \infty)$, $s \geq 0$, and $f(\cdot)$ in C_X ,

$$\begin{aligned}
f[m(s, m(t_n + t, \varphi))] &\rightarrow f[m(s, \psi(t))] \\
f[m(s, m(t_n + t, \varphi))] &= f[m(0, m(t_n + t + s, \varphi))] \\
&\rightarrow f[m(0, \psi(t + s))] = f[\psi(t + s)],
\end{aligned}$$

which implies that $\psi(t + s, \cdot) = m(s, \psi(t), \cdot)$ since the continuous functions determine the measures uniquely. Thus $\{\psi(t)\}$ obeys the law of the process and each $\psi(t)$ is in an invariant set.

Let $f[\psi_n(\cdot)]$ converge for each $f(\cdot) \in C_X$, as $n \rightarrow \infty$, where $\psi_n(\cdot) \in W(\varphi)$ (thus $\{\psi_n\}$ are weakly bounded). There is a measure $\psi(\cdot)$ for which $f[\psi_n(\cdot)] \rightarrow f[\psi(\cdot)]$ on C_X , and $\psi(G) = 1$. We need to show that $\psi(\cdot) \in W(\varphi)$. For each n , $m(t_i(n), \varphi, \cdot) \xrightarrow{W} \psi_n(\cdot)$ as $i \rightarrow \infty$, for some real sequence $t_i(n) \rightarrow \infty$. Since

$$\begin{aligned}
\lim_n \lim_i \int f(x) m(t_i(n), \varphi, dx) &= \lim_n \int f(x) \psi_n(dx) \\
&= \int f(x) \psi(dx),
\end{aligned}$$

for each $f \in \mathcal{F}$, we can extract a subsequence $\{t_\alpha\}$ of the double sequence $\{t_i(n)\}$ for which $m(t_\alpha, \varphi, \cdot) \xrightarrow{W} \psi(\cdot)$.

Only the last assertion of the theorem remains to be proved. Suppose that there is a sequence $\{t_n\}$ so that for any subsequence $\{t'_n\}$, and some $f(\cdot)$ in $\mathcal{F}$ or C_X ,

$$(*) \quad \limsup_n \inf_{\psi \in W(\varphi)} |f[m(t'_n, \varphi)] - f[\psi(\cdot)]| > 0.$$

By weak boundedness of $\{m(t'_n, \varphi, \cdot)\}$, there is a subsequence which converges

weakly to some $\psi(\cdot) \in \mathcal{M}$. This $\psi(\cdot)$ must also be in $W(\varphi)$, a contradiction to (*). Q.E.D.

Theorem 5. Assume (B1)-(B3) of Theorem 4. Then

(i) $x_t \xrightarrow{P} \bar{S}(W(\varphi)) \equiv C$, the closure of the support set of the invariant set $W(\varphi)$, I.e., $P_\varphi\{\inf_{y \in C} |x_t - y| > \epsilon\} \rightarrow 0$ as $t \rightarrow \infty$, for any $\epsilon > 0$.

(ii) Let $k(\cdot)$ be a real valued, non-negative and continuous function on X and let $k(x_t) \xrightarrow{P} 0$. Let G_n denote compact sets in X for which $m(t, \varphi, X - G_n) < \epsilon_n \rightarrow 0$, $G_{n+1} \supset G_n$. Then x_t converges in probability to the largest support set of an invariant set whose support is contained in $\lim_n G_n \cap \{x: k(x) = 0\}$.

Proof. (i) Let $N_\epsilon(C)$ denote an ϵ -neighborhood of C . We will show that, for each $\epsilon > 0$,

$$(*) \quad \lim_{t \rightarrow \infty} P_\varphi\{x_t \in X - \bar{N}_\epsilon(C)\} = 0,$$

since (*) implies (i). Suppose (*) is violated. Then there are $t_n \rightarrow \infty$ and $\epsilon_0 > 0$ so that $P_\varphi\{x_{t_n} \in X - \bar{N}_\epsilon(C)\} \geq \epsilon_0 > 0$. There is a function $f(\cdot) \in C_X$ satisfying $0 \leq f(x) \leq 1$, $f(x) = 0$ on $\bar{N}_{\epsilon/2}(C)$, $f(x) = 1$ on $X - N_\epsilon(C)$. For some subsequence $\{t'_n\}$ of $\{t_n\}$, $m(t'_n, \varphi, \cdot)$ converges weakly to a $\psi(\cdot)$ in $W(\varphi)$ and $\int [m(t'_n, \varphi)] \rightarrow \int [\psi(\cdot)] \geq \epsilon_0 > 0$. Thus $X - N_\epsilon(C)$, which is disjoint from $\bar{C}$, contains some point in the support set of $\psi(\cdot)$, a contradiction to the definition of C .

(ii) follows easily from (i), and the proof is omitted. Q.E.D.

Discussion of the Conditions (B1)-(B3) of Theorem 4.

Under the conditions of Theorem 1, if $\bar{Q}_\lambda$ is compact, then the measures for the stopped process are weakly bounded, and we can apply the invariance theorem to the stopped process. If the conditions of Theorem 1 hold for all $\lambda < \infty$, and each $\bar{Q}_\lambda$ is compact, then $\{m(t, \varphi, \cdot)\}$ is weakly bounded. Usually, the function $k(\cdot)$ in Theorem (5) is the $k(\cdot)$ of Theorem 3. Furthermore, even if each Q_λ is not bounded, it may be that the measures for the process stopped on exit from Q_λ are weakly bounded. See Example 2 in [6].

Theorem 6. (B3) holds for a Feller process on any topological state space.

Proof. Let $\varphi_n(\cdot) \xrightarrow{w} \varphi(\cdot)$. We must show that

$$(*) \quad \int f(x) m(t, \varphi_n, dx) - \int f(x) m(t, \varphi, dx) \rightarrow 0$$

for all $f(\cdot) \in C_X$. Write (*) as

$$\begin{aligned} & \int [f(y) m(t, x, dy)] (\varphi_n(dx) - \varphi(dx)) \\ &= \int h_t(x) [\varphi_n(dx) - \varphi(dx)] \rightarrow 0. \end{aligned}$$

$m(t, x, \cdot)$ denotes the measure with initial condition x and $h_t(x) = E_x f(x_t)$ which is in C_X by the Feller property, and the convergence follows since $\varphi_n(\cdot) \xrightarrow{w} \varphi(\cdot)$. Q.E.D.

Remark. Theorem 6 implies that condition (B3) is not very restrictive.

Theorem 7. Let

$$(*) \quad P_x\{|x_t - x| > \epsilon\} \rightarrow 0$$

as $t \rightarrow 0$, uniformly for x in any compact set. For each real $T > 0$ and compactum $K \subset X$, let the family $\{m(t, x, \cdot), x \in K, t \leq T\}$ be weakly bounded. Then (B2) holds.

Proof. Let $\{\varphi_\alpha\}$ denote a weakly bounded set of measures. Then the second hypothesis implies that the family $\{m(t, \varphi_\alpha, \cdot), t \leq T, \text{ all } \alpha\}$ is weakly bounded (we omit the proof, which is not hard).

Write, for $t \geq 0, s \geq 0, s+t \leq T$,

$$\begin{aligned} & | \int f(x) [m(t+s, \varphi_\alpha, dx) - m(t, \varphi_\alpha, dx)] | \\ & \leq \int |E_x f(x_s) - f(x)| m(t, \varphi_\alpha, dx) \\ & = \int_{G'} |E_x f(x_s) - f(x)| m(t, \varphi_\alpha, dx) + \int_{X-G'} |E_x f(x_s) - f(x)| m(t, \varphi_\alpha, dx). \end{aligned}$$

Choose compact G' to make the second term less than $\frac{\epsilon}{2}$, for all α , $t \leq T$. Then, using the first hypothesis, choose $s_0 > 0$ so that $|E_x f(x_s) - f(x)| \leq \frac{\epsilon}{2}$ for $s \leq s_0$ and $x \in G'$, thus proving right continuity of $E_x f(x_t)$.

To prove left continuity, write, for $T \geq t-s \geq 0, s \geq 0$,

$$\begin{aligned}
& \left| \int f(x) [m(t, \varphi_\alpha, dx) - m(t-s, \varphi_\alpha, dx)] \right| \\
& \leq \int_{G'} |E_x f(x_s) - f(x)| m(t-s, \varphi_\alpha, dx) + \int_{X-G'} |E_x f(x_s) - f(x)| m(t-s, \varphi_\alpha, dx).
\end{aligned}$$

Choose compact G' so that the second term is $\leq \frac{\epsilon}{2}$ for $0 \leq t-s \leq T$, and all α , and then choose s_0 so that $|E_x f(x_s) - f(x)| \leq \frac{\epsilon}{2}$ for $s \leq s_0$ and all $x \in G'$. Q.E.D.

III. Examples.

Example 1. A relatively simple example is the diffusion process given by the Itô equation

$$\begin{aligned}
dx_1 &= x_2 dt \\
dx_2 &= -g(x_1) dt - ax_2 dt - x_2 cdz
\end{aligned}$$

where

$$\int_0^t g(s) ds \rightarrow \infty \text{ as } t \rightarrow \infty, \quad sg(s) > 0, \quad s \neq 0, \quad g(0) = 0,$$

and $g(\cdot)$ satisfies a local Lipschitz condition. Let Q be a bounded open set in R^2 . Then x_t can be defined up until the first exit time from Q , and the stopped process is a continuous Feller process and (B1)-(B3) hold. The function

$$V(x) = x_2^2 + 2 \int_0^{x_1} g(s) ds$$

is in $\mathcal{D}(\tilde{A}_Q)$ and, for $x \in Q$,

$$\tilde{A}_Q V(x) = x_2^2 (c^2 - 2a).$$

Let $c^2 \leq 2a$. Then

$$P_x\left\{\sup_{\infty > t \geq 0} V(x_t) \geq \lambda\right\} \leq V(x)/\lambda \rightarrow 0 \text{ as } \lambda \rightarrow \infty,$$

and x_t can be uniquely defined on $[0, \infty)$ w.p.l., even though $g(\cdot)$ does not satisfy a global Lipschitz condition. It is a continuous Feller process and (B1)-(B3) hold.

Let $c^2 < 2a$. Then $x_{2t} \rightarrow 0$ w.p.l. and by Theorem 1, x_t tends in probability to the smallest invariant set whose support satisfies $x_{2t} = 0$, for all t . Thus $x_t \xrightarrow{P} 0$. This and the w.p.l. convergence of $V(x_t)$ implies that $x_t \rightarrow 0$ w.p.l.

Example 2. For the second example, we take a problem arising in the identification of the coefficients of a linear differential equation.

The system to be identified is the scalar input, scalar output asymptotically stable, reduced form, system

$$(1) \quad \left(\frac{d}{dt} + \sum_{i=0}^{n-1} a_i \frac{d^i}{dt^i}\right)y = \left(\sum_{i=0}^m b_i \frac{d^i}{dt^i}\right)u, \quad n > m,$$

where $u(t)$ is the input. We wish to know the a_i, b_i . The input $u(t)$ is $\sum_i c_i \bar{u}_i(t)$, where $\bar{u}(t)$ is a stationary Markov process. The "equation error" method of P. M. Lion ("Rapid Identification of Linear and Nonlinear Systems", Proc. 1966 Joint Automatic Control Systems Conference, University of Washington, Seattle) will be used. For this method, some estimate of the derivatives of a smoothed input and output are needed.

Let $H(s)$ denote a transfer function the degree of whose denominator exceeds the degree of the numerator by at least n . For any

real number c , define the "derivatives of the smoothed u, y " as

$$y_k(s) = H(s)(s+c)^k y(s), \quad k = 0, \dots, n$$

$$u_k(s) = H(s)(s+c)^k u(s), \quad k = 0, \dots, n$$

and the equation error $\epsilon(t)$ as

$$(2) \quad \epsilon(t) = y_n(t) + \sum_0^{n-1} \alpha_i y_i(t) + \sum_0^m \beta_i u_i(t)$$

where $\{\alpha_i, \beta_i\}$ are to be prescribed. Let the system $(y, y^{(1)}, \dots, y^{(n-1)})$ be state variablized by the minimal order, (with asymptotically stable A^y) $\dot{x}^y = A^y x^y + B^y u$, $y = H^y x^y$, and write

$$y(s) = \frac{N(s)}{D(s)} u(s) + \sum_0^{n-1} \frac{Q_i(s)}{D(s)} x_i(0)$$

where the last term goes to zero exponentially.

Let us impose the following conditions:

(C1) $\bar{u}(t)$ is a right continuous stationary Feller strong Markov process with $E|\bar{u}(t)|^2 = M_0 < \infty$. Thus, the paths are Laplace transformable paths w.p.l. In particular, $\int_0^\infty e^{-kt} |\bar{u}(t)| < \infty$ w.p.l. for all $k > 0$.

(C2) $P_{\bar{u}}\{\sup_{\delta > h > 0} |\bar{u}(h) - \bar{u}| > \epsilon\} \rightarrow 0$ as $\delta \rightarrow 0$ uniformly for the initial condition $\bar{u} = \bar{u}(0)$ in any compact region.

(C3) $E[\bar{u}(t+\tau)\bar{u}'(t)|\bar{u}(s), s \leq 0] \rightarrow \bar{R}(\tau)$, the covariance of the $\bar{u}(t)$ processes. Let $E\bar{u}(t) = 0$. (This condition is not essential.)

(C4) $S_u(\omega)$, the spectral density of $u(t)$, is nonzero over some interval.

There are real numbers $\{\alpha_i, \beta_i\}$ so that $\epsilon(t) \equiv 0$ if all $x_i(0) = 0$. To see this write the Laplace transform of (2) where we have

$$\begin{aligned} \epsilon(s) &= y_n(s) + \sum_0^{n-1} \alpha_i^0 y_i(s) + \sum_0^m \beta_i^0 u_i(s) \\ &= H(s) \left[(s+c)^n \frac{N(s)}{D(s)} + \sum_0^{n-1} \alpha_i^0 (s+c)^i \frac{N(s)}{D(s)} + \sum_0^m \beta_i^0 (s+c)^i \right] u(s) = 0 \end{aligned}$$

if

$$\frac{N(s)}{D(s)} = - \frac{\sum_0^m \beta_i^0 (s+c)^i}{\sum_0^n \alpha_i^0 (s+c)^i}, \quad \alpha_n^0 = \alpha_n = 1.$$

For $\{\alpha_i^0, \beta_i^0\}$ used in (2), $\epsilon(t) \rightarrow 0$ exponentially. In fact, we suppose that the systems generating $y_i(t)$, $u_i(t)$ are connected to their inputs at $t = 0$, and that their initial conditions do not depend on the process $\bar{u}(t)$. Then $\epsilon(t)$ is non-random. The condition can be relaxed to allow for random $y_i(0)$, $u_i(0)$, at some extra complication in the analysis.

The parameter adjustment procedure is

$$\begin{aligned} \dot{\alpha}_j &= - \frac{k}{2} \frac{\partial \epsilon^2}{\partial \alpha_j} = -k \epsilon y_j \\ (3) \quad \dot{\beta}_j &= - \frac{k}{2} \frac{\partial \epsilon^2}{\partial \beta_j} = -k \epsilon u_j. \end{aligned}$$

Define the column vectors

$$z = (\alpha_0 - \alpha_0^0, \dots, \alpha_{n-1} - \alpha_{n-1}^0, \dots, \beta_m - \beta_m^0)$$

$$w = (y_0, \dots, y_{n-1}, u_0, \dots, u_m).$$

Then

$$(4) \quad \dot{z} = -kw\epsilon = -kw\{[y_0\alpha_0^0 + \dots + y_{n-1}\alpha_{n-1}^0 + \dots + u_m\beta_m^0] \\ + y_n + [y_0\alpha_0^0 + \dots + u_m\beta_m^0] - [y_0\alpha_0^0 + \dots + u_m\beta_m^0]\}$$

$$(5) \quad = -kww'z + \tilde{\delta}_t$$

where $\tilde{\delta}_t = -kw[y_n + y_0\alpha_0^0 + \dots + u_m\beta_m^0]$.

We can assume that the $y_k(t)$, $u_k(t)$ are the outputs for asymptotically stable systems of the form $\dot{x}^{y_k} = A^{y_k}x^{y_k} + B^{y_k}y$, $y_k(t) = H^{y_k}x^{y_k}$, etc. Thus all y_k , u_k , y , u , z are state variabilized, and the composite state variabilization, namely $x(t)$, is a right continuous strong Markov process and Feller.

Furthermore, it is uniformly stochastically continuous in the sense of (C2).⁺

Let $E|z(0)|^2 < \infty$. Let $\Phi(t,s)$ denote the fundamental matrix solution of $\dot{z} = -kww'z$. Then $|\Phi(t,s)| \leq 1$ and

$$|z(t)| \leq |z(0)| + \int_0^t |\delta_s| ds$$

⁺E.g., let $\dot{x} = Ax + By$, where y satisfies (C2). Then we only need

$$P_{x,y}\left\{\sup_{\delta \geq h > 0} \left|\int_0^h e^{A(h-\tau)} B y(\tau) d\tau\right| > \epsilon\right\} \rightarrow 0$$

uniformly in (x,y) in compact intervals, as $\delta \rightarrow 0$.

which, together with the bound $E|w_s||w_t| \leq M < \infty$ for some M , yields that $E|z(t)|^2 \leq M_1$ for some $M_1 < \infty$.

Next, let us introduce the Liapunov technique. Let

$$V = z'z.$$

Then

$$\begin{aligned} (6) \quad \dot{V} &= -2k(z'w)^2 - k(z'w)[y_n + w'\alpha^0] \\ &\equiv -2k(z'w)^2 + \rho_t, \end{aligned}$$

where $\alpha^0 = (\alpha_0^0, \dots, \alpha_{n-1}^0, \beta_0^0, \dots, \beta_m^0)$. Note that, since $V(z(t))$ is differentiable, $[E_{x_t} V(z(t+\Delta)) - V(z(t))]/\Delta \rightarrow \dot{V}(z(t))$ w.p.l. V does not involve the possibly non-differentiable components (namely $\bar{u}(t)$) of $x(t)$.

(a) $E|z'(t)w'(t)| \leq E^{1/2}|z(t)|^2 E^{1/2}|w(t)|^2$ is uniformly bounded and $y_n(t) + w(t)\alpha^0$ is not random. Hence $\int_0^\infty |\rho_t| dt < \infty$ w.p.l. and $\int_0^\infty E|\rho_t| dt < \infty$.

(b) The $z(t)$ process is weakly bounded, since $E|z(t)|^2 \leq M_1 < \infty$; hence the $\{x(t)\}$ process is weakly bounded.

(c) From (6) and (a), $z'(t)w(t)$ is square integrable on $\Omega \times [0, \infty)$ and $\int_t^\infty (z'(\tau)w(\tau))^2 d\tau \rightarrow 0$ w.p.l. as $t \rightarrow \infty$. The components of $z(t)$ and $w(t)$ satisfy the uniform stochastic continuity condition.

This, together with the weak boundedness and the convergence

$\int_t^\infty (z'(\tau)w(\tau))^2 d\tau \rightarrow 0$ w.p.l. imply that $z'(t)w(t) \xrightarrow{P} 0$. (Indeed $z'(t)w(t)$ can be shown to converge to zero w.p.l.)

(d) Writing

$$\epsilon = [y_n + \sum_{i=0}^{n-1} \alpha_i^0 y_i + \sum_{i=0}^m \beta_i^0 u_i] = z'w + y_n - y_n$$

we have that $\epsilon(t) \xrightarrow{P} 0$, and is square integrable.

(e) The measures of $x(t)$ tend to an invariant set of measures. This invariant set must be consistent with $\epsilon(t) = 0$ w.p.1. for each t . Thus, by (4), for the invariant set, $z(t) = \tilde{z}$, a random variable.

For the measures in the invariant set

$$(7) \quad 0 = y_n(t) + \sum_{i=0}^{n-1} \tilde{\alpha}_i y_i(t) + \sum_{i=0}^m \tilde{\beta}_i u_i(t) \equiv \tilde{\epsilon}(t)$$

for some set of random variables $\tilde{z} = (\tilde{\alpha}_0, \dots, \tilde{\alpha}_{n-1}, \tilde{\beta}_0, \dots, \tilde{\beta}_m) = (\alpha_0^0, \dots, \beta_m^0)$.

Write

$$0 = E[\tilde{\epsilon}(t)\tilde{\epsilon}(t+\tau) | \tilde{\alpha}_i, \tilde{\beta}_i, \text{ for all } i] = R_t(\tau).$$

Let $y(t)$ denote the components of $x(t)$ without $z(t)$. The probability law of the process $x(t)$ implies that $P_{y,z}\{y(t) \in A\} = P_y\{y(t) \in A\}$. This and (C3) imply that the limit $R(\tau)$ of $R_t(\tau)$, as $t \rightarrow \infty$, is the same as if the $\tilde{\alpha}_i, \tilde{\beta}_i$ were not random. Then, using the stationarity of the $\bar{u}(t)$, $y(t)$, $y_k(t)$, $u_k(t)$ processes, the Fourier transform of $R(\tau)$ is

$$(8) \quad 0 = \left| \sum_{j=0}^n \tilde{\alpha}_j (i\omega + c)^j \frac{N(i\omega)}{D(i\omega)} + \sum_{j=0}^m \tilde{\beta}_j (i\omega + c)^j \right|^2 |H(i\omega)|^2 S_u(\omega).$$

[Consider $R(\tau)$ as the covariance of the output of an asymptotically stable linear system with input $u(t)$ - the transfer function of which is

$$\sum_0^n \tilde{\alpha}_j(s+c)^j H(s) \frac{N(s)}{D(s)} + \sum_0^m \tilde{\beta}_j(s+c)^j H(s).]$$

But (8) and (C4) imply that $\tilde{\alpha}_j = \alpha_j^0$, $\tilde{\beta}_j = \beta_j^0$, and the demonstration is complete.

Bibliography

- [1] H. J. Kushner, Stochastic Stability and Control, Academic Press, New York, 1967.
- [2] H. J. Kushner, "The concept of invariant set for stochastic dynamical systems and applications to stochastic stability", in Stochastic Optimization and Control, H. Karreman, Editor, John Wiley and Sons, New York, 1968.
- [3] H. J. Kushner, Introduction to Stochastic Control Theory, Holt, Rinehart and Winston, New York, 1971.
- [4] M. Loeve, Probability Theory, 3rd Edition, Van Nostrand, Princeton, 1963.
- [5] E. G. Dynkin, Markov Processes, Springer-Verlag, Berlin, 1965.
- [6] H. J. Kushner, "On the stability of processes defined by stochastic difference-differential equations", Journal of Differential Equations, Vol. 4, No. 3, July 1968, pp. 424-443.
- [7] J. P. LaSalle, "The extent of asymptotic stability", Proc. of the Nat. Acad. of Sciences, 46, (1960), p. 365.
- [8] J. K. Hale, "Sufficient conditions for stability and instability of autonomous functional-differential equations", J. Diff. Eqns., 1, 1965.
- [9] N. Dunford, J. T. Schwartz, Linear Operators, Part I, Interscience; Wiley, New York, 1958.
- [10] I. I. Gikhman, A. V. Skorokhod, Introduction to the Theory of Random Processes, Saunders, Phila., 1969.