Stochastic Differential Equations:

A Dynamical Systems Approach

Except where reference is made to the work of others, the work described in this dissertation is my own or was done in collaboration with my advisory committee. This dissertation does not include proprietary or classified information.

Blane Jackson Hollingsworth

Certificate of Approval:

Georg Hetzer
Professor
Mathematics and Statistics

Paul Schmidt, Chair
Professor
Mathematics and Statistics

Ming Liao
Professor
Mathematics and Statistics

Wenxian Shen
Professor
Mathematics and Statistics

Joe F. Pittman
Interim Dean
Graduate School

Stochastic Differential Equations: A Dynamical Systems Approach

Blane Jackson Hollingsworth

A Dissertation

Submitted to

the Graduate Faculty of

Auburn University

in Partial Fulfillment of the

Requirements for the

Degree of

Doctor of Philosophy

Auburn, Alabama
May 10, 2008

Stochastic Differential Equations:

A Dynamical Systems Approach

Blane Jackson Hollingsworth

_____

Signature of Author

_____

Date of Graduation

Vita

Blane Hollingsworth was born in Huntsville, Alabama in 1976. His parents are Dianne and Sonny Hollingsworth. He attended the University of Alabama in Huntsville from 1994 to 2000, receiving both his B.S. and M.A. degrees in mathematics. In fall of 2000, he entered the Ph.D. program at Auburn University.

Dissertation Abstract

Stochastic Differential Equations:

A Dynamical Systems Approach

Blane Jackson Hollingsworth

Doctor of Philosophy, May 10, 2008
(B.S., University of Alabama in Huntsville, 1998)
(M.A., University of Alabama in Huntsville, 2000)

121 Typed Pages

Directed by Paul Schmidt

The relatively new subject of stochastic differential equations has increasing importance in both theory and applications. The subject draws upon two main sources, probability/stochastic processes and differential equations/dynamical systems. There exists a significant "culture gap" between the corresponding research communities. The objective of the dissertation project is to present a concise yet mostly self-contained theory of stochastic differential equations from the differential equations/dynamical systems point of view, primarily incorporating semigroup theory and functional analysis techniques to study the solutions. Prerequisites from probability/stochastic processes are developed as needed. For continuous-time stochastic processes whose random variables are (Lebesgue) absolutely continuous, the Fokker-Planck equation is employed to study the evolution of the densities, with applications to predator-prey models with noisy coefficients.

Acknowledgments

No one deserves more thanks than Dr. Paul Schmidt for his patience and guidance throughout this endeavor. Dr. Georg Hetzer, Dr. Ming Liao, and Dr. Wenxian Shen are all deserving of thanks as well, as I have taken one or more important classes from each of them and they are members of my committee. Also, I'd like to thank Dr. Olav Kallenberg, who guided me through the chapters on stochastic differential equations in his book during independent study. Finally, I would like to thank my parents for all their love and support.

Style manual or journal used Journal of Approximation Theory (together with the style known as "aums"). Bibliograpy follows van Leunen's *A Handbook for Scholars*.

Computer software used The document preparation package TEX (specifically LATEX) together with the departmental style-file `aums.sty`.

Table of Contents

Chapter 1

Introduction and Preliminaries

1.1 Stochastic Processes and Their Distributions

Let $(\Omega, \mathcal{A}, P)$ be a probability space, $(S, \mathcal{B})$ a measurable space, and X an S-valued random variable on Ω, that is, a mapping from Ω into S that is measurable with respect to the σ-algebras $\mathcal{A}$ and $\mathcal{B}$. By the distribution of X, denoted by P^X, we mean the image of the probability measure P under the mapping X, that is, the probability measure on $\mathcal{B}$, defined by $P^X(B) := P(X \in B) := P(X^{-1}(B))$ for $B \in \mathcal{B}$. (Here, as in the sequel, we take some liberties in our terminology. To be precise, we should of course refer to X as an $(S, \mathcal{B})$-valued random variable on $(\Omega, \mathcal{A})$ and to P^X as its P-distribution.)

Now let T be a non-empty set and $X = (X_t)_{t \in T}$ a family of S-valued random variables on Ω; we call X a stochastic process on Ω, with state space S and index set T. Clearly, X can be thought of as a mapping from Ω into the Cartesian product S^T, defined by $X(\omega) := X^\omega := (X_t(\omega))_{t \in T}$ for $\omega \in \Omega$. The image X^ω of a point $\omega \in \Omega$ is called the path of ω; the set S^T, endowed with the product σ-algebra induced by $\mathcal{B}$, is called the path space of X. With slight abuse of notation, we denote the product σ-algebra of S^T by $\mathcal{B}^T$. Since $\mathcal{B}^T$ is generated by the coordinate projections $\Pi_t : S^T \to S$, defined by $\Pi_t(x) := x_t$ for $x \in S^T$ and $t \in T$, and since $X_t = \Pi_t \circ X$ for $t \in T$, measurability of X with respect to the σ-algebras $\mathcal{A}$ and $\mathcal{B}^T$ is equivalent to the measurability of X_t for every $t \in T$. In other words, X is an S^T-valued random variable on Ω. Its distribution P^X, a probability measure on $\mathcal{B}^T$, is called the joint distribution of the random variables X_t, $t \in T$.

It follows from a standard uniqueness theorem of measure theory that the probability measure P^X is uniquely determined by the values

$$P^X\Big(\bigcap_{t\in F}\Pi_t^{-1}(B_t)\Big) = P\Big(\bigcap_{t\in F}X_t^{-1}(B_t)\Big) = P(X_t\in B_t\,\forall\, t\in F),$$

where F varies over the non-empty finite subsets of T and $(B_t)_{t\in F}$ over the corresponding finite families of sets in $\mathcal{B}$. In particular, even if T is infinite, the distribution of the family $(X_t)_{t\in T}$ is uniquely determined by the distributions of the "finite subfamilies" $(X_t)_{t\in F}$ with $\emptyset\neq F\subset T$ finite, that is, by the probability measures $Q_{t_1,\ldots,t_n} := P^{(X_{t_1},\ldots,X_{t_n})}$ with $n\in\mathbb{N}$ and $(t_1,\ldots,t_n)\in T^n$ injective (that is, $t_1,\cdots,t_n$ are pairwise distinct); these are called the finite joint distributions of the random variables X_t, $t\in T$, or the finite-dimensional distributions of the process X.

Note that for each $n\in\mathbb{N}$ and $(t_1,\ldots,t_n)\in T^n$ injective, $Q_{t_1,\ldots,t_n}$ is a probability measure on the product σ-algebra $\mathcal{B}^n$ of S^n induced by $\mathcal{B}$. Clearly, if $B_1,\ldots,B_n\in\mathcal{B}$ and π is a permutation of the set $\{1,\ldots,n\}$, then

$$\begin{aligned}Q_{t_1,\ldots,t_n}\big(B_{\pi^{-1}(1)}\times\cdots\times B_{\pi^{-1}(n)}\big) &= P\big(X_{t_i}\in B_{\pi^{-1}(i)}\,\forall\, i\in\{1,\ldots,n\}\big)\\ &= P\big(X_{t_{\pi(j)}}\in B_j\,\forall\, j\in\{1,\ldots,n\}\big) = Q_{t_{\pi(1)},\ldots,t_{\pi(n)}}\big(B_1\times\cdots\times B_n\big).\end{aligned}$$

Also, if $n\geq 2$ and $B_n = S$, then

$$\begin{aligned}Q_{t_1,\ldots,t_n}\big(B_1\times\cdots\times B_n\big) &= P\big(X_{t_i}\in B_i\,\forall\, i\in\{1,\ldots,n\}\big)\\ &= P\big(X_{t_i}\in B_i\,\forall\, i\in\{1,\ldots,n-1\}\big) = Q_{t_1,\ldots,t_{n-1}}\big(B_1\times\cdots\times B_{n-1}\big).\end{aligned}$$

Under certain restrictions on the state space S, a theorem due to Kolmogorov ensures, roughly speaking, that *any* family of probability measures $Q_{t_1,\ldots,t_n}$, consistent with the

above conditions, is in fact the family of finite-dimensional distributions of a stochastic process on some probability space $(\Omega, \mathcal{A}, P)$; recall that $(S, \mathcal{B})$ is a Polish space if $\mathcal{B}$ is the Borel σ-algebra generated by a complete and separable metric topology on S. Then we have

Theorem 1.1 (Kolmogorov)**.** *Suppose $(S, \mathcal{B})$ is a Polish space (that is, $\mathcal{B}$ is the Borel σ-algebra generated by a complete and separable metric topology on S), T is a non-empty set, and for each $n \in \mathbb{N}$ and $(t_1, \dots, t_n) \in T^n$ injective, $Q_{t_1,\dots,t_n}$ is a probability measure on $\mathcal{B}^n$ (the product σ-algebra of S^n, which in this case coincides with the Borel σ-algebra generated by the product topology of S^n). Further suppose that the following two conditions are satisfied for all $n \in \mathbb{N}$, $(t_1, \dots, t_n) \in T^n$ injective, and $B_1, \dots, B_n \in \mathcal{B}$:*

(a) If π is a permutation of $\{1, \dots, n\}$, then

$$Q_{t_1,\dots,t_n}\big(B_{\pi^{-1}(1)} \times \cdots \times B_{\pi^{-1}(n)}\big) = Q_{t_{\pi(1)},\dots,t_{\pi(n)}}\big(B_1 \times \cdots \times B_n\big).$$

(b) If $n \geq 2$ and $B_n = S$, then

$$Q_{t_1,\dots,t_n}\big(B_1 \times \cdots \times B_n\big) = Q_{t_1,\dots,t_{n-1}}\big(B_1 \times \cdots \times B_{n-1}\big).$$

Then there exists a probability space $(\Omega, \mathcal{A}, P)$, along with a family $X = (X_t)_{t \in T}$ of S-valued random variables on Ω, such that $Q_{t_1,\dots,t_n} = P^{(X_{t_1},\dots,X_{t_n})}$ for all $n \in \mathbb{N}$ and $(t_1, \dots, t_n) \in T^n$ injective.

Note that while neither the probability space $(\Omega, \mathcal{A}, P)$ nor the process X are uniquely determined, the distribution P^X is. We refer to [3, Section 35] for a detailed exposition of these issues and a proof of Kolmogorov's theorem (see in particular Theorem 35.3 and

Corollary 35.4 ibidem). For most purposes, the distribution of a stochastic process is much more important than the process itself. This implies, of course, that both have the same state space and index set, but the underlying probability spaces may be different.

Definition 1.1. *Two processes are called equivalent if both have the same distribution.*

Definition 1.2. *Two processes* $X = (X_t)_{t\in T}$ *and* $Y = (Y_t)_{t\in T}$ *over the same probability space* $(\Omega, \mathcal{A}, P)$*, with the same state space and index set, are called modifications of each other if* $P(\bigcup_{t\in T}(X_t \neq Y_t)) = 0$.

It is easily verified that any two processes that are modifications of each other have the same finite-dimensional distributions and are thus equivalent.

Now suppose that X is a stochastic process over $(\Omega, \mathcal{A}, P)$, with state space $(S, \mathcal{B})$, index set T, and distribution $Q = P^X$. Then X is equivalent to the process $\Pi := (\Pi_t)_{t\in T}$ on $(S^T, \mathcal{B}^T, Q)$. To see this, note that the coordinate projections Π_t, for $t \in T$, are S-valued random variables on S^T and that Π, as a mapping from S^T into S^T, is the identity map: $\Pi(\omega) := \Pi^\omega := (\Pi_t(\omega))_{t\in T} = (\omega_t)_{t\in T} = \omega$ for all $\omega \in S^T$; hence, $Q^\Pi = Q$.

Definition 1.3. *The process* Π *as above is called the canonical process with distribution* Q.

We can think of the canonical process as the standard representative of the equivalence class of stochastic processes with distribution Q. Using this terminology, the assertion of (1.1) may be stated as follows: There exists a unique probability measure Q on $\mathcal{B}^T$ such that the given probability measures $Q_{t_1,\dots,t_n}$ coincide with the finite-dimensional distributions of the canonical process Π with distribution Q.

In the following we assume that the state space $(S, \mathcal{B})$ is Polish (as in (1.1)) and that X is a so-called continuous-time process, that is, the index set T is $\mathbb{R}_+$. For each $\omega \in \Omega$,

the path X^ω is then a curve in S, parametrized with $t \in \mathbb{R}_+$. If the curve X^ω is continuous for every (or P-almost every) $\omega \in \Omega$, we say that X has continuous (or almost surely continuous) paths. If X has almost surely continuous paths, an obvious and inconsequential modification of the underlying probability space will turn X into a process with continuous paths. Also, any process with almost surely continuous paths admits a modification with continuous paths.

Saying that X has continuous paths is equivalent to saying that X maps Ω into the subspace $C := C(\mathbb{R}_+, S)$ of $S^{\mathbb{R}_+}$, that is, the subspace of continuous mappings from $\mathbb{R}_+$ into S. This space is, in general, not measurable as a subset of $S^{\mathbb{R}_+}$; in fact, $C \notin \mathcal{B}^{\mathbb{R}_+}$ unless S is a singleton (see [3, Corollary 38.5]). However, C is a Polish space under the topology of uniform convergence on compact subsets of $\mathbb{R}_+$, and the trace σ-algebra $C \cap \mathcal{B}^{\mathbb{R}_+} := \{C \cap B \,|\, B \in \mathcal{B}^{\mathbb{R}_+}\}$ coincides with the Borel σ-algebra generated by this topology. Also, C inherits a topology from $S^{\mathbb{R}_+}$ (the product topology, which coincides with the topology of pointwise convergence on $\mathbb{R}_+$), and the trace σ-algebra $C \cap \mathcal{B}^{\mathbb{R}_+}$ coincides with the Borel σ-algebra generated by that topology as well (see [3, Theorem 38.6]).

Now suppose that X is a continuous-time process with Polish state space $(S, \mathcal{B})$ and distribution Q and that X is equivalent to a process with continuous paths. Then X is in fact equivalent to the process $\widetilde{\Pi} := (\Pi_t|_C)_{t\in\mathbb{R}_+}$ on $(C, C \cap \mathcal{B}^{\mathbb{R}_+}, \widetilde{Q})$, where $C := C(\mathbb{R}_+, S)$ and $\widetilde{Q}$ is defined by $\widetilde{Q}(C \cap B) := Q(B)$ for $B \in \mathcal{B}^{\mathbb{R}_+}$. That $\widetilde{Q}$ is well defined as a probability measure on the trace σ-algebra $C \cap \mathcal{B}^{\mathbb{R}_+}$ follows from the (non-trivial) fact that $Q(B) = 1$ for all $B \in \mathcal{B}^{\mathbb{R}_+}$ with $B \supset C$ (in other words, C has Q-outer measure 1). For the proof, we refer to [3, Sections 38–39], in particular Theorems 38.2–3 and Lemma 39.2 ibidem. To see that X and $\widetilde{\Pi}$ are equivalent, observe that $\widetilde{\Pi}$, as a mapping from C into $S^{\mathbb{R}_+}$, is simply the

restriction of the identity map of $S^{\mathbb{R}_+}$ to C. Thus, $\widetilde{Q}^{\widetilde{\Pi}}(B) = \widetilde{Q}(\widetilde{\Pi}^{-1}(B)) = \widetilde{Q}(C \cap B) = Q(B)$ for all $B \in \mathcal{B}^{\mathbb{R}_+}$; that is, $\widetilde{Q}^{\widetilde{\Pi}} = Q$.

Definition 1.4. *The process $\widetilde{\Pi}$ as above is called the C-canonical process with distribution Q.*

Whenever an equivalence class of continuous-time processes with Polish state space contains a process with continuous paths, we think of the associated C-canonical process $\widetilde{\Pi}$ (rather than the canonical process Π) as the standard representative of the equivalence class.

In the next two sections, we discuss semigroups, which will be used in the short term to prescribe a family of measures that satisfies Kolmogorov's theorem and hence allows us to construct Brownian motion.

1.2 Semigroups of Linear Operators

Let X be a Banach space. A family $T := (T_t) = (T_t)_{t\in\mathbb{R}_+}$ of bounded linear operators $T_t : X \to X$ is called a semigroup of linear operators (or, more simply, a semigroup) if $T_0 = id_X$ and $T_{t+s} = T_t S_s$ for all $t, s \in \mathbb{R}_+$. If $\lim_{t\to 0^+} \| x - T_t x \| = 0$ for all $x \in X$, then we say T is strongly continuous. The infinitesimal generator (or, more simply, the generator) of a strongly continuous semigroup T is the operator $A : D(A) \subset X \to X$ defined by

$$Ax := \lim_{t\to 0^+} \frac{T_t x - x}{t}$$

for all $x \in D(A)$, the set of $x \in X$ such that the limit exists. We say a semigroup T is a γ-contraction semigroup if, for some nonnegative constant γ, $\| T_t \| \leq e^{\gamma t}$ for all $t \geq 0$, where $\| T_t \|$ is the operator norm of T_t. We say T is a contraction semigroup if $\gamma = 0$.

Call $C_b(\mathbb{R}^n, \mathbb{R})$ the space of bounded, continuous functions mapping $\mathbb{R}^n$ into $\mathbb{R}$, and call $C_0(\mathbb{R}^n, \mathbb{R})$, the subset of $C_b(\mathbb{R}^n, \mathbb{R})$ such that $\lim_{|x|\to\infty} f(x) = 0$. Equip $C_0(\mathbb{R}^n, \mathbb{R})$ with the sup norm to make it a Banach space.

Definition 1.5. *A contraction semigroup of linear operators T on $C_0(\mathbb{R}^n, \mathbb{R})$ is called a Feller semigroup if*

1. *for every $t \geq 0$, T_t maps $C_0(\mathbb{R}^n, \mathbb{R})$ into itself, and*
2. *$\lim_{t\to 0} T_t f(x) = f(x)$ for all $f \in C_0(\mathbb{R}^n, \mathbb{R})$ and $x \in \mathbb{R}^n$.*

It can be shown ([9, Theorem 19.6]) that Feller semigroups are strongly continuous.

Strong continuity is quite valuable due to the following theorem (see [17, Theorem 2.3.2]):

Theorem 1.2. *Any strongly continuous semigroup (G_t) with infinitesimal generator A has the property that, for any $x \in D(A)$, $G_t x \in D(A)$ for all $t \in \mathbb{R}_+$, $t \mapsto G_t x$ is C^1, and*

$$\frac{d}{dt}(G_t x) = AG_t x = G_t Ax.$$

In another way, $u : t \mapsto T_t x$ solves the initial value problem $\dot{u} = Au$, $u(0) = x$. So, formally, $u(t)$ should be of the form e^{tA}, that is, $T_t x = e^{tA}x$ or $T_t = e^{tA}$. We would like to have a way to guarantee that a given operator A, generally unbounded, indeed will be the generator of a strongly continuous semigroup.

Let X be a normed linear space, $A : D(A) \subset X \to X$ a linear operator. Consider the equation

$$Ax = y.$$

To guarantee the existence and uniqueness of a solution $x \in D(A)$, for every $y \in X$, and the continuous dependence of x on y, the operator A must be one-to-one and onto, with a bounded inverse A^{-1}. Assuming X to be complete and A to be closed, the latter is automatic, by the open-mapping theorem. More generally, consider the equation

$$(A - \lambda I)x = y,$$

where $I := id_X$ and $\lambda \in \mathbb{C}$. Existence, uniqueness, and continuous dependence are guaranteed if λ belongs to the resolvent set of A, as defined below.

Definition 1.6. *For a Banach space X and a closed linear operator $A : D(A) \subset X \to X$, define $\rho(A)$, the resolvent set of A, by $\rho(A) := \{\lambda \in \mathbb{C} \mid A - \lambda I$ is one-to-one and onto$\}$. Then define $R(\lambda; A)$, the resolvent of A, by $R(\lambda; A) := (A - \lambda I)^{-1}$.*

Theorem 1.3 (Hille-Yosida)**.** *For a Banach space X, a closed, densely defined linear operator $A : D(A) \subset X \to X$ is the infinitesimal generator of a strongly continuous semigroup of contractions if and only if*

1. *$(0, \infty) \subset \rho(A)$, and*

2. *for each $\lambda > 0$, $\| R(\lambda; A) \| \leq \frac{1}{\lambda}$.*

We refer to [17, pp. 51-56] for the proof.

In the next section, we present a discussion of semigroups of kernels with the construction of Brownian motion in mind.

1.3 Kernels and Semigroups of Kernels

Let $(\Omega_1, \mathcal{A}_1)$ and $(\Omega_2, \mathcal{A}_2)$ be given measurable spaces.

Definition 1.7. *A function $k : \Omega_1 \times \mathcal{A}_2 \to \overline{\mathbb{R}}_+$ with the properties*

1. *$k(\cdot, A_2)$ is $\mathcal{A}_1$-measurable for all $A_2 \in \mathcal{A}_2$,*
2. *$k(\omega, \cdot)$ is a (probability) measure on $\mathcal{A}_2$ for all $\omega \in \Omega_1$,*

is called a (probability) kernel from $(\Omega_1, \mathcal{A}_1)$ to $(\Omega_2, \mathcal{A}_2)$.

We also call a probability kernel a Markov kernel or say that the kernel is Markovian. Further, if $(\Omega_1, \mathcal{A}_1)$ equals $(\Omega_2, \mathcal{A}_2)$, we call k a kernel on $(\Omega_1, \mathcal{A}_1)$, or simply a kernel on Ω_1.

Let us establish some notation here; call $\overline{\mathcal{B}}$ the Borel σ-algebra on $\overline{\mathbb{R}}$, call $\mathcal{B}_+$ the Borel σ-algebra on $\mathbb{R}_+$, and call $\mathcal{B}^n$ the Borel σ-algebra on $\mathbb{R}^n$ for any $n \in \mathbb{N}$. Call λ^n the Lebesgue measure on $(\mathbb{R}^n, \mathcal{B}^n)$; we may simply call $\lambda := \lambda^n$ when n is understood. Given $x \in \mathbb{R}^n$, call ϵ_x the point mass at x, that is, the measure that satisfies $\epsilon_x(A) = 1$ if $x \in A$ and $\epsilon_x(A) = 0$ else, for $A \in \mathcal{B}^n$.

Let $(\Omega, \mathcal{A}, \mu)$ be a σ-finite measure space and let $\mathcal{M}(\Omega, \mathcal{A})$ denote the space of all $\overline{\mathbb{R}}$-valued functions on Ω that are measurable with respect to $\mathcal{A}$ and $\overline{\mathcal{B}}$. For $p \in [1, \infty)$, let $\mathcal{L}^p(\Omega, \mathcal{A}, \mu)$ denote the space of functions f belonging to $\mathcal{M}(\Omega, \mathcal{A})$ such that $|f|^p$ is integrable (with respect to μ); call $\mathcal{L} := \mathcal{L}^1$. Let $\mathcal{L}^\infty(\Omega, \mathcal{A}, \mu)$ denote the space of functions f belonging to $\mathcal{M}(\Omega, \mathcal{A})$ such that the essential supremum of $|f|$ is finite. When the associated σ-algebra and measure are understood, we may abbreviate $\mathcal{L}^p(\Omega)$ and $\mathcal{L}^\infty(\Omega)$ for $\mathcal{L}^p(\Omega, \mathcal{A}, \mu)$ and $\mathcal{L}^\infty(\Omega, \mathcal{A}, \mu)$, respectively; we frequently understand $\mathbb{R}_+ \times \Omega$ to have σ-algebra $\mathcal{B}_+ \times \mathcal{A}$ and measure $\lambda_+ \times \mu$ (where λ_+ is Lebesgue measure on $(\mathbb{R}_+, \mathcal{B}_+)$.

Given f and g in $\mathcal{M}(\Omega, \mathcal{A})$, we say that f is equivalent to g (with respect to μ) if $\mu(f \neq g) = 0$. Note that if f is integrable and equivalent to g, then g is also integrable. Denote the equivalence classes of $\mathcal{M}(\Omega, \mathcal{A})$ and $\mathcal{L}^p(\Omega, \mathcal{A}, \mu)$ by $M(\Omega, \mathcal{A}, \mu)$ and $L^p(\Omega, \mathcal{A}, \mu)$, respectively; we frequently "identify" an equivalence class with an arbitrary member. Also, if $f \in \mathcal{M}(\Omega, \mathcal{A})$ and f is nonnegative, we say $f \in \mathcal{M}_+(\Omega, \mathcal{A})$; we give the analogous meaning to $\mathcal{L}^p_+, M_+$, and L^p_+.

Now, a kernel k from $(\Omega_1, \mathcal{A}_1)$ to $(\Omega_2, \mathcal{A}_2)$ determines a mapping K of $\mathcal{M}_+(\Omega_2, \mathcal{A}_2)$ into $\mathcal{M}_+(\Omega_1, \mathcal{A}_1)$, defined by

$$(Kf_2)(\omega_1) := \int f_2(\omega_2) k(\omega_1, d\omega_2),$$

for $\omega_1 \in \Omega_1$ and $f_2 \in \mathcal{M}_+(\Omega_2, \mathcal{A}_2)$. Let us refer to K as the integral operator associated with k. Note that for any $A_2 \in \mathcal{A}_2$, $K1_{A_2} = k(\cdot, A_2)$. In particular, $K1_{\Omega_2} = 1_{\Omega_1}$ if and only if k is Markovian.

Kernels may be composed in the following way: for $i = 1, 2$, let k_i be a kernel from $(\Omega_i, \mathcal{A}_i)$ to $(\Omega_{i+1}, \mathcal{A}_{i+1})$. We may define the composition $k_1 k_2$ in terms of the composition of the associated integral operators K_1 and K_2:

$$(k_1 k_2)(\cdot, A_3) := K_1 K_2 1_{A_3}.$$

Then $k_1 k_2$ is a kernel from $(\Omega_1, \mathcal{A}_1)$ to $(\Omega_3, \mathcal{A}_3)$, and we have

$$(k_1 k_2)(\omega_1, A_3) = \int k_1(\omega_1, d\omega_2) k_2(\omega_2, A_3),$$

for all $\omega_1 \in \Omega_1$ and $A_3 \in \mathcal{A}_3$. Observe that if k_1, k_2 are Markovian then so is $k_1 k_2$. We need the composition of kernels to define semigroups of kernels.

Definition 1.8. *If (P_t) is a family of kernels on a measurable space $(S, \mathcal{B})$ and if $P_{s+t} = P_s P_t$ for all $s, t \geq 0$, then we say $(P_t) := (P_t)_{t \in \mathbb{R}_+}$ is a semigroup of kernels on S.*

We remark that a semigroup of kernels satisfies $P_{s+t}(x, B) = \int P_s(x, dy) P_t(y, B)$ for $x \in S, B \in \mathcal{B}$, often called the Chapman-Kolmogorov property.

Definition 1.9. *A semigroup of kernels (P_t) is called normal if $P_0(x, \cdot) = \epsilon_x$ for all $x \in S$. We call (P_t) Markovian if each kernel P_t is Markovian.*

Now, let (P_t) be a semigroup of kernels on $(\mathbb{R}^n, \mathcal{B}^n)$.

Definition 1.10. *(P_t) is called translation-invariant if $P_t(x, B) = P_t(x + z, B + z)$ for all $x, z \in \mathbb{R}^n, t \geq 0$, and $B \in \mathcal{B}^n$.*

It can be shown that translation-invariant semigroups of kernels must be normal (see [3, 29.7 and p.311]). The following proposition demonstrates the importance of these semigroups.

Proposition 1.1. *Given a translation-invariant (P_t) as above,*

1. *define $T := (T_t)_{t \geq 0}$ by*

$$T_t f = \int f(y) P_t(\cdot, dy), \tag{1.1}$$

for any $t \in \mathbb{R}_+$ and $f \in L^\infty(\mathbb{R}^n)$. Then T is a contraction semigroup of linear operators on $L^\infty(\mathbb{R}^n)$;

2. *define* $(\mu_t) := (\mu_t)_{t\geq 0}$ *by* $\mu_t(B) := P_t(0, B)$ *for all* $B \in \mathcal{B}^n$. *Then* (μ_t) *is a convolution semigroup of measures on* $\mathcal{B}^n$, *that is, for all* $s, t \geq 0$, *for all* $B \in \mathcal{B}^n$, (μ_t) *satisfies*

$$\mu_{s+t}(B) = \int \mu_s(dy)\mu_t(B - y) = (\mu_s * \mu_t)(B).$$

Proof. The second claim is simple, so we only address the first claim (which in fact does not require translation invariance of (P_t)). It remains to note that T is indeed a semigroup; (P_t) satisfies the Chapman-Kolmogorov property and T_0 is the identity mapping since (P_t) is normal:

$$T_0 f(x) = \int f(y)P_0(x, dy) = \int f(y)\delta_x(dy) = f(x).$$

Note that $\|T_t f\| \leq \|f\|$ since (P_t) is Markovian, so that T is a contraction semigroup. □

Conversely, given a convolution semigroup of measures (μ_t) on $\mathcal{B}^n$, if we define $P_t(x, B) := \mu_t(B - x)$ for all $t \geq 0, x \in \mathbb{R}^n, B \in \mathcal{B}^n$, then (P_t) is a translation-invariant semigroup of kernels on $\mathbb{R}^n$ ([3, pp. 310-311]). Notice that (P_t) is a translation-invariant Markov semigroup iff (μ_t) is a convolution semigroup of probability measures.

At this point, an intuitive interpretation of a translation-invariant Markov semigroup is helpful. Think of $P_t(x, B)$ as the probability that a randomly moving particle starting at x at time 0 is in the set B at time t. We see that the semigroup property means there is no memory, in the sense that we need not understand the history of the particle's movement, rather, we only need to know where it is at time t to yield the probability that it is in some set at time $t + s$. Thinking that the particle is "in dy" at time t, we can see from the Chapman-Kolmogorov property that $P_{t+s}(x, B) = \int P_t(x, dy)P_s(y, B)$, or that the probability a particle is in a set B at time $t + s$ can be obtained from $P_t(x, dy)$ (which we

think of as the "present") and $P_s(y, B)$ (which is the probability that the particle starting at y ends up in B at time s). This semigroup reasoning is similar to concepts in deterministic dynamical systems, which will be discussed later.

Armed with this intuitive understanding of translation-invariant Markovian semigroups of kernels, we realize the next step: that translation-invariant Markovian semigroups of kernels lead to measures which satisfy the hypotheses of Kolmogorov's theorem, and hence, lead to the construction of stochastic processes (in particular, Brownian motion) which will model random particle motion in a natural way. The idea is, if we take times $t_1 < t_2 < \cdots < t_k$ and sets $B_1, B_2, \cdots, B_k$ in $\mathcal{B}^n$, we may construct the iterated integral

$$\int_{B_1} \int_{B_2} \cdots \int_{B_k} P_{t_k - t_{k-1}}(x_{k-1}, dx_k) P_{t_{k-1} - t_{k-2}}(x_{k-2}, dx_{k-1}) \cdots P_{t_1}(x_0, dx_1). \tag{1.2}$$

For a particle starting at x_0, this integral models random particle motion without memory, in the sense that it gives the probability that at times $t_1, t_2, \cdots, t_k$, the particle is found successively in $B_1, B_2, \cdots, B_k$. We could even, by tacking on another integral in (1.2), impose that the particle's initial location is random; let μ be a probability measure on $\mathcal{B}^n$ that describes the distribution of the initial location of the particle. Then we would integrate over $\mathbb{R}^n$ with respect to μ over the variable x_0:

$$\int_{\mathbb{R}^n} \int_{B_1} \cdots \int_{B_k} P_{t_k - t_{k-1}}(x_{k-1}, dx_k) \cdots P_{t_1}(x_0, dx_1) \mu(dx_0). \tag{1.3}$$

Then it can be shown [3, 36.4] that given (P_t) and μ as above, for $x := (x_1, x_2, \cdots, x_k)$ and for any $B \in \bigotimes_{i=1}^{k} \mathcal{B}_i$ (where $\mathcal{B}_i = \mathcal{B}^n$ for all $1 \leq i \leq k$), the measures $P_{t_1,t_2,\cdots,t_k}$, defined by

$$P_{t_1,t_2,\cdots,t_k}(B) := \int_{\mathbb{R}^n} \int_{B_1} \cdots \int_{B_k} 1_B(x) P_{t_k - t_{k-1}}(x_{k-1}, dx_k) \cdots P_{t_1}(x_0, dx_1)\mu(dx_0), \qquad (1.4)$$

for $B \in \mathcal{B}^n$, satisfy the hypotheses of Kolmogorov's theorem. The family of measures in (1.4) are thus the finite-dimensional distributions of some stochastic process with state space $\mathbb{R}^n$.

The canonical process X associated with this stochastic process has a distribution which depends only on (P_t) and μ, so let us denote this distribution by P^μ. This means

$$P^\mu(X_{t_1} \in B_1, X_{t_2} \in B_2, \cdots, X_{t_k} \in B_k) =$$

$$\int_{\mathbb{R}^n} \int_{B_1} \int_{B_2} \cdots \int_{B_k} P_{t_k - t_{k-1}}(x_{k-1}, dx_k) \cdots P_{t_0}(x_0, dx_1)\mu(dx_0)$$

holds for all $B_1, B_2, \cdots, B_k$ in $\mathcal{B}^n$. Also, the P^μ-distribution of X_0 is μ, and so we may refer to μ as the initial distribution of the process.

Processes constructed as in the above enjoy some useful and intuitive properties.

Definition 1.11. *A process X with state space $(\mathbb{R}^n, \mathcal{B}^n)$ has stationary increments if there is a family of probability measures (μ_t) on $\mathcal{B}^n$ such that $\mu_{t-s} = P^{X_t - X_s}$; this means that the distribution of $X_t - X_s$ depends only on $t - s$.*

Definition 1.12. *A process X with state space $(\mathbb{R}^n, \mathcal{B}^n)$ has independent increments if $X_{t_0}, X_{t_1} - X_{t_0}, \cdots, X_{t_k} - X_{t_{k-1}}$ are all independent for any $t_0, t_1, \ldots, t_k \in \mathbb{R}_+$ with $t_0 < t_1 < \ldots < t_k$, for any $k \geq 1$.*

It can be shown ([3, 37.2]) that the canonical process derived from a translation-invariant Markov semigroup of kernels (P_t) and initial distribution μ has stationary and independent increments.

In the next section, we will explain conditional probability, martingales, and Markov processes, and then we will be able to prescribe a particular (P_t) so that we can construct Brownian motion.

1.4 Conditional Expectation, Martingales, and Markov Processes

Let $(\Omega, \mathcal{A}, \mu)$ be a σ-finite measure space, $f \in \mathcal{L}(\Omega, \mathcal{A}, \mu)$. Then for any $A \in \mathcal{A}$, we define

$$\mu_f(A) := \int_A f d\mu.$$

We say that μ_f is the signed measure that has density f with respect to μ; this implies the relation

$$\int g d\mu_f = \int g f d\mu,$$

for all $g \in \mathcal{L}(\Omega, \mathcal{A}, \mu_f)$. Further, note that $g \in \mathcal{L}(\Omega, \mathcal{A}, \mu_f)$ iff $gf \in \mathcal{L}(\Omega, \mathcal{A}, \mu)$. Finally, note that μ_f is a finite signed measure on $(\Omega, \mathcal{A})$ that is absolutely continuous with respect to μ, that is, $\mu_f(A) = 0$ whenever $A \in \mathcal{A}$ and $\mu(A) = 0$.

Conversely, given any finite signed measure ν on $(\Omega, \mathcal{A})$ that is absolutely continuous with respect to μ, there exists by the Radon-Nikodym theorem a function $f \in \mathcal{L}(\Omega, \mathcal{A}, \mu)$, unique up to modification on a μ-null set, such that $\nu = \mu_f$. The equivalence class of all f such that $\nu = \mu_f$ is called the Radon-Nikodym derivative of ν with respect to μ and is denoted $\frac{d\nu}{d\mu}$. Note that if f is any representative of $\frac{d\nu}{d\mu}$, we have $\int g d\nu = \int g f d\mu$ for all $g \in \mathcal{L}(\Omega, \mathcal{A}, \nu)$, or (formally) $d\nu = f d\mu$; we frequently "identify" $\frac{d\nu}{d\mu}$ with an arbitrary representative. This justifies the "differential" notation $\frac{d\nu}{d\mu}$ for Radon-Nikodym derivatives. We also have a number of rules for Radon-Nikodym derivatives that are reminiscent of the rules of differential calculus, for example, the chain rule: if μ_1 is a finite signed measure on $(\Omega, \mathcal{A})$, if μ_2 and μ_3 are finite measures on $(\Omega, \mathcal{A})$, if μ_1 is absolutely continuous with respect to μ_2, and if μ_2 is absolutely continuous with respect to μ_3, then μ_1 is absolutely

continuous with respect to μ_3, and

$$\frac{d\mu_1}{d\mu_3} = \frac{d\mu_1}{d\mu_2}\frac{d\mu_2}{d\mu_3}.$$

In particular, $\frac{d\mu_2}{d\mu_3}\frac{d\mu_3}{d\mu_2} = 1$ when μ_2 and μ_3 are both absolutely continuous with respect to each other.

Now, let $(\Omega, \mathcal{A}, P)$ be a probability space, let $\xi \in \mathcal{L}(\Omega, \mathcal{A}, P)$, and let $\mathcal{F} \subset \mathcal{A}$ be a σ-algebra. Then P_ξ, the signed measure that has density ξ with respect to P, restricts to a measure on $\mathcal{F}$, namely, $P_\xi|_{\mathcal{F}}$, which is absolutely continuous with respect to $P|_{\mathcal{F}}$, the restriction of P to $\mathcal{F}$. This leads to the definition of conditional expectation.

Definition 1.13. *The conditional expectation of ξ given $\mathcal{F}$, denoted $E_{\mathcal{F}}\xi$ or $E(\xi|\mathcal{F})$, is the Radon-Nikodym derivative of $P_\xi|_{\mathcal{F}}$ with respect to $P|_{\mathcal{F}}$.*

Note that $E(\xi|\mathcal{F})$ is the unique member of $L(\Omega, \mathcal{F}, P|_{\mathcal{F}})$ such that

$$\int_F E(\xi|\mathcal{F})dP = \int_F \xi dP,$$

for all $F \in \mathcal{F}$.

Definition 1.14. *The expected value of ξ, denoted by $E\xi$, is defined as*

$$E\xi := \int_\Omega \xi dP.$$

The conditional expectation of ξ given an event $A \in \mathcal{A}$ with $P(A) > 0$, denoted by $E(\xi|A)$, is defined as

$$E(\xi|A) := \frac{1}{P(A)} \int_A \xi dP.$$

It is helpful to consider examples. We see if $\mathcal{F}$ is the σ-algebra induced by ξ, or if ξ has an $\mathcal{F}$-measurable version, then $E_{\mathcal{F}}\xi = \xi$ P-a.s.; in this case we have the "pull out" property $E_{\mathcal{F}}\xi\eta = \xi E_{\mathcal{F}}\eta$ P-a.s., for any $\eta \in L(\Omega)$. If instead $\mathcal{F} = \{\emptyset, \Omega\}$, or if ξ is independent of $\mathcal{F}$, then $E_{\mathcal{F}}\xi = E\xi$ P-a.s. Along these lines, for $A \in \mathcal{A}$ such that $0 < P(A) < 1$, if we take $\mathcal{F} = \{\emptyset, A, A^c, \Omega\}$, then $E_{\mathcal{F}}\xi = E(\xi|A)1_A + E(\xi|A^c)1_{A^c}$ P-a.s. Also, if $A \in \mathcal{F}, P(A) > 0$, and A has no proper nonempty subset belonging to $\mathcal{F}$, then $E_{\mathcal{F}}\xi|_A = E(\xi|A)$ P-a.s.

We use conditional expectation to define conditional probability; observe that $E1_A = \int_\Omega 1_A dP = P(A)$.

Definition 1.15. *The conditional probability given $\mathcal{F}$, denoted $P_{\mathcal{F}}$, is defined by*

$$P_{\mathcal{F}}(A) := E_{\mathcal{F}}(1_A),$$

for all $A \in \mathcal{A}$.

Note that $P_{\mathcal{F}}$ is not a probability measure, rather, it maps members of $\mathcal{A}$ into $\overline{\mathbb{R}}$-valued random variables on Ω, with the property that $\int_F P_{\mathcal{F}}(A)dP = P(A \cap F)$, for all $F \in \mathcal{F}$.

We will use conditional expectation to define martingales, but first we need some definitions.

Definition 1.16. *Given a measurable space $(\Omega, \mathcal{A})$, a family of σ-algebras $\mathcal{F} := \{\mathcal{F}_t\}_{t\geq 0}$ such that $\mathcal{F}_s \subset \mathcal{F}_t$ for $s \leq t$ with $\mathcal{F}_t \subset \mathcal{A}$ for all $t \geq 0$ is called a filtration of $\mathcal{A}$.*

For simplicity we usually just call $\mathcal{F}$ a filtration. Now let X be a continuous-time stochastic process on $(\Omega, \mathcal{A})$ with state space $(S, \mathcal{B})$.

Definition 1.17. *We call $\sigma(X_s|s \leq t)$ the σ-algebra generated by $(X_s)_{s\leq t}$, that is, the smallest σ-algebra that contains $X_s^{-1}(B)$ for every $B \in \mathcal{B}^n$ and $s \leq t$.*

We call $\mathcal{F}(X) := (\mathcal{F}_t(X))_{t\in\mathbb{R}_+}$ *the filtration generated (or induced) by* X*, where* $\mathcal{F}_t(X) = \sigma(X_s|s \leq t)$ *for each* t*.*

We say X *is adapted to a filtration* $\mathcal{F}$ *if* X_t *is* $\mathcal{F}_t$*-measurable for all* t*.*

Observe that $\mathcal{F}(X)$ is the smallest filtration for which X is adapted. If $\mathcal{F}$ is understood we may simply say that X is adapted.

We see that filtrations add more and more sets (or at least, no less sets) as time increases; by increasing the size of a σ-algebra, the potential for the process to take new values is increased. For example, a measurable $\mathbb{R}$-valued function on Ω that has only one value only generates the trivial σ-algebra (ϕ, Ω). A measurable function taking two values, say, $f(\omega) = 1$ when $\omega \in A$ and $f(\omega) = 0$ when $\omega \in A^c$, generates the σ-algebra $\{\phi, A, A^c, \Omega\}$. Thus the increasing in the filtration describes the "increase of randomness," and the size of $\mathcal{F}_t$ is indicative of the possible deviation of X_t from its expected value.

Now let X have state space $(\overline{\mathbb{R}}, \overline{\mathcal{B}})$.

Definition 1.18. *We say* X *is an integrable process, or simply,* X *is integrable, if* X_t *is an integrable random variable for each* t*.*

Given a filtration $\mathcal{F}$*, we say* X *is a martingale with respect to* $\mathcal{F}$ *if* X *is an integrable, adapted process that satisfies,* P*-a.s.,*

$$X_s = E(X_t|\mathcal{F}_s),$$

for every t*, for* $s \leq t$*.*

For an example of a martingale, fix $\xi \in \mathcal{L}(\Omega, \mathcal{A}, P)$ and a filtration $\mathcal{F}$. Define a continuous-time process M by

$$M_t = E(\xi|\mathcal{F}_t),$$

for every t. Then M is integrable and adapted to $\mathcal{F}$. For $s \leq t$, we have $\mathcal{F}_s \subset \mathcal{F}_t$, so $E_{\mathcal{F}_s} E_{\mathcal{F}_t} \xi = E_{\mathcal{F}_s} \xi$ P-a.s. Thus we have the P-a.s. relation

$$M_s = E(\xi|\mathcal{F}_s) = E(E(\xi|\mathcal{F}_t)|\mathcal{F}_s) = E(M_t|\mathcal{F}_s).$$

Thus, M is a martingale. Intuitively, martingales are "fair games" in the sense that the expected value of "winnings" at a later time are exactly the value of "winnings" at present.

Next, let $\mathcal{F}_1, \mathcal{F}_2$, and $\mathcal{G}$ be sub σ-algebras of $\mathcal{A}$.

Definition 1.19. *The σ-algebras $\mathcal{F}_1$ and $\mathcal{F}_2$ are called conditionally independent given $\mathcal{G}$, denoted $\mathcal{F}_1 \perp_{\mathcal{G}} \mathcal{F}_2$, if a.s.,*

$$P_{\mathcal{G}}(F_1 \cap F_2) = P_{\mathcal{G}}(F_1) P_{\mathcal{G}}(F_2)$$

for all $F_1 \in \mathcal{F}_1$, $F_2 \in \mathcal{F}_2$.

We now define Markov processes.

Definition 1.20. *For X a continuous-time process on $(\Omega, \mathcal{A})$ and a filtration $\mathcal{F}$ of $\mathcal{A}$, we call X a Markov process if it is adapted to $\mathcal{F}$ and if for all $s, t \in \mathbb{R}_+$ with $s \leq t$, $\mathcal{F}_s$ and $\sigma(X_t)$ are conditionally independent given $\sigma(X_s)$.*

Intuitively, for Markov processes one may think that the past is independent of the future given the present, in the sense that knowing the state X_s makes the future predictions X_t independent of the "history" $\mathcal{F}_s$.

Markov processes are precisely those processes which are generated by translation-invariant Markovian semigroups of kernels (with respect to the induced filtration; the non-trivial proof can be found in [3, Theorem 42.3]). Since the semigroup property is essential both to dynamical systems and the construction of a Markov process, one can interpret a Markov process as a randomized dynamical system. As we will see, Markov processes are of value in understanding the dynamics generated by solutions of stochastic differential equations (much like the dynamics of deterministic differential equations).

In the next section, we will motivate the need for Brownian motion and prescribe a special translation-invariant Markovian semigroup of kernels in order to construct it. We will further prove some useful properties of Brownian motion.

1.5 Brownian Motion

We will now proceed to prescribe the specific Markov semigroup of kernels (P_t) to construct Brownian motion. We will first motivate our selection of (P_t) by returning to our intuition of how particles undergo random motion. Consider the "drunken sailor" problem, where a drunken sailor stands on the origin 0 and starts taking unit length steps in random directions. After each step, he randomly steps in a different direction. The question is, "Where does he end up after n steps?"

The obvious answer is that we do not know; his position is described by a random variable. He is expected to be where he started, as he has the same chance of going left as right, or forward as backward. But the variance depends directly on the number of steps; he cannot stray far in a short number of steps, for example, so one could expect a low variance in this case. So what is the distribution of this random variable?

The key is the Central Limit Theorem; one fairly simple version is in [9, Proposition 5.9], which says that for independent, identically distributed $\mathbb{R}^d$-valued random variables $\xi, \xi_1, \xi_2, \ldots$ with $E\xi = 0$ and $E\xi^2 = 1$, then as $n \to \infty$, $n^{-\frac{1}{2}} \sum_{k \leq n} \xi_k$ converges in distribution to a standard normally distributed random variable ζ, that is, a normally distributed random variable with mean 0 and variance 1. We may say for brevity that ζ is $N(0, 1)$. So, in the drunken sailor problem, the random variable describing where a sailor will end up after a large enough number of steps is normal with mean 0 and variance n (see e.g. [3, p.221-p.226]).

Now, one can think of n as time moving continuously rather than as a discrete number of steps; call it t now. So, imagine a continuous-time stochastic process X having initial distribution ϵ_x. This represents the initial location of a particle at x known with probability

1, where the densities (assuming they exist) of X_t as time increases "flatten" into a Gaussian curve, successively getting "more flat" the more time increases.

Imagine now that the sailor is not drunk, but in a heavy crowd, so that he is being pushed around in a random direction. This is essentially the same problem, but it makes more sense in a physical interpretation; particles are interacting with other particles, being bumped into other particles which in turn bump into other particles ad infinitum. This model of particle movement is called Brownian motion, and it is a stochastic process where, at time t, each random variable B_t has distribution $N(0,t)$.

This type of random interference can be thought to perturb a trajectory as well, not just a stationary object. For example, if a ball is thrown, one can model its path. But now suppose there is lots of wind blowing in random directions; where does the ball go? To describe this, we incorporate a "noise term" in the differential equation. Quite sensibly, this term should somehow be based on Brownian motion, which changes the otherwise deterministic trajectory of the ball into a continuous-time stochastic process.

Recall that $N(m,t)$ as a probability measure over $(\mathbb{R}, \mathcal{B})$ has (Lebesgue) density

$$g_{m,t}(x) := (\frac{1}{2\pi t})^{\frac{1}{2}} e^{\frac{-(x-m)^2}{2t}},$$

and observe that $N(0,s) * N(0,t) = N(0,s+t)$. Define the Brownian convolution semigroup of measures (μ_t) on $\mathbb{R}^d$ by setting μ_t equal to the product measure (d-many times) of $N(0,t)$ in $\mathbb{R}$ for each t, that is, $\mu_t := \bigotimes_{i=1}^{d} N(0,t)$. Then we can define the translation-invariant Markov semigroup of kernels (P_t) by

$$P_t(x,A) := \int_A (\frac{1}{2\pi t})^{\frac{1}{2}} e^{\frac{-(y-x)^2}{2t}} dy,$$

for any $t \in \mathbb{R}_+$, $x \in \mathbb{R}^d, A \in \mathcal{B}$. After defining our initial condition $\mu := \mu_0 := \epsilon_x$, we may construct a process $\tilde{B}$ as in Section 1.3. If we write $\tilde{B} = (\tilde{B^1}, \tilde{B^2}, \cdots, \tilde{B^d})$, then $\tilde{B^i}$ and $\tilde{B^j}$ are independent for all $i \neq j$ (see [9, Lemma 3.10]).

We now prove $\tilde{B}$ has a continuous version using the following result from [3, Theorem 39.3].

Theorem 1.4 (Continuous Paths)**.** *For a continuous-time stochastic process X on $(\Omega, \mathcal{A})$ with state space $(\mathbb{R}^d, \mathcal{B}^d)$, if for some positive constants a, b, and C, the inequality*

$$E(|X_t - X_s|^a) \leq C|t - s|^{b+1} \tag{1.5}$$

holds for all $s, t \in \mathbb{R}^+$, then X has a continuous version.

We use the following lemma to verify (1.5) for $\tilde{B}$.

Lemma 1.1. *For $\tilde{B}$ as above,*

$$E(|\tilde{B}_t - \tilde{B}_s|^4) = d(d+2)(t-s)^2, \tag{1.6}$$

Proof. This claim follows from the property that $\tilde{B}_{t-s}$ is equal in distribution to $(t-s)^{\frac{1}{2}}\tilde{B}_1$ (called the scaling property) and the following recursion for $N(0,1)$-distributed $\mathbb{R}$-valued random variables ξ on $(\Omega, \mathcal{A})$ (which is easy to prove using integration by parts; see [3, 4.20]):

$$\int_{\mathbb{R}} x^{2n} g_{0,1} dx = E(\xi^{2n}) = (2n-1)E(\xi^{2n-2}), \tag{1.7}$$

for any $n \in \mathbb{R}$. Now, to prove (1.6), we see that, for $(\tilde{B}_t - \tilde{B}_s)_i$ the i-th component of $\tilde{B}_t - \tilde{B}_s$, and for $1 \leq i \leq d$,

$$E(|\tilde{B}_t - \tilde{B}_s|)^4 = E([(\tilde{B}_t - \tilde{B}_s)_1^2 + (\tilde{B}_t - \tilde{B}_s)_2^2 + \cdots + (\tilde{B}_t - \tilde{B})_d^2]^2),$$

and by stationarity and scaling, the above equals

$$(t-s)^2 E([Z_1^2 + Z_2^2 + \cdots + Z_d^2]^2),$$

where $Z := (Z_1, Z_2, \cdots, Z_d)$ is an $\mathbb{R}^d$-valued $N(0,1)$-distributed random variable. By algebra and the independence, the above equals

$$(t-s)^2[\sum_{i=1}^{d} E(Z_i^4) + \prod_{i \neq j} 2E(Z_i^2)E(Z_j^2)]. \tag{1.8}$$

Now, $E(Z_i)^2 = 1$ for all $1 \leq i \leq n$, and by the recursion (1.7), $E(Z_i)^4 = 3E(Z_i)^2$, so (1.8) becomes $(3d + d(d-1))(t-s)^2$, which is $d(d+2)(t-s)^2$, so (1.6) holds. □

So, it is a simple corollary to select $a = 4, b = 1, C = d(d+2)$ and thus satisfy (1.5), so we indeed have a continuous modification B of $\tilde{B}$.

Definition 1.21. *B as above is defined to be a Brownian motion.*

B is unique (up to equivalence to another C-canonical process); in another way, we may interpret Brownian motion to be a probability measure P^{B^0} (called Wiener measure) on the path space $(C(\mathbb{R}_+, \mathbb{R}^n), \mathcal{B}(C(\mathbb{R}_+, \mathbb{R}^n))$.

By construction, Brownian motion is a Markov process; it is easy to see that one-dimensional Brownian motion is also a martingale (with respect to the induced filtration)

since, a.s.,

$$E(B_t|\mathcal{F}_s) = E(B_t + B_s - B_s|\mathcal{F}_s) = E(B_s|\mathcal{F}_s) + E(B_t - B_s|\mathcal{F}_s)$$

$$= B_s + E(B_t - B_s) = B_s,$$

since B has independent increments, $E(B_t - B_s) = E(B_t) - E(B_s) = 0$, and $E_{\mathcal{F}}(X) = X$ a.s. when X is an $\mathcal{F}$-measurable random variable. This means B has stationary increments as well, as we argued in the section on kernels.

Since $\mu_0 = \epsilon_x$, we sometimes write B^x instead of B to emphasize the starting point, and hence we sometimes refer to B^x as a Brownian motion starting at x; if otherwise not stated, we assume the Brownian motion starts at zero. Now we observe that the variance of $B_t^x - B_s^x := B_s - B_t$ is $t - s$. This is because

$$\text{var}(B_t - B_s) = \text{var}(B_{t-s} - x) = E[(B_{t-s} - x)^2] - E(B_{t-s} - x)^2$$

$$= E[B_{t-s}^2 - xB_{t-s} + x^2] - 0 = E[B_{t-s}^2] - x^2 + x^2 = t - s,$$

since B_t has variance t for any t.

In the next chapter, we will see how to integrate with respect to a Brownian motion; this will prove essential to the definition of a stochastic differential equation.

Chapter 2

Ito Integrals and Stochastic Differential Equations

2.1 The Ito Integral

Let $(\Omega, \mathcal{A}, P)$ be a probability space and X a continuous-time, real-valued stochastic process on Ω. Assuming that the paths of X are differentiable, we can define the time-derivative $\dot{X}$ of X by

$$\dot{X}(t,\omega) := \frac{d}{dt} X^{\omega}(t),$$

for $t \in \mathbb{R}_+$ and $\omega \in \Omega$. It is easy to see that the mappings $\dot{X}_t = \dot{X}(t,\cdot)$ are measurable for all $t \in \mathbb{R}_+$, so that $\dot{X}$ is a stochastic process. Unfortunately, differentiability of the paths is a very restrictive assumption. For example, the paths of a one-dimensional Brownian motion B on Ω are continuous but nowhere differentiable [10, Theorem 2.9.18]. Thus, the time-derivative $\dot{B}$ of B, frequently referred to as "white noise," does not exist in the naïve sense. Nevertheless, "white noise" plays an important role in the theory of stochastic differential equations.

By way of motivation, consider a simple scalar ODE,

$$\dot{x} = r(x)x$$

for a function $x : \mathbb{R}_+ \to \mathbb{R}$, where $r : \mathbb{R} \to \mathbb{R}$ is a given, sufficiently smooth function. We can interpret $x(t)$, for $t \in \mathbb{R}_+$, as the density of a population at time t, in which case $r(x)$ represents the per-capita growth rate of the population as a function of its density. The growth rate of any real population is subject to random fluctuations; to model these, we

would like to add "white noise" to the function r. On a purely formal level, this idea leads to a "stochastic differential equation" of the form

$$\dot{X} = \big(r(X) + W\big)X, \tag{2.1}$$

where $W = \dot{B}$ is the (formal) time-derivative of a Brownian motion B. The "solutions" of this "stochastic differential equation" should, of course, be continuous-time stochastic processes X rather than functions $x : \mathbb{R}_+ \to \mathbb{R}$. Since already the "antiderivatives" of W (one-dimensional Brownian motions) have nowhere differentiable paths, we cannot hope to find stochastic processes X that satisfy (2.1) in the naïve sense, that is,

$$\frac{d}{dt}X^\omega(t) = \big(r(X^\omega(t)) + W^\omega(t)\big)X^\omega(t) \tag{2.2}$$

for all $t \in \mathbb{R}_+$ and $\omega \in \Omega$, where $(\Omega, \mathcal{A}, P)$ is the underlying probability space; instead, we have to develop a notion of "weak" or "generalized" solutions of (2.1). The first step, still on a purely formal level, is to rewrite (2.2) as an "integral equation,"

$$X^\omega(t) = X^\omega(0) + \int_0^t \big(r(X^\omega(s)) + W^\omega(s)\big)X^\omega(s)\,ds, \tag{2.3}$$

for $t \in \mathbb{R}_+$ and $\omega \in \Omega$. The most problematic term in (2.3) is, of course, the one involving W (the formal time-derivative of B). This raises the question of how to make sense of integrals of the form $\int_0^t X^\omega(s)\, W^\omega(s)ds$, where $t \in \mathbb{R}_+$, $W = \dot{B}$, for some one-dimensional Brownian motion B on Ω, and X is a continuous-time, real-valued process on Ω. Note that,

formally,

$$\int_0^t X^\omega(s)\, W^\omega(s) ds = \int_0^t X^\omega(s) \dot{B}^\omega(s)\, ds = \int_0^t X^\omega(s)\, dB^\omega(s)$$

for all $t \in \mathbb{R}_+$ and $\omega \in \Omega$. The integral on the right appears to be a Riemann-Stieltjes integral involving the real-valued functions X^ω and B^ω, but unfortunately, the paths of B are not of bounded variation on compact subintervals of $\mathbb{R}_+$ [9, Corollary 13.10]. Thus, the integral does not exist, in general, in the classical Riemann-Stieltjes sense, no matter what assumptions we make about the process X. Nevertheless, it is possible to rigorously define the integral

$$(I_t X)(\omega) = \int_0^t X^\omega(s)\, dB^\omega(s),$$

for $t \in \mathbb{R}_+$, $\omega \in \Omega$, and a reasonably large class of continuous-time, real-valued processes X on Ω, in such a way that $I_t X$ is measurable for every $t \in \mathbb{R}_+$. The process $Y := (I_t X)_{t \in \mathbb{R}_+}$ then qualifies as a weak or generalized antiderivative of WX (that is, a solution of the "stochastic differential equation" $\dot{Y} = WX$). In fact, there are several ways of doing this. Our definition will be based on the use of left Riemann-Stieltjes sums and leads to the so-called Ito integral. Other choices are possible; for example, the use of mid-point Riemann-Stieltjes sums leads to the so-called Stratonovich integral [10, p. 350].

For all of the following, suppose that B is a one-dimensional Brownian motion on Ω and that X is a continuous-time, real-valued stochastic process on Ω. Also, suppose that X is adapted to the filtration $\mathcal{F}(B)$. As discussed in the section on conditional probabilities, this has the interpretation that the random variable X_t, for $t \in \mathbb{R}_+$, is "no more random" than the Brownian motion B up to time t, certainly a reasonable assumption if we think of X as the solution of a "stochastic differential equation" whose randomness is produced by

B. (For now, we ignore the effect of a "random initial condition" $X_0 = X^0$ a.s., where X^0 is a given random variable on Ω, on the solution X.)

Now fix a, $b \in \mathbb{R}_+$ with $a < b$. We wish to define the Ito integral

$$(I_{a,b}X)(\omega) = \int_a^b X^\omega(t)\, dB^\omega(t),$$

for $\omega \in \Omega$, under suitable additional assumptions on X. To that end, endow the interval $[a, b]$ with the Borel σ-algebra $\mathcal{B}_{[a,b]}$ and the Lebesgue-Borel measure $\lambda_{[a,b]}$. The Cartesian product $[a, b] \times \Omega$ is then naturally endowed with the product σ-algebra $\mathcal{B}_{[a,b]} \times \mathcal{A}$ and the product measure $\lambda_{[a,b]} \times P$. Given any filtration $\mathcal{F}$ of $\mathcal{A}$, let $L^p_{\mathcal{F}}([a, b] \times \Omega)$, for $p \geq 1$, denote the set of all (equivalence classes of) $\mathcal{F}$-adapted functions in $L^p([a, b] \times \Omega)$ (that is, functions $Y \in L^p([a, b] \times \Omega)$ such that $Y_t = Y(t, \cdot)$ is $\mathcal{F}_t$-measurable for every $t \in [a, b]$); we are most interested in $L^2_{\mathcal{F}}([a, b] \times \Omega)$.

Lemma 2.1. *For any filtration $\mathcal{F}$ of $\mathcal{A}$, $L^2_{\mathcal{F}}([a, b] \times \Omega)$ is a closed linear subspace of $L^2([a, b] \times \Omega)$.*

Proof. That $L^2_{\mathcal{F}}([a, b] \times \Omega)$ is closed is the only nonobvious part; to see this let $(Y_n)_{n\in\mathbb{N}} \in L^2_{\mathcal{F}}([a, b] \times \Omega)^{\mathbb{N}}$, let $Y \in L^2([a, b] \times \Omega)$, and let $Y_n \to Y$ in $L^2([a, b] \times \Omega)$. Then there is a subsequence $(Y_{k_n})_{n\in\mathbb{N}}$ of $(Y_n)_{n\in\mathbb{N}}$ that converges to Y pointwise almost everywhere. Modifying the functions Y_{k_n} on a set of measure zero if necessary, we may assume that $Y_{k_n}(t, \omega) \to Y(t, \omega)$ for all $t \in [a, b]$ and $\omega \in \Omega$. But then, for every $t \in [a, b]$, $Y(t, \cdot)$ is the pointwise limit of the $\mathcal{F}_t$-measurable functions $Y_{k_n}(t, \cdot)$, and thus, $\mathcal{F}_t$-measurable; that is, $Y \in L^2_{\mathcal{F}}([a, b] \times \Omega)$. □

Definition 2.1. *We call a measurable function Y on $[a,b]\times\Omega$ simple if it can be written in the form*

$$Y(t,\omega)=\sum_{j=1}^{n}Y_j(\omega)1_{[t_{j-1},t_j)}(t),$$

for $t\in[a,b]$ and $\omega\in\Omega$, where n is a positive integer, $(t_j)_{j=0}^n$ is a partition of the interval $[a,b]$, and $(Y_j)_{j=1}^n$ is a sequence of measurable functions on Ω.

Given a filtration $\mathcal{F}$ of $\mathcal{A}$, such Y will belong to $L^2_{\mathcal{F}}([a,b]\times\Omega)$ if and only if Y_j is square-integrable and $\mathcal{F}_{t_{j-1}}$-measurable for every $j\in\{1,\dots,n\}$. The set of all simple functions in $L^2_{\mathcal{F}}([a,b]\times\Omega)$ can be shown to be dense in $L^2_{\mathcal{F}}([a,b]\times\Omega)$ [12, pp. 18-20].

It is obvious how to define the Ito integral $I_{a,b}X$ if $X\in L^2_{\mathcal{F}}([a,b]\times\Omega)$ is simple; given a representation of the form $X(t,\omega)=\sum_{j=1}^n X_j(\omega)1_{[t_{j-1},t_j)}(t)$, for $t\in[a,b]$ and $\omega\in\Omega$, with $n\in\mathbb{N}$, $(t_j)_{j=0}^n$ a partition of $[a,b]$, and $(X_j)_{j=1}^n\in L^2(\Omega)^n$ where X^j is $\mathcal{F}_{t_{j-1}}$-adapted for all $j\in\{1,2,\cdots,n\}$, we let

$$(I_{a,b}X)(\omega):=\sum_{j=1}^{n}X_j(\omega)\big(B^\omega(t_j)-B^\omega(t_{j-1})\big),$$

for $\omega\in\Omega$. The sum on the right-hand side is independent of the representation of X and coincides with the left Riemann-Stieltjes sum of X^ω with respect to B^ω for any partition of $[a,b]$ that is a refinement of the partition $(t_j)_{j=0}^n$.

Theorem 2.1 (Ito Isometry for Simple Functions)**.** *Let $\mathcal{F}=\mathcal{F}(B)$ and let $X\in L^2_{\mathcal{F}}([a,b]\times\Omega)$ be simple. Then $I_{a,b}X\in L^2(\Omega)$ with*

$$\|I_{a,b}X\|^2_{L^2(\Omega)}=\sum_{j=1}^{n}(t_j-t_{j-1})\|X_j\|^2_{L^2(\Omega)}=\|X\|^2_{L^2([a,b]\times\Omega)}.$$

Proof. First note that

$$\|I_{a,b}X\|^2_{L^2(\Omega)} = \int_\Omega \big(\sum_{j=1}^n X_j(B_{t_j} - B_{t_{j-1}})\big)^2 dP$$

$$= \sum_{i,j=1}^n \int_\Omega X_i X_j (B_{t_i} - B_{t_{i-1}})(B_{t_j} - B_{t_{j-1}}) dP$$

$$= \sum_{i,j=1}^n E\big(X_i X_j (B_{t_i} - B_{t_{i-1}})(B_{t_j} - B_{t_{j-1}})\big),$$

where E denotes expectation with respect to P. Next, realize that $i \neq j$ (say, without loss of generality, $i < j$) implies

$$E\big(X_i X_j (B_{t_i} - B_{t_{i-1}})(B_{t_j} - B_{t_{j-1}})\big) = 0. \tag{2.4}$$

This is because X_i, X_j, and $(B_{t_i} - B_{t_{i-1}})$ are $\mathcal{F}_{t_{i-1}}$-measurable and because of independent increments (so that $E(B_{t_j} - B_{t_{j-1}}) = E_{\mathcal{F}_{t_{i-1}}}(B_{t_j} - B_{t_{j-1}})$). Therefore, by definition of conditional probability and use of the the "pull out" property (see the section on conditional probability), we have

$$E\big(X_i X_j (B_{t_i} - B_{t_{i-1}})(B_{t_j} - B_{t_{j-1}})\big)$$

$$= E\big(E_{\mathcal{F}_{t_{i-1}}}(X_i X_j (B_{t_i} - B_{t_{i-1}})(B_{t_j} - B_{t_{j-1}}))\big)$$

$$= E\big(X_i X_j (B_{t_i} - B_{t_{i-1}}) E_{\mathcal{F}_{t_{i-1}}}(B_{t_j} - B_{t_{j-1}})\big)$$

$$= E\big(X_i X_j (B_{t_i} - B_{t_{i-1}}) E\big(B_{t_j} - B_{t_{j-1}}\big)\big)$$

$$= E\big(X_i X_j (B_{t_i} - B_{t_{i-1}})\big) E(B_{t_j} - B_{t_{j-1}}).$$

Since $E(B_{t_j} - B_{t_{j-1}}) = 0$, we have shown (2.4).

Next, when $i = j$,

$$E(X_i X_j (B_{t_i} - B_{t_{i-1}})(B_{t_j} - B_{t_{j-1}})) = E(X_i^2)(t_i - t_{i-1}), \tag{2.5}$$

since by the same argument used to show (2.4),

$$E(X_i^2 (B_{t_i} - B_{t_{i-1}})^2) = E(X_i^2) E(B_{t_i} - B_{t_{i-1}})^2,$$

and since B has stationary increments,

$$E(B_{t_i} - B_{t_{i-1}})^2 = E(B_{t_i - t_{i-1}})^2 = var(B_{t_i - t_{i-1}}) = t_i - t_{i-1}.$$

Finally, by combining (2.4) and (2.5), we see that

$$\|I_{a,b} X\|^2_{L^2(\Omega)} = \sum_{i,j=1}^{n} E(X_i X_j (B_{t_i} - B_{t_{i-1}})(B_{t_j} - B_{t_{j-1}}))$$

$$= \sum_{i=1}^{n} E(X_i^2)(t_i - t_{i-1}) = \|X\|^2_{L^2([a,b]\times\Omega)}.$$

□

Therefore, $I_{a,b}$ is a (linear) isometry from a dense (linear) subspace of $L^2_{\mathcal{F}}([a,b] \times \Omega)$ into $L^2(\Omega)$; as such, it has a unique extension to a linear isometry $I_{a,b} : L^2_{\mathcal{F}}([a,b] \times \Omega) \to L^2(\Omega)$. This defines the Ito integral $I_{a,b} X$ for every $X \in L^2_{\mathcal{F}}([a,b] \times \Omega)$, and we have the Ito isometry,

$$\|I_{a,b} X\|_{L^2(\Omega)} = \|X\|_{L^2([a,b]\times\Omega)}.$$

We will use the symbol $\int_a^b X_t\,dB_t$ to denote the Ito integral $I_{a,b}X$.

Stated rigorously:

Definition 2.2. *Let $\mathcal{F} := \mathcal{F}(B)$. For every $X \in L^2_{\mathcal{F}}([a,b]\times\Omega)$, the Ito integral $\int_a^b X_t\,dB_t$ exists and is defined by*

$$\int_a^b X(t,\omega)\,dB_t(\omega) := \lim_{n\to\infty}\int_a^b Y_n(t,\omega)\,dB_t(\omega)$$

(convergence in $L^2(\Omega)$), where $(Y_n)_{n\in\mathbb{N}}$ is any sequence of simple functions that approach X in $L^2_{\mathcal{F}}([a,b]\times\Omega)$.

Note that, due to Fubini's theorem, $\int_a^b X_t^2\,dt$ is an integrable function on Ω, with $\int_\Omega\big(\int_a^b X_t^2\,dt\big)dP = \|X\|^2_{L^2([a,b]\times\Omega)}$. This allows us to write the Ito isometry in the form

$$E\Big(\int_a^b X_t\,dB_t\Big)^2 = E\Big(\int_a^b X_t^2\,dt\Big).$$

Now, if X is a continuous-time, real-valued process on Ω such that $X \in L^2_{\mathcal{F}}([0,t]\times\Omega)$ for every $t \in \mathbb{R}_+$ then $I_tX = \int_0^t X_s dB_s$ is defined for every $t \in \mathbb{R}_+$, and $IX := (I_tX)_{t\in\mathbb{R}_+}$ is a stochastic process. It can be shown that IX is a martingale with respect to $\mathcal{F}$, and as a consequence, has a modification with continuous paths (see [12, pp.22-26] for more details). In the future, we will assume without saying that IX has continuous paths.

Definition 2.2 is enough to make sense of $\int_0^t X_sW_s ds = \int_0^t X_s dB_s$ on the right hand side of (2.3), provided that $X \in L^2_{\mathcal{F}}([0,t]\times\Omega)$. This condition needs to be part of the notion of a "solution" of equation (2.3). As discussed earlier, $\mathcal{F} = \mathcal{F}(B)$-adaptedness of X is a reasonable requirement as long as the "randomness" of X is "caused" solely by B.

Now, instead of solving just one scalar "stochastic differential equation", we would like to solve coupled systems of such equations. By way of motivation, consider the system

$$\dot{X}^i = (r^i(X) + W^i)X^i, \qquad 1 \leq i \leq n \tag{2.6}$$

for an $\mathbb{R}^n$-valued process $X = (X^1, X^2, \cdots, X^n)$, where

$$r = (r^1, r^2, \cdots, r^n) \ : \ \mathbb{R}^n \ \rightarrow \ \mathbb{R}^n$$

is a sufficiently smooth vector field, $B := (B^1, B^2, \cdots, B^n)$ is an n-dimensional Brownian motion, and $W = (W^1, W^2, \cdots, W^n) = (\dot{B}^1, \dot{B}^2, \cdots \dot{B}^n)$. In integral form, equation (2.6) reads

$$X_t^i = X_0^i + \int_0^t r^i(X_s)X_s^i ds + \int_0^t X_s^i dB_s^i, \qquad 1 \leq i \leq n. \tag{2.7}$$

Using Definition (2.2), the integral on the far right would make sense if we could assume that $X^i \in L^2_{\mathcal{F}(B_i)}([0,t] \times \Omega)$. Unfortunately, this is not a reasonable assumption: due to the coupling of the equations, X^i is affected by *all* components of B; thus X^i should be $\mathcal{F}(B)$-adapted, but cannot be expected to be $\mathcal{F}(B^i)$-adapted! Luckily, the assumption $\mathcal{F} = \mathcal{F}(B)$ in Theorem 2.1 and Definition 2.2 (where B is a one-dimensional Brownian motion) can be relaxed — it is enough to assume that *$\mathcal{F}$ is a filtration of $\mathcal{A}$ such that B is a martingale with respect to $\mathcal{F}$.*

Under this assumption (clearly satisfied if $\mathcal{F} = \mathcal{F}(\mathcal{B})$), the proof of the Ito isometry for simple functions (Theorem 2.1) still goes through (note that $E_{\mathcal{F}}(B^i_{t_j} - B^i_{t_{j-1}}) = 0$, for

$1 \leq j \leq n$, and then, so does the entire construction, culminating in Definition 2.2 (for more details, see ([12, p.24]))

Now, if $B = (B^1, B^2, \cdots, B^n)$ is an n-dimensional Brownian motion, then each component B^i is a martingale with respect to $\mathcal{F} = \mathcal{F}(B)$. This is true since $B_t^i - B_{t-c}^i$ is independent of $\mathcal{F}_{t-c}$ for $c \in (0, t)$, which means

$$E(B_t^i|\mathcal{F}_{t-c}) = E(B_t^i - B_{t-c}^i + B_{t-c}^i|\mathcal{F}_{t-c}) = 0 + B_{t-c}^i. \tag{2.8}$$

As a consequence, the integral on the far right of equation (2.7) is defined if
$X^i \in L^2_{\mathcal{F}(B)}([0, t] \times \Omega)$ for $1 \leq i \leq n$, as desired.

Note that, in vector notation, the system can be written as

$$\dot{X} = U(X) + V(X)W, \tag{2.9}$$

where $U(X) = (r^1(X)X^1, r^2(X)X^2, \cdots, r^n(X)X^n)$ and $V(X)$ is the diagonal $n \times n$-matrix whose diagonal entries are $X^1, \cdots, X^n$. Of course, we would like to consider more general systems of the form (2.9), with arbitrary (sufficiently smooth) functions $U : \mathbb{R}^n \to \mathbb{R}^n$ and $V : \mathbb{R}^n \to \mathbb{R}^{n \times n}$. Also we would like to allow for the possibility that only some of the equations are affected by white noise, say, the first d equations, where $1 \leq d \leq n$. In this case,

$$W = (W^1, \cdots, W^d, 0, \cdots, 0) = (\dot{B}^1, \cdots, \dot{B}^d, 0 \cdots, 0),$$

where B is a d-dimensional Brownian motion. Only the first d columns of V are then relevant, and we may as well assume that $V : \mathbb{R}^n \to \mathbb{R}^{n \times d}$.

Under these assumptions, the integral version of equation (2.9) is

$$X_t = X_0 + \int_0^t U(X_s)ds + \int_0^t V(X_s)dB_s. \tag{2.10}$$

Of course, the integrals are understood "componentwise", that is,

$$\int_0^t U(X_s)ds = \left(\int_0^t U^i(X_s)ds\right)_{i=1}^n,$$

$$\int_0^t V(X_s)dB_s = \left(\int_0^t \sum_{j=1}^d V^{ij}(X_s)dB_s^j\right)_{i=1}^n = \left(\sum_{j=1}^d \int_0^t V^{ij}(X_s)dB_s^j\right)_{i=1}^n.$$

The second integral is well defined, provided that $V^{ij}(X) \in L^2_{\mathcal{F}}([0,t]\times\Omega)$ for all $1 \le i \le n$, $1 \le j \le d$, where $\mathcal{F}$ is a filtration of $\mathcal{A}$ such that each component of B is a martingale with respect to $\mathcal{F}$. Note that if X is $\mathcal{F}$-adapted and V is continuous, then $V(X)$ is $\mathcal{F}$-adapted.

This motivates the following definition.

Definition 2.3. *Let $(\Omega, \mathcal{A}, P)$ be a probability space, $n \in \mathbb{N}$, $d \in \{1, \cdots, n\}$. Let B be a d-dimensional Brownian motion on Ω, $\mathcal{F}$ a filtration of $\mathcal{A}$ such that each component of B is a martingale with respect to $\mathcal{F}$. Let U be an $\mathbb{R}^n$-valued process on Ω, V an $\mathbb{R}^{n\times d}$-valued process on Ω such that $U^i \in L^2_{\mathcal{F}}([0,t]\times\Omega)$ and $V^{ij} \in L^2_{\mathcal{F}}([0,t]\times\Omega)$ for all $1 \le i \le n$, $1 \le j \le d$, $t \in \mathbb{R}_+$.*

If X^0 is an $\mathbb{R}^n$-valued random variable on Ω, the process X, defined by

$$X_t = X^0 + \int_0^t U_s ds + \int_0^t V_s dB_s, \tag{2.11}$$

for $t \in \mathbb{R}_+$, is called a stochastic integral generated by (U, V). The set of all *stochastic integrals generated by (U, V) is denoted by*

$$\int U_t dt + \int V_t dB_t; \tag{2.12}$$

with slight abuse of language, we call this the *stochastic integral generated by (U, V).*

Formally, the process X defined by (2.11) is an "antiderivative" of $U + V\dot{B}$, that is, a solution of the "stochastic differential equation"

$$\dot{X} = U + V\dot{B}, \tag{2.13}$$

or, in differential notation,

$$dX_t = U_t dt + V_t dB_t. \tag{2.14}$$

In the same sense, the stochastic integral (2.12) is the set of all "antiderivatives" (the "indefinite integral") of $U + V\dot{B}$, that is, the "general solution" of (2.13)/(2.14). The formal expression $U_t dt + V_t dB_t$ is called the stochastic differential generated by (U, V).

We note that the assumptions on V in Definition 2.3 are needed to guarantee the existence of the second integral in (2.11). They also guarantee that the process $\left(\int_0^t V_s dB_s\right)_{t\in\mathbb{R}_+}$ is $\mathcal{F}$-adapted and square-integrable in the sense that $\int_0^t V_s dB_s \in L^2(\Omega)$ for all $t \in \mathbb{R}_+$. The assumptions on U are stronger than necessary to guarantee the existence of the first integral in (2.11); in fact, $U^i \in L^1([0, t] \times \Omega)$ for all $1 \leq i \leq n$ and $t \in \mathbb{R}_+$ would be sufficient.

However, the stronger assumptions on U guarantees that the process $\left(\int_0^t U_s ds\right)_{t\in\mathbb{R}_+}$ is $\mathcal{F}$-adapted and square-integrable in the sense that $\int_0^t U_s ds \in L^2(\Omega)$ for all $t \in \mathbb{R}_+$. Indeed, we have the following lemma.

Lemma 2.2. *Under the assumptions of Definition 2.3, the processes* $\left(\int_0^t U_s ds\right)_{t\in\mathbb{R}_+}$ *and* $\left(\int_0^t V_s dB_s\right)_{t\in\mathbb{R}_+}$ *are well-defined,* $\mathcal{F}$*-adapted and square-integrable in the sense that* $\int_0^t U_s ds$, $\int_0^t V_s dB_s \in L^2(\Omega)$ *for all* $t \in \mathbb{R}_+$.

Corollary 2.1. *Assume the hypotheses of Definition 2.3 with* $\mathcal{F} = \mathcal{F}(B)$ *and let* $X^0 \in L^2(\Omega)$. *Then the process* X *defined by (2.11) is* $\mathcal{F}(B, X^0)$*-adapted and square-integrable in the sense that* $X_t \in L^2(\Omega)$ *for all* $t \in \mathbb{R}_+$.

Let us return to the integral equation (2.10), that is, the integral version of the "stochastic differential equation"

$$dX_t = U(X_t)dt + V(X_t)dB_t. \tag{2.15}$$

It is natural to seek a solution X of (2.15) subject to an initial condition of the form

$$X_0 = X^0, \tag{2.16}$$

where X^0 is a given $\mathbb{R}^n$-valued random variable on Ω. The integral equation corresponding to (2.15/(2.16)) reads

$$X_t = X^0 + \int_0^t U(X_s)ds + \int_0^t V(X_s)dB_s \tag{2.17}$$

Due to the random initial condition, a solution of (2.17) cannot be expected to be $\mathcal{F}(B)$-adapted, but should be $\mathcal{F}(B, X^0)$-adapted. The same would then hold for $U(X)$ and $V(X)$. However, under this assumption, the Ito integral in (2.17) is defined only if each component of B is a martingale with respect to $\mathcal{F}(B, X^0)$. This is the case if X^0 and B are *independent*; this follows from an argument similar to (2.8) and is reasonable intuitively, as we expect that the randomness of the initial condition should have nothing to do with an arbitrarily given Brownian motion. Along these lines, note that if X were only $\mathcal{F}(B)$-adapted, then X^0 being independent of B would force X^0 to be a.s. constant!

The above consideration motivates the following version of Corollary (2.1).

Corollary 2.2. *Assume the hypotheses of Definition 2.3 with $\mathcal{F} := \mathcal{F}(B, X^0)$, where $X^0 \in L^2(\Omega)$ is independent of B. Then the process X defined by (2.11) is $\mathcal{F}(B, X^0)$-adapted and square-integrable in the sense that $X_t \in L^2(\Omega)$ for all $t \in \mathbb{R}_+$.*

Now we move to the next section, where we formally define stochastic differential equations, define the solution to a stochastic differential equation, and discuss the existence and uniqueness of solutions.

2.2 Stochastic Differential Equations and their Solutions

As we discussed in the previous section, (2.13) or (2.14) is a stochastic analog of the deterministic antidifferentiation problem $dx = f(t)dt$ or $\dot{x} = f(t)$, where $f : \mathbb{R}_+ \to \mathbb{R}^n$ is a given, sufficiently regular function. To arrive at the stochastic analog of $dx = f(t, x)dt$ or $\dot{x} = f(t, x)$, where $f : \mathbb{R}_+ \times \mathbb{R}^n \to \mathbb{R}^n$ is a given, sufficiently regular function, we need to discuss the composition of stochastic processes.

Let $(\Omega, \mathcal{A})$, $(S, \mathcal{S})$, and $(S', \mathcal{S}')$ be measurable spaces, let H be a continuous-time S-valued stochastic process on $(\Omega, \mathcal{A})$ and let G be a continuous-time S'-valued stochastic process on $(S, \mathcal{S})$. We now (with slight abuse of notation) define the composition of G with H, denoted by $G \circ H$.

Definition 2.4. *For G and H as above, we define the composition $G \circ H$ to be the process defined by $(G \circ H)_t := G_t \circ H_t$, for all $t \in \mathbb{R}_+$.*

In this way, $G \circ H$ is a continuous-time S'-valued stochastic process on $(\Omega, \mathcal{A})$.

Now, if X is a stochastic integral, then X is a continuous-time $\mathbb{R}^n$-valued stochastic process on $(\Omega, \mathcal{A}, P)$. So, take $U : \mathbb{R}_+ \times \mathbb{R}^n \to \mathbb{R}^n$ to be measurable with respect to the second variable (so U is a continuous-time $\mathbb{R}^n$-valued stochastic process on $(\mathbb{R}^n, \mathcal{B}^n)$). Then $U \circ X$ is an $\mathbb{R}^n$-valued process on Ω, and we have

$$(U \circ X)(t, \omega) = (U_t \circ X_t)(\omega) = U_t(X_t(\omega)) = U_t(X(t, \omega)) = U(t, X(t, \omega)),$$

for all $t \in \mathbb{R}_+$ and $\omega \in \Omega$.

Similarly, take $V : \mathbb{R}_+ \times \mathbb{R}^n \to \mathbb{R}^{n \times d}$ to be measurable with respect to the second variable. Then $V \circ X$ is an $\mathbb{R}^{n \times d}$-valued process on Ω, and at least formally, we can consider the stochastic differential equation

$$dX_t = U(t, X_t)dt + V(t, X_t)dB_t, \tag{2.18}$$

or the equivalent integral equation

$$X_t = X_0 + \int_0^t U(s, X_s)ds + \int_0^t V(s, X_s)dB_s. \tag{2.19}$$

This motivates the definition of solution to a stochastic differential equation.

Definition 2.5. *We say X is a solution to (2.18) if X is any continuous-time $\mathbb{R}^n$-valued process X on Ω such that $(U \circ X, V \circ X)$ satisfy the hypotheses of Definition 2.3 and X satisfies (2.19) for all $t \in \mathbb{R}_+$.*

We remark that this definition makes sense due to Lemma 2.2. We will soon give particular conditions on U and V in order to guarantee that a unique solution to (2.18) exists. For now, assume that U and V are appropriate enough for (2.18) to make sense. We can now rigorously impose an initial condition to (2.18), state the definition of the stochastic initial value problem, and define the notion of solution.

Let X be a solution to (2.18) as above, and suppose we are given an $\mathbb{R}^n$-valued random variable $X^0 = (X_0^1, X_0^2, \cdots, X_0^n)$ such that $X_0^i \in L^2(\Omega)$ for each $1 \leq i \leq n$ and such that X^0 is independent of B. Recalling the argument preceding Corollary 2.2, if $X_0 = X^0$ a.s. then we specify $\mathcal{F}$ to be $\mathcal{F}(X^0, B)$, where $\mathcal{F}(X^0, B) = \{\mathcal{F}_t(X^0, B)\}_{t \in \mathbb{R}_+}$ and where $\mathcal{F}_t(X^0, B)$ is the σ-algebra generated by X^0 and $\{B_s | s \leq t\}$, for every $t \in \mathbb{R}_+$. Motivated by this and Corollary 2.2, the following definition is justified.

Definition 2.6. *Given $X^0 \in L^2(\Omega)$, independent of B, we call*

$$dX_t = U(t, X_t)dt + V(t, X_t)dB_t, \tag{2.20}$$

$$X_0 = X^0 \quad a.s.$$

a (strong) stochastic initial value problem, and we say X is a (strong) solution to (2.20) if X is a solution to $dX_t = U(t, X_t)dt + V(t, X_t)dB_t$ in the sense of Definition 2.5 and satisfies $X_0 = X^0$ a.s.

In Problem (2.20) the Brownian motion B is given in advance and we are seeking a solution X. A "weak" version of (2.20) would require that, along with X, we find a Brownian motion B on a probability space $(\Omega, \mathcal{A}, P)$ and a filtration $\mathcal{F}$ such that each component of B is a martingale with respect to $\mathcal{F}$. Then, since $(\Omega, \mathcal{A}, P)$ is not given, we cannot impose an initial condition as in (2.20), but we can impose an initial distribution μ. This leads to the following definition.

Definition 2.7. *We call*

$$dX_t = U(t, X)dt + V(t, X)dB_t, \tag{2.21}$$

$$P^{X_0} = \mu \quad a.s.$$

a weak stochastic initial value problem, and we say $(X, B, \mathcal{F})$ is a weak solution to (2.21) if B is an $\mathbb{R}^d$-valued Brownian motion such that B^i is a martingale with respect to $\mathcal{F}$ for all $1 \leq i \leq d$, and X is a solution to $dX_t = U(t, X)dt + V(t, X)dB_t$ in the sense of Definition 2.5 and satisfies $\mu = P^{X_0}$.

Clearly, a strong initial value problem induces a weak initial value problem (by replacing the given initial condition with its distribution and then removing the given probability space and Brownian motion); if that strong initial value problem has a solution then clearly the induced weak initial value problem must also have a solution. Also, if a weak initial value problem has a solution, then there is at least one associated strong initial value problem (by taking the Brownian motion B in the weak problem's solution as the given Brownian motion in the strong problem and constructing a random variable X^0 over B's accompanying probability space such that X^0 has distribution μ). Further, the existence

of a weak solution X to a weak initial value problem does not necessarily imply that X is a strong solution to the induced strong initial value problem; this is believable simply because X may be adapted to some filtration $\mathcal{F}$ but not to $\mathcal{F}(X^0, B)$.

From a modeling perspective, we really have weak problems (as no one can realistically present up front the specific representation of the white noise involved). Weak solutions are also useful because there are examples of weak initial value problems which have a weak solution but no strong solutions (see [10, pp. 301-302]). We drop the adjective weak or strong when there is no ambiguity.

Along these lines, there are also strong and weak notions of uniqueness.

Definition 2.8. *We say that the strong initial value problem* (2.20) *has the strong uniqueness property if any two solutions X and $\tilde{X}$ are modifications of each other.*

For convenience we often say that X is a strongly unique solution, or that X is strongly unique. Strong uniqueness is often called pathwise uniqueness.

Definition 2.9. *We say that the weak initial value problem* (2.21) *has the weak uniqueness property if any two solutions $(X, B, \mathcal{F})$ and $(\tilde{X}, \tilde{B}, \tilde{\mathcal{F}})$ are equivalent in the sense that $P^X = P^{\tilde{X}}$.*

Again, for convenience, we often say that X is a weakly unique solution or X is weakly unique, and since we may identify any weak solution with its (unique) distribution, we sometimes call P^X the weak solution. Weak uniqueness is often called uniqueness in distribution.

Analogously, we can have an initial value problem starting at any time $s > 0$.

Now we have the ingredients to present an existence and uniqueness theorem, which gives us at least one way to place conditions on U and V to guarantee the situation of Definition 2.5 is satisfied (one proof can be found in [12, Theorem 5.5]):

Theorem 2.2 (Existence/Uniqueness)**.** *Let* $(\Omega, \mathcal{A}, P)$ *be a measure space,* $n \in \mathbb{N}$*,* $d \in \{1, \ldots, n\}$*,* B *a* d*-dimensional Brownian motion on* Ω*,* $U : \mathbb{R}_+ \times \mathbb{R}^n \to \mathbb{R}^n$ *and* $V : \mathbb{R}_+ \times \mathbb{R}^n \to \mathbb{R}^{n \times d}$ *measurable functions,* $X^0 \in L^2(\Omega)$*,* X^0 *independent of* B*,* $\mathcal{F} = \mathcal{F}(B, X^0)$*.*

Assume that there exist positive constants C *and* D *such that, for all* $t \in \mathbb{R}_+$ *and* $x, y \in \mathbb{R}^n$*,*

$$|U(t,x)| + |V(t,x)| \leq C(1 + |x|), \tag{2.22}$$

where $|\cdot|$ *is the Euclidean norm, and*

$$|U(t,x) - U(t,y)| + |V(t,x) - V(t,y)| \leq D|x - y|. \tag{2.23}$$

Then the initial value problem (2.20) *has a strongly unique strong solution* X*.*

Before proving this important theorem, we remark that (2.22) is imposed to avoid that X explodes, i.e., that there is a finite time T_0 such that $P(\lim_{t \to T_0} |X_t(\cdot)| = \infty) > 0$ (see [9, Lemma 21.6]) while (2.23) ensures uniqueness. Compare this to the deterministic case, where an at most linear growth estimate insures that solutions do not explode (see [1, Theorem 7.6]) and a Lipschitz condition guarantees uniqueness. The idea of the proof is similar to the deterministic case; let us only consider the scalar case.

Proof. First, we show uniqueness. Suppose two solutions X and Y exist, having initial values X^0 and Y^0, respectively. Then we can estimate $E|X_t - Y_t|^2$ for a fixed t by using the inequality $(x+y+z)^2 \leq 3(x^2+y^2+z^2)$, the Ito isometry, the Cauchy-Schwarz inequality, and (2.23) so that Gronwall's inequality applies. This yields an inequality of the form

$$E|X_t - Y_t|^2 \leq 3E|X^0 - Y^0|^2 e^{Kt},$$

where K is a constant depending only on D and T. Assuming $X^0 = Y^0$ then implies that $P(|X_t - Y_t| = 0) = 1$ (recall t is fixed). We can repeat this argument for all rational t and then use that stochastic integrals are continuous in time to obtain $P(\bigcup_{t\in[0,T]} |X_t - Y_t| = 0) = 1$, which means X and Y are modifications of each other. This shows the strong uniqueness.

To show existence, first define the iterations $Y_t^{(0)} := X_0$ and

$$Y_t^{(k+1)} := X_0 + \int_0^t U(s, Y_s^{(k)})ds + \int_0^t V(s, Y_s^{(k)})dB_s, \tag{2.24}$$

for $k \in \mathbb{Z}_+$. We claim that (2.24) is well defined; first note that $Y_t^{(k)}$ is $\mathcal{F}_t(X^0, B)$-measurable for each $k \in \mathbb{Z}_+$ and for all $t \in [0,T]$. Next, we have by a similar calculation to that in the uniqueness proof, by (2.22), and by Fubini, that

$$E|Y_t^{(1)} - Y_t^{(0)}|^2 \leq 2C^2(t+t^2)(1+E|X_0|^2) \leq L_1 t, \tag{2.25}$$

where L_1 only depends on C, T, and $E|X_0|^2$. Therefore (2.24) makes sense for $k = 1$. One can proceed by induction to show that (2.24) makes sense for all k; we can estimate

$E|Y_t^{(2)} - Y_t^{(1)}|^2$ similarly and use (2.23) to yield an inequality of the form

$$E|Y_t^{(2)} - Y_t^{(1)}|^2 \leq 3(1+T)D^2 \int_0^t E|Y_s^{(1)} - Y_s^{(0)}|^2 ds \tag{2.26}$$

$$\leq 3(1+T)D^2 \int_0^t L_1 s \; ds \leq L_2 \frac{t^2}{2},$$

where L_2 only depends on C, D, T, and $E|X_0|^2$. Iterating this, we can estimate $E|Y_t^{(k+1)} - Y_t^{(k)}|^2$ similarly:

$$E|Y_t^{(k+1)} - Y_t^{(k)}|^2 \leq 3(1+T)D^2 \int_0^t E|Y_s^{(k)} - Y_s^{(k-1)}|^2 ds \leq \frac{L_{k+1} t^{k+1}}{(k+1)!} \tag{2.27}$$

where L_{k+1} is a constant depending only on T, C, D, and $E|X_0|^2$; in fact, $L_{k+1} = L_1(3(1+T)D^2)^k$, for $k \in \mathbb{Z}_+$. Since $t \leq T$, this inequality shows $\{Y_t^{(k)}\}_{k\in\mathbb{Z}_+}$ is a Cauchy sequence in $L^2([0,T]\times\Omega)$, so for every $t \in [0,T]$, $\{Y_t^{(k)}\}_{k\in\mathbb{Z}_+}$ has a $\mathcal{F}_t(X^0, B)$-measurable limit X_t. In fact, this convergence is uniform; from (2.27), we apply the inequality (see [6, Theorem 2.8])

$$P(\sup_{[a,b]} |\int_a^b f(s) dB_s| > r) \leq \frac{1}{r^2} E(\int_a^b f^2(s) ds)$$

yielding

$$P(\sup_{[0,T]} |Y_t^{k+1} - Y_t^k| > \frac{1}{k^2}) \leq \frac{L_{k+1} t^{k+1}}{k+1!} k^4.$$

Since $\sum_{k=1}^{\infty} \frac{L_{k+1}t^{k+1}}{k+1!}k^4$ converges, by the Borel-Cantelli Lemma (see e.g. [6, Theorem 1.1]) there exists a sufficiently large $M \in \mathbb{N}$ such that, for all $m \geq M$,

$$P(\sup_{[0,T]} |Y_t^{m+1} - Y_t^m| > \frac{1}{m^2}) = 0.$$

Therefore the convergence of $\{Y_t^{(k)}\}_{k\in\mathbb{Z}_+}$ to X_t is uniform, which means

$$X_t = X_0 + \lim_{k\to\infty}(\int_0^t U(s, Y_s^{(k)}) + \int_0^t V(s, Y_s^{(k)})) = X_0 + \int_0^t U(s, X_s) + \int_0^t V(s, X_s)dB_s,$$

so X is indeed a solution. □

Unless we say otherwise, we assume (2.22) and (2.23) hold when we discuss stochastic differential equations.

As we discussed before, a strong initial value problem induces a weak initial value problem; it can be shown that if a strong initial value problem enjoys the strong uniqueness property, then the strong initial value problem and its induced weak initial value problem have the weak uniqueness property (see e.g. [10, pp. 306-310]).

We will soon discuss how deterministic dynamical systems generalize to the stochastic case, but before this, we reserve the next section for Ito's formula, which allows us to calculate specific examples of Ito integrals and hence solutions to stochastic differential equations.

2.3 Ito's Formula and Examples

Equipped with Ito integral and stochastic differential equation concepts, we now focus on explicitly calculating Ito integrals and solutions to stochastic differential equations. The tool we need is Ito's formula, which is essentially a stochastic analog of the chain rule.

First, we prove Ito's formula in one dimension, and then we present the n-dimensional version and study some examples. Let X be the stochastic integral generated by (U, V) (for U, V satisfying the assumptions in Defintion 2.3) and let $g \in C^2(\mathbb{R}_+ \times \mathbb{R})$. Then the process $(g(t, X_t))_{t \in \mathbb{R}_+}$ is also a 1-dimensional stochastic integral, and for all t,

$$g(t, X_t) = g(0, X_0) + \int_0^t \Big(\frac{\partial g}{\partial s}(s, X_s) + U_s \frac{\partial g}{\partial x}(s, X_s) + \frac{1}{2} V_s^2 \frac{\partial^2 g}{\partial x^2}(s, X_s) \Big) ds + \int_0^t V_s \frac{\partial g}{\partial x}(s, X_s) dB_s,$$

which we call Ito's formula.

Notice the "extra" term $\int_0^t \frac{1}{2} V_s^2 \frac{\partial^2 g}{\partial x^2} ds$; such a term is often called a "correction term." We can, in fact, recover the "natural" form of the chain rule by using the Stratonovich integral (which differs from the Ito integral in that it uses the midpoint instead of the left endpoint), but Stratonovich integrals "look into the future" and (among other things) do not enjoy the martingale property.

To see where this extra term comes from, let us examine a Taylor expansion of $g(t, X_t)$; it is enough to assume that g, $\frac{\partial g}{\partial t}$, $\frac{\partial g}{\partial x}$, and $\frac{\partial^2 g}{\partial x^2}$ are bounded, for if we can prove it in this case, then we can take sequences of bounded functions g_n, $\frac{\partial g_n}{\partial t}$, $\frac{\partial g_n}{\partial x}$, and $\frac{\partial^2 g_n}{\partial x^2}$ to uniformly approximate a C^2 function g and $\frac{\partial g}{\partial t}$, $\frac{\partial g}{\partial x}$, and $\frac{\partial^2 g}{\partial x^2}$, respectively on compact subsets of $\mathbb{R}_+ \times \mathbb{R}$ (by Stone-Weierstrass) and then the uniform convergence allows the limit to carry through

the integral (for the stochastic term, use Ito's isometry). Recall that the norm of any partition P of $[0,t]$ is defined to be $\| P \| = \max_{1\leq i\leq n}(t_i - t_{i-1})$, and let $P = (t_j)_{j=0}^n$ be a partition of $[0,t]$ with sufficiently small norm. Then, carrying out a Taylor expansion of $g(t, X_t)$,

$$\begin{aligned} g(t,X_t) &= g(0,X_0) + \sum_{j=1}^{n} \big(g(t_j, X_{t_j}) - g(t_{j-1}, X_{t_{j-1}})\big) \\ &= g(0,X_0) + \sum_{j=1}^{n} \bigg(\frac{\partial g}{\partial t}(t_{j-1}, X_{t_{j-1}})(t_j - t_{j-1}) \\ &+ \frac{\partial g}{\partial x}(t_{j-1}, X_{t_{j-1}})(X_{t_j} - X_{t_{j-1}}) \bigg) + \frac{1}{2}\sum_{j=1}^{n} \bigg(\frac{\partial^2 g}{\partial t^2}(t_{j-1}, X_{t_{j-1}})(t_j - t_{j-1})^2 \\ &+ 2\frac{\partial^2 g}{\partial x \partial t}(t_{j-1}, X_{t_{j-1}})(X_{t_j} - X_{t_{j-1}})(t_j - t_{j-1}) \\ &+ \frac{\partial^2 g}{\partial x^2}(t_{j-1}, X_{t_{j-1}})(X_{t_j} - X_{t_{j-1}})^2 \bigg) + \sum_{j=1}^{n} R_j, \end{aligned}$$

where R_j in the remainder $\sum_{j=1}^n R_j$ takes the form

$$R_j = \sum_{\{|\alpha|\geq 3\}} \frac{\partial^{\alpha_1}}{\partial t^{\alpha_1}} \frac{\partial^{\alpha_2}}{\partial x^{\alpha_2}} \frac{g(t_{j-1}, X_{t_{j-1}})}{\alpha_1! \alpha_2!} (t_j - t_{j-1})^{\alpha_1} (X_{t_j} - X_{t_{j-1}})^{\alpha_2}.$$

Let us approximate each term, using that the norm of the partition is small. For the first-order terms, we see

$$\sum_j \frac{\partial g}{\partial t}(t_j - t_{j-1}) \approx \int_0^t \frac{\partial g}{\partial t}(s, X_s) ds,$$

and

$$\sum_j \frac{\partial g}{\partial x}(X_{t_j} - X_{t_{j-1}}) \approx \int_0^t \frac{\partial g}{\partial x}(s, X_s) dX_s,$$

where

$$\int_0^t \frac{\partial g}{\partial x}(s, X_s) dX_s = \int_0^t \frac{\partial g}{\partial x}(s, X_s) U(s, X_s) ds + \int_0^t \frac{\partial g}{\partial x}(s, X_s) V(s, X_s) dB_s,$$

For the second order terms, only the term $\sum_{j=1}^n \frac{\partial^2 g}{\partial x^2}(t_{j-1}, X_{t_{j-1}})(X_{t_j} - X_{t_{j-1}})^2$ is not approximately zero. We expand this term:

$$\sum_{j=1}^n \frac{\partial^2 g}{\partial x^2}\Big((U_{t_j} - U_{t_{j-1}})^2 (t_j - t_{j-1})^2 + (V_{t_j} - V_{t_{j-1}})^2 (B_{t_j} - B_{t_{j-1}})^2 + \tag{2.28}$$

$$(U_{t_j} - U_{t_{j-1}})(V_{t_j} - V_{t_{j-1}})(t_j - t_{j-1})(B_{t_j} - B_{t_{j-1}})\Big).$$

We claim that (2.28) has only one term that is not approximately zero, namely, $\sum_{j=1}^n \frac{\partial^2 g}{\partial x^2}(V_{t_j} - V_{t_{j-1}})^2 (B_{t_j} - B_{t_{j-1}})^2$, and it satisfies

$$\| \sum_{j=1}^n \frac{\partial^2 g}{\partial x^2}(V_{t_j} - V_{t_{j-1}})^2 (B_{t_j} - B_{t_{j-1}})^2 - \int_0^t \frac{\partial^2 g}{\partial x^2} V_s^2 ds \|_{L^2} \approx 0. \tag{2.29}$$

For details on how to prove (2.29), see [12, p.32]; we present a similar and more transparent argument to clearly convey the essence of Ito's formula. To this end, we now show that

$$\lim_{\|P\| \to 0} \| \sum_{j=1}^n [(B_{t_j} - B_{t_{j-1}})^2 - (t_j - t_{j-1})] \|_{L^2} = 0. \tag{2.30}$$

To see this, call $M_j = (B_{t_j} - B_{t_{j-1}})^2 - (t_j - t_{j-1})$. Then $EM_j = 0$ because $E(B_{t_j} - B_{t_{j-1}})^2 = t_j - t_{j-1}$. Further, $E(M_j)^2 = 2(t_j - t_{j-1})$ because

$$\begin{aligned} E(M_j)^2 &= E((B_{t_j} - B_{t_{j-1}})^4 - (B_{t_j} - B_{t_{j-1}})^2(t_j - t_{j-1}) + (t_j - t_{j-1})^2) \\ &= 3(t_j - t_{j-1})^2 - 2(t_j - t_{j-1})^2 + (t_j - t_{j-1})^2, \end{aligned}$$

by (1.5) (in the section on Brownian motion).

Now, since Brownian motion has independent increments, each M_j is independent, which means

$$E(\sum_{j=1}^{n} M_j)^2 = E\sum_{j=1}^{n}(M_j)^2 = \sum_{j=1}^{n} 2(t_j - t_{j-1})^2,$$

which clearly goes to zero as the norm of the partition goes to zero.

So, we have taken the Taylor expansion of $g(t, X_t)$ and approximated each term when the norm of the partition is small; combining all of them yields Ito's formula, as claimed. Using differential notation, we can express (2.30) as "$(dB_t)^2 = dt$". Also, we may write "$(dt)^2 = dtdB_t = 0$," since terms containing $(t_j - t_{j_1})^2$ or $(t_j - t_{j_1})(B_{t_j} - B_{t_{j_1}})$ go to zero when the norm of the partition goes to zero; thus we could write

$$\text{“}(dX_t)^2 = U^2(dt)^2 + UVdtdB_t + V^2(dB_t)^2 = V^2dt.\text{”}$$

We could then rewrite Ito's formula more conveniently:

$$dg(t, X_t) = \frac{\partial g}{\partial t}(t, X_t)dt + \frac{\partial g}{\partial x_i}(t, X_t)dX_t + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}(t, X_t)(dX_t)^2. \tag{2.31}$$

We may extend this result to n-dimensions: taking X to be an n-dimensional stochastic integral and taking $g \in C^2(\mathbb{R}_+ \times \mathbb{R}^n, \mathbb{R}^n)$, where $X = (X^i)_{i=1}^n$ and $g = (g^i)_{i=1}^n$, then $(g(t, X_t))_{t\in\mathbb{R}_+}$ is an n-dimensional stochastic integral, and for all $t \in \mathbb{R}_+$,

$$dg^k(t, X_t) = \frac{\partial g^k}{\partial t}(t, X_t)dt + \sum_{i=1}^{n} \frac{\partial g^k}{\partial x_i}(t, X_t)dX_t^i \tag{2.32}$$

$$+ \sum_{i,j=1}^{n} \frac{1}{2} \frac{\partial^2 g^k}{\partial x_i \partial x_j}(t, X_t)dX_t^i dX_t^j,$$

for $1 \leq k \leq n$ (understanding $dB_t^i dB_t^j = \delta_{ij}dt$ and $dtdB_t^i = 0$).

We remark that Ito's formula can be stated and proved in higher generality using martingale theory (as in ([9, Theorem 17.18])), but we do not need such a level of abstraction in the specialized environment of stochastic differential equations.

Now for some examples; let B be a real-valued Brownian motion on a probability space $(\Omega, \mathcal{A}, P)$. Also, let $X_0 = x$ a.s. always be the initial condition (for some $x \in \mathbb{R}$).

To start, let us observe an integration by parts formula: for $f \in C^1(\mathbb{R}_+)$,

$$\int_0^t f(s)dB_s = f(t)B_t - \int_0^t B_s df(s), \tag{2.33}$$

for all $t \in \mathbb{R}_+$. This follows from Ito's formula; we think that $\int_0^t f(s)dB_s$ should somehow yield a "$f(t)B_t$" sort of term, so we take $g(t, x) = f(t)x$. Then it is clear that $\frac{\partial}{\partial t}g(t, x) = f'(t)x$, $\frac{\partial}{\partial x}g(t, x) = f(t)$, and $\frac{\partial^2}{\partial x^2}g(t, x) = 0$. Since $g(t, B_t) = f(t)B_t$, by Ito's formula,

$$d(g(t, B_t)) = d(f(t)B_t) = f'(t)B_t dt + f(t)dB_t + 0,$$

which easily yields (2.33). Note that f only depends on time, and note that (similarly to the above argument)

$$\int_0^t f(s)dX_s = (f(t)X_t - f(0)X_0) - \int_0^t X_s df(s) \tag{2.34}$$

holds for any 1-dimensional stochastic integral X (for a more general result, see [9, Theorem 17.16]).

Along these lines, consider

$$dX_t = dB_t - X_t dt. \tag{2.35}$$

Calling $g(t,x) = e^t x$ and using Ito's formula yields

$$d(e^t X_t) = e^t dX_t + e^t X_t dt = e^t dB_t,$$

and it is easy to see that the solution is

$$X_t = e^{-t}x + \int_0^t e^{(s-t)} dB_s.$$

Similarly,

$$dX_t = \sigma dB_t - bX_t dt,$$

where σ, b are constants, has solution

$$X_t = e^{-bt}x + \sigma \int_0^t e^{b(s-t)} dB_s.$$

Next, we have a basic example of an Ito integral calculation:

$$\int_0^t B_s dB_s = \frac{1}{2}(B_t^2 - t). \tag{2.36}$$

We expect the $\frac{1}{2}B_t^2$ term from deterministic calculus, but we inherit the extra $\frac{1}{2}t$ term from the stochastic case (often called a correction term). To use Ito's formula, take $g(t,x) = \frac{1}{2}x^2$ and $X_t = B_t$, which yields

$$dg(t, B_t) = B_t dB_t + \frac{1}{2}(dB_t)^2,$$

which means

$$\frac{1}{2}dB_t^2 = \int_0^t B_s dB_s + \frac{1}{2}\int_0^t ds,$$

and this easily reduces to (2.36).

This example shows how the realization $(dB_t)^2 = dt$ helps calculate Ito integrals. Notice that the integral of a purely random process ends up having a deterministic part; for deeper insight along these lines, see [9, pp. 339-340].

Along these lines, let us solve

$$dX_t = bX_t dt + \sigma X_t dB_t, \tag{2.37}$$

where σ and b are positive constants. The noiseless case would produce the solution $X_t = e^{bX_t}$, so we may expect something like this but with a correction term. Since (2.37) can be interpreted as

$$\int_0^t \frac{dX_s}{X_s} = bt + \sigma B_t,$$

we apply Ito's formula (using that $(dt)^2 = 0$):

$$d(\ln(X_t)) = \frac{dX_t}{X_t} - \frac{(dX_t)^2}{2X_t^2} = \frac{dX_t}{X_t} - \frac{\sigma^2 X_t^2 dt}{2X_t^2} = \frac{dX_t}{X_t} - \frac{1}{2}\sigma^2 dt.$$

Thus, solving for $\frac{dX_t}{X_t}$ and integrating, we get

$$bt + \sigma B_t = \ln(\frac{X_t}{x}) + \frac{1}{2}\sigma^2 t,$$

or

$$X_t = xe^{(b-\frac{1}{2}\sigma^2)t+\sigma B_t}.$$

Notice that we recover the proper solution for the noiseless case when $\sigma := 0$, and that we have solved the one-dimensional equation $\dot{X} = (r(X) + W)X$ when $r(X)$ is constant.

For the next example we study the logistics equation

$$dX_t = (aX_t - b(X_t)^2)dt + \sigma X_t dB_t, \tag{2.38}$$

for a, b, and σ positive constants. We remark that only a is perturbed; if b is perturbed, then the probability that solutions do not explode in finite time is zero (except, of course, for the trivial solution $x = 0$; see [2, p. 99]).

To solve (2.38), substitute $Y_t = \frac{-1}{X_t}$ to get

$$dY_t = (-aY_t - b)dt - \sigma Y_t dB_t,$$

which is solved in a similar fashion to (2.37):

$$Y_t = e^{-[(a-\frac{1}{2}\sigma^2)t+\sigma dB_t]}[-\frac{1}{x} - b\int_0^t e^{(a-\frac{1}{2}\sigma^2)s+\sigma dB_s}ds].$$

Transforming back to X_t gives

$$X_t = \frac{xe^{(a-\frac{1}{2}\sigma^2)t+\sigma B_t}}{1 + xb\int_0^t e^{(a-\frac{1}{2}\sigma^2)s+\sigma dB_s}ds}.$$

Observe that in the noiseless case one yields the familiar elementary solution

$$X_t = \frac{xK}{x - (K-x)e^{-at}},$$

for $K = \frac{b}{a}$ and all $t \in \mathbb{R}_+$.

Finally, let us move to a two-dimensional system; now let B be an $\mathbb{R}^2$-valued Brownian motion on a probability space $(\Omega, \mathcal{A}, P)$. We will "reverse engineer" the classical system where

$$dx_1 = -x_2 dt$$

$$dx_2 = x_1 dt,$$

with initial value $x(0) = (x_1(0), x_2(0)) = (1, 0) \in \mathbb{R}^2$. It is well-known that the solution is $x(t) = (\cos(t), \sin(t)) = e^{it}$. Now pick $g(t, x) = e^{ix}$, so that

$$g(t, B) = e^{iB} = (\cos B, \sin B) := (X^1, X^2).$$

Then by Ito's formula, we see

$$dX_t^1 = -\sin(B_t)dB_t - \frac{1}{2}\cos(B_t)dt = -X^2 dB_t - \frac{1}{2}X^1 dt$$

$$dX_t^2 = \cos(B_t)dB_t - \frac{1}{2}\sin(B_t)dt = X^1 dB_t + \frac{1}{2}X^2 dt$$

is the stochastic system (with initial condition $X_0 = (1, 0)$ a.s.) whose solution is a one-dimensional Brownian motion traveling around the unit circle.

Chapter 3

Dynamical Systems and Stochastic Stability

3.1 "Stochastic Dynamical Systems"

In this section we present an overview of dynamical systems and their "stochastic" analogies. Let X be a metric space and let $S : D(S) \subset \mathbb{R} \times X \to X$. For $t \in \mathbb{R}$, let $S_t := S(t, \cdot)$.

Definition 3.1. *We say S is a local dynamical system or local flow if $D(S)$ is open in $\mathbb{R} \times X$, S is continuous and*

i) $S_0 = id_S$, and

ii) $S_{t+s} = S_t \circ S_s$ for all $t, s \in \mathbb{R}$.

We say S is a global flow if S is a local flow with $D(S) = \mathbb{R} \times X$. We say S is a local (global) semiflow if the above conditions hold with $\mathbb{R}$ replaced by $\mathbb{R}_+$.

Definition 3.2. *We say that a set $A \subset X$ is positively invariant for a local semiflow S if $S_t(A \cap D(S_t)) \subset A$ for all $t \in \mathbb{R}_+$. We say $\overline{x}$ is an equilibrium point if, for every $t \in \mathbb{R}_+$, $\overline{x} \in D(S_t)$ and $S_t(\overline{x}) = \overline{x}$.*

Recall that autonomous ordinary differential equations generate flows: let $b \in C^1(\mathbb{R}^n, \mathbb{R}^n)$ and consider the system $\dot{u} = b(u)$. Given any initial value $x \in \mathbb{R}^n$ there exists a largest open interval of existence $I_x \subset \mathbb{R}$ containing 0 such that the system $\dot{u} = b(u)$ has a unique solution $u^x \in C^1(I_x, \mathbb{R}^n)$ with $u^x(0) = x$. The system $\dot{u} = b(u)$ generates a local solution flow $S : D(S) \subset \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n$ with $D(S) := \{(t, x) \in \mathbb{R} \times \mathbb{R}^n | t \in I_x\}$ where $S(t, x) := u^x(t)$ for all $(t, x) \in D(S)$; $D(S)$ is open and S is continuous by [1, Theorem

8.3], S is $C^1(D(S), \mathbb{R}^n)$ since $b \in C^1(\mathbb{R}^n, \mathbb{R}^n)$ by [1, Theorem 9.5], and the uniqueness of the solution guarantees the group property: $S_{t+s} = S_t S_s$ for all $s, t \in \mathbb{R}_+$. Similarly, we can obtain the local solution semiflow $S|_{D(S)\cap(\mathbb{R}_+\times\mathbb{R}^n)}$. Observe that $\frac{\partial}{\partial t}S(t,x) = b(S(t,x))$ for all $(t,x) \in D(S)$.

Assume $\dot{u} = b(u)$ generates a global solution flow S. Then $S_t \in C^1(\mathbb{R}^n, \mathbb{R}^n)$ for all $t \in \mathbb{R}$ and S_t is invertible with $S_t^{-1} = S_{-t} \in C^1(\mathbb{R}^n, \mathbb{R}^n)$. So, for each $t \in \mathbb{R}$, $S_t : \mathbb{R}^n \to \mathbb{R}^n$ is a C^1-diffeomorphism. This means that for any $A \in \mathcal{B}^n$ we have the change of variable formula

$$\int_{S_t^{-1}(A)} f(x)dx = \int_A f(S_t^{-1}(y))|\det(D(S_t^{-1}(y)))|dy \tag{3.1}$$

$$= \int_A f(S_{-t}(y))|\det(DS_{-t})(y)|dy$$

where $D(S_t^{-1})$ is the (nonzero) Jacobian matrix for S_t^{-1}. Thus, $\lambda(A) = 0$ implies $\lambda(S_t^{-1}(A)) = 0$ for all t; we shall soon see the importance of this property.

Analogously, compare the above to the autonomous case of stoachastic differential equations. For $(U, V) \in \mathcal{S}^{n,d}_{\mathcal{F},B}$, this amounts to making U and V constant in time; call $U_t := b$ for all $t \in \mathbb{R}_+$ and $V_t := \sigma$ for all $t \in \mathbb{R}_+$, where $b : \mathbb{R}^n \to \mathbb{R}^n$ is $(\mathcal{B}^n, \mathcal{B}^n)$-measurable and $\sigma : \mathbb{R}^n \to \mathbb{R}^{n\times d}$ is $(\mathcal{B}^n, \mathcal{B}^{n\times d})$-measurable. We remark that this may seem like a strange and sudden notation change, but it is quite common in the literature to use "σ, b" notation and so we adhere to this convention now, specially reserving it for the autonomous case (even though many authors use σ and b for more general cases).

We shall now see how solutions in the autonomous case enjoy some of the "nice dynamical systems properties" that we hope for. Considering only degenerate initial distributions,

we see that (2.20) takes the form

$$dX_t = b(X_t)dt + \sigma(X_t)dB_t, \tag{3.2}$$

$$X_0 = x \text{ a.s.},$$

with $x \in \mathbb{R}^n$. To emphasize the initial condition in the solution's expression, call $X^{0,x}$ the strong solution to (3.2), where B is given, and $X^{0,x}$ is $\mathcal{F}(B)$-adapted. Let us study the induced weak problem

$$dX_t = b(X_t)dt + \sigma(X_t)dB_t, \tag{3.3}$$

$$P^{X_0} = \epsilon_x.$$

We will now provide two weak solutions to (3.3) that are "time shifts" of each other and use the weak uniqueness property to conclude that they must have the same distribution.

Obviously, $(X^{0,x}, B, \mathcal{F}(B))$ is a weak solution to (3.3).

Now consider the process $(X^{s,x}_{t+s})_{t\in\mathbb{R}_+}$, that is, the shifted version of $X^{0,x}$ as above where $X^{s,x}_s = x$ a.s. We would like to shift back in time by s to solve (3.3); observe that one cannot simply shift the Brownian motion B in time without affecting the variance, so to "shift" B, we need to define a new Brownian motion with the appropriate distribution (this is why we must appeal to the weak problem!). So, define $\tilde{B}_t := B_{t+s} - B_s$ for all $t \geq 0$. Then $\tilde{B} = (\tilde{B}_t)_{t\in\mathbb{R}_+}$ is a Brownian motion starting at zero, has the same distribution as $(B_{t+s})_{t\in\mathbb{R}_+}$ by stationary increments, and $\tilde{\mathcal{F}}(\tilde{B})$ is defined by $\tilde{\mathcal{F}}_t(\tilde{B}) := \mathcal{F}_{t+s}(B)$ for all

$t \in \mathbb{R}_+$. Now, by definition of $X^{s,x}$ and shifting,

$$X^{s,x}_{t+s} = x + \int_s^{t+s} b(X^{s,x}_r)dr + \int_s^{t+s} \sigma(X^{s,x}_r)dB_r$$

$$= x + \int_0^{s} b(X^{s,x}_{r+s})dr + \int_0^{s} \sigma(X^{s,x}_{r+s})d\tilde{B}_r.$$

This means $((X^{s,x}_{t+s})_{t\in\mathbb{R}_+}, \tilde{B}, \tilde{\mathcal{F}}(\tilde{B}))$ is also a weak solution of (3.3). Thus, by the weak uniqueness,

$$(X^{0,x}_t)_{t\in\mathbb{R}_+} \sim^d (X^{s,x}_{t+s})_{t\in\mathbb{R}_+}, \tag{3.4}$$

which leads us to the following definition:

Definition 3.3. *We say a process X is time-homogeneous if X satisfies (3.4).*

We call this the diffusion case, often referring to the solution X as a diffusion; we think of x as a particle that would move with velocity b (b is sometimes called the "drift" coefficient) except that random collisions with other particles (say, the collisions occur with some kind of "intensity" σ, also called a "diffusion" coefficient) may cause interference. As we will see later, there is an intimate relationship between stochastic differential equations and second-order partial differential equations, which is one reason why the term "diffusion" is used.

Again, time-homogeneity has much to do with the "nice dynamical systems properties" we want; we may think of a diffusion as a "stochastic semiflow." In fact, in probability theory, a diffusion refers to a Markov process with continuous paths (with perhaps some extra properties); it therefore is not a surprise that a solution to a stochastic differential

equation with initial values that are a.s. constant will be a Markov process (see [12, Theorem 7.2]).

So what if the initial condition is a nondegenerate random variable? Then we have a semiflow action on a set of probability measures. More precisely, recall that $C_b(\mathbb{R}^n, \mathbb{R})$ is the set of bounded, continuous functions mapping $\mathbb{R}^n$ into $\mathbb{R}$, and call $\mathcal{M}_{\mathcal{B}^n}$ the set of all finite Borel measures. Equip $C_b(\mathbb{R}^n, \mathbb{R})$ with the the sup norm, that is, the norm $\|f\| = \sup_x |f(x)|$, to make it a Banach space. Then $\mathcal{M}_{\mathcal{B}^n}$ is a subset of $C_b^*(\mathbb{R}^n, \mathbb{R})$, the dual space of $C_b(\mathbb{R}^n, \mathbb{R})$, so if we equip $C_b^*(\mathbb{R}^n, \mathbb{R})$ with the weak* topology, $\mathcal{M}_{\mathcal{B}^n}$ inherits it. It can be shown that $\mathcal{M}_{\mathcal{B}^n}$ is metrizable as a complete metric space (see [13, p.371]), therefore a dynamical system could be defined over $\mathcal{M}_{\mathcal{B}^n}$. In fact, we are most interested in the (Lebesgue) absolutely continuous measures; this allows us the luxury of using semigroup theory to "fluctuate functions" rather than "fluctuate measures."

Now define a family $U = (U_t)_{t \geq 0}$ such that for each $t \in \mathbb{R}_+$, $U_t : \mathcal{M}_{\mathcal{B}^n} \to \mathcal{M}_{\mathcal{B}^n}$ by

$$U_t \mu := \int P_t(x, \cdot) \mu(dx). \tag{3.5}$$

Then U is the dual semigroup to T, as $\langle f, U_t \mu \rangle = \langle T_t f, \mu \rangle$ for all $f \in C_0(\mathbb{R}^n, \mathbb{R})$ and all $\mu \in C_0^*(\mathbb{R}^n, \mathbb{R})$, where $\langle f, \mu \rangle_{C_0(\mathbb{R}), C_0^*(\mathbb{R})} := \int f d\mu$. Note that U is a semidynamical system on $\mathcal{M}_{\mathcal{B}^n}$: $\mu \in \mathcal{M}_{\mathcal{B}^n}$ implies $U_t \mu \in \mathcal{M}_{\mathcal{B}^n}$ for every $t \in \mathbb{R}_+$. Also, for any $B \in \mathcal{M}_{\mathcal{B}^n}$, we have

$$U_0 \mu(B) = \int P_0(x, B) \mu(dx) = \int 1_B(x) \mu(dx) = \mu(B)$$

and

$$U_{s+t} \mu(B) = \int P_{s+t}(x, B) \mu(dx) = \int \left(\int P_t(y, B) P_s(x, dy) \right) \mu(dx) =$$

$$\int P_t(y,B)\int P_s(x,dy)\mu(dx) = \int P_t(y,B)U_s\mu(dy) = U_t(U_s\mu(B)).$$

All we lack is the continuity, the proof of which can be found in ([13, pp. 370-371]). Notice that if μ is a probability measure, then $U_t\mu$ is a probability measure for every $t \in \mathbb{R}_+$ since T is a contraction semigroup.

3.2 Koopman and Frobenius-Perron Operators: The Deterministic Case

In this section, we define the Koopman and Frobenius-Perron operators, which are useful in understanding how deterministic cases for differential equations extend to stochastic ones. Primarily, we are interested in describing the distribution of a (continuous-time $\mathbb{R}^n$-valued) solution process X of a stochastic differential equation via semigroup theory; in the case where the distributions P^{X_t} have densities for every t, one can represent the flow as a semigroup of linear operators on $L^1(\mathbb{R}^n)$ whose generator is a second-order differential operator on $\mathbb{R}^n$. This leads to the set-up of a partial differential equation called the Fokker-Planck equation, which describes the fluctuation of the densities of the distributions of X (assuming that the random variable X_t has a Lebesgue density for all $t \in \mathbb{R}_+$).

A brief outline of the procedure is as follows. First, we define the Koopman operator and the Frobenius-Perron operators. We show they are adjoint and derive the infintesimal generators for each in the case of a deterministic ordinary differential equation. We then make a stochastic generalization of these operators and mimic the deterministic case, employing stochastic calculus. Finally, we yield the form for the infintesimal generator of the semigroup describing the solution process X and use its adjoint to obtain the Fokker-Planck equation.

Given a measureable space $(X, \mathcal{F})$ equipped with a signed measure ρ, let $(Y, \mathcal{G}, \bar{\nu})$ be a (σ-finite) measure space, and let $S : X \to Y$ be measurable.

Definition 3.4. *We define the image measure of ρ under S by*

$$\rho^S(G) = \rho(S^{-1}(G)),$$

for any $G \in \mathcal{G}$.

A useful characterization which follows from the definition is

$$\int g d\rho^S = \int (g \circ S) d\rho, \tag{3.6}$$

for g nonnegative and measurable. Also, a measurable g is ρ^S-integrable iff $g \circ S$ is ρ-integrable, in which case (3.6) holds.

Definition 3.5. *S is nonsingular if $\bar{\nu}(G) = 0$ implies $\rho(S^{-1}(G)) = 0$ for every $G \in \mathcal{G}$.*

So, if S is nonsingular, then ρ^S is absolutely continuous with respect to $\bar{\nu}$ and thus, has a Radon-Nikodym derivative $\frac{d\rho^S}{d\bar{\nu}}$ (which is in $L^1(\bar{\nu}) := L^1(Y, \mathcal{G}, \bar{\nu})$ iff ρ^S is finite).

Now let $(X, \mathcal{F}, \bar{\mu})$ be another σ-finite measure space, and let $f \in L^1(\bar{\mu}) := L^1(X, \mathcal{F}, \bar{\mu})$, and recall that $\bar{\mu}_f(A) := \int_A f d\bar{\mu}$, for $A \in \mathcal{F}$.

We now define the Frobenius-Perron operator P (associated with S as above) by applying the image measure construction to the signed measure $\bar{\mu}_f$; denote this by $\bar{\mu}_f^S = (\bar{\mu}_f)^S$. Note that since f is $\bar{\mu}$-integrable and

$$\bar{\mu}_f^S(G) = \bar{\mu}_f(S^{-1}(G)) = \int_{S^{-1}(G)} f d\bar{\mu}$$

holds for any $G \in \mathcal{G}$, then $\bar{\mu}_f^S$ is a finite signed measure.

Definition 3.6. *The operator P, where $P : L^1(\bar{\mu}) \to L^1(\bar{\nu})$ is defined by*

$$Pf = \frac{d(\bar{\mu}_f^S)}{d\bar{\nu}},$$

is called the Frobenius-Perron operator (associated to S).

We obtain this by taking $f \in L^1(\bar{\mu})$, creating $\bar{\mu}_f$, associating it to $\bar{\mu}_f^S$, and using the nonsingularity of S to take the the Radon-Nikodym derivative with respect to $\bar{\nu}$. In another way, for $G \in \mathcal{G}$,

$$\bar{\mu}_f^S(G) = \bar{\mu}_f(S^{-1}(G)) = \int_{S^{-1}(G)} f d\bar{\mu} = \int_G Pf d\bar{\nu}.$$

In fact, what happens in general is that S is causing a change in measure, so one can think of $\mu \mapsto \mu^S$ as a mapping from $\mathcal{M}_{\mathcal{F}}$ into $\mathcal{M}_{\mathcal{G}}$, where $\mathcal{M}_{\mathcal{A}}$ denotes the set of all finite signed measures on a σ-algebra $\mathcal{A}$. Define, for a given measure γ_0 on $\mathcal{A}$,

$$\mathcal{M}_{\mathcal{A}}^{\gamma_0} = \{\gamma \in \mathcal{M}_{\mathcal{A}} | \gamma << \gamma_0\},$$

which is the set of all finite signed measures which are absolutely continuous with respect to γ_0 over $\mathcal{A}$. By Radon-Nikodym, there is a one-to-one correspondence between elements of $\mathcal{M}_{\mathcal{A}}^{\gamma_0}$ and $L^1(\gamma_0)$. Also, if S is nonsingular, then $\mu \mapsto \mu^S$ is a mapping from $\mathcal{M}_{\mathcal{F}}^{\bar{\mu}}$ into $\mathcal{M}_{\mathcal{G}}^{\bar{\nu}}$; by the above one-to-one correspondence, this mapping can be identified with the Frobenius-Perron operator.

Closely related to this concept is the Koopman operator.

Definition 3.7. *The operator U, where $U : L^\infty(\bar{\nu}) \to L^\infty(\bar{\mu})$, defined by $Ug = g \circ S$, is called the Koopman operator (associated to S).*

Clearly, U is nonnegative in the sense that $g \geq 0$ implies $Ug \geq 0$ for all $g \in L^\infty(\bar{\nu})$, and U is a bounded linear operator with operator norm 1. Further, P is a nonnegative bounded linear operator with operator norm 1. To see that P is nonnegative, let $f \in L^1(\bar{\mu})$ be nonnegative everywhere and suppose that Pf is negative over a set $G \in \mathcal{G}$ of positive measure. This implies

$$\int_{S^{-1}(G)} f d\bar{\mu} = \int_G Pf d\bar{\nu} < 0,$$

which contradicts that f is nonnegative. To see that the operator norm of P is 1, observe for nonnegative $f \in L^1(\bar{\mu})$

$$\|Pf\|_{L^1(\nu)} = \int_Y Pf d\bar{\nu} = \int_X f d\bar{\mu} = \|f\|_{L^1(\mu)}.$$

This extends easily to the case of arbitrary $f \in L^1(\bar{\mu})$, using the decomposition of f into positive and negative parts.

Lemma 3.1. *The Koopman operator is the adjoint of the Frobenius-Perron operator.*

Proof. By (3.6), for all $f \in L^1(\bar{\mu}), g \in L^\infty(\bar{\nu})$,

$$< Pf, g >_{L^1(\bar{\nu}),L^\infty(\bar{\nu})} = \int (Pf) g d\bar{\nu} = \int \frac{d\bar{\mu}_f^S}{d\bar{\nu}} g d\bar{\nu}$$

$$= \int g d\bar{\mu}_f^S = \int (g \circ S) d\bar{\mu}_f = \int (g \circ S) f d\bar{\mu} = < f, Ug >_{L^1(\bar{\mu}),L^\infty(\bar{\mu})} .$$

□

Now, let $X = Y$ be a metric space, and let $\mathcal{F} = \mathcal{G} = \mathcal{B}$, where $\mathcal{B}$ is the Borel σ-algebra on X, and let $\bar{\mu}, \bar{\nu}$ be two σ-finite measures on $\mathcal{B}$. Now we take $S : \mathbb{R}_+ \times X \to X$ to be a nonsingular semidynamical system, that is, S is a semidynamical system such that $S_t : X \to X$ is nonsingular for all t. Then we can define $P_t f := \frac{d\bar{\mu}_f^{S_t}}{d\bar{\nu}}$ and $U_t g := g \circ S_t$, for each $t \in \mathbb{R}_+$.

More specifically, let $X = Y = \mathbb{R}^n$, let $\bar{\mu} = \bar{\nu} = \lambda^n$, and let $\mathcal{F} = \mathcal{G} = \mathcal{B}^n$. As in the previous section, let $b \in C^1(\mathbb{R}^n, \mathbb{R}^n)$ and let $\dot{y} = b(y)$ generate a global solution flow S; recall that S is nonsingular due to the change of variable formula (3.1). We now observe that $P := \{P_t\}_{t \geq 0}$ and $U := \{U_t\}_{t \geq 0}$ (both of which are associated to S) are in fact semigroups; let $t, s \geq 0$, and let $g \in L^\infty(\mathbb{R}^n)$. U is a semigroup because

$$U_{t+s}(g) = U((S_{t+s}(g)) = U(S_t(S_s(g))) = U_t(S_s(g)) = U_s(U_t(g)),$$

and clearly, $U_0(g) = g$. Also, P is a semigroup, since for any $f \in L^1(\mathbb{R}^n)$ and $A \in \mathcal{B}^n$

$$\int_A P_{t+s} f = \int_{S_{t+s}^{-1}(A)} f = \int_{S_s^{-1}(S_t^{-1}(A))} f = \int_{S_t^{-1}(A)} P_s f = \int_A P_s(P_t f),$$

so $P_{t+s} f = P_s(P_t f)$ (and clearly, $P_0 f = f$). We have already observed that U_t and P_t have operator norm equal to one for each $t \geq 0$, therefore P and U are in fact contraction semigroups.

The next property we need is strong continuity. Call $C_c^1(\mathbb{R}^n)$ the space of functions in $C^1(\mathbb{R}^n)$ that have compact support.

Lemma 3.2. *P is a strongly continuous semigroup.*

Proof. By (3.1), we have (a.s.)

$$P_t f(x) = f(S_{-t}x)|\det((DS_{-t})(x))|$$

for all $f \in C_c^1(\mathbb{R}^n)$, and

$$\lim_{t\to 0^+} f(S_{-t}x)|\det((DS_{-t})(x))| = f(x).$$

We claim that the limit is uniform in x. To see this, define K_t for $t \in [0,1]$ to be the support of $P_t f$, so K_0 is the support of f. Then $K_t = S_t(K_0)$ for all $t \in [0,1]$, since $|\det(DS_{-t})(x)|$ is never zero and

$$P_t f(x) \neq 0 \Leftrightarrow f(S_{-t}x) \neq 0 \Leftrightarrow S_{-t}x \in K_0 \Leftrightarrow x \in S_t(K_0).$$

But K_0 is compact and S is continuous on $\mathbb{R}_+ \times \mathbb{R}^n$, so $K := \bigcup_{t\in[0,1]} K_t = S([0,1] \times K_0)$ is compact. Therefore, $(t,x) \mapsto f(S_{-t}x)$ is uniformly continuous on $[0,1] \times K$, and so is $(t,x) \mapsto f(S_{-t}x)|\det((DS_{-t})(x))|$. Finally, realizing that if $h : [0,1] \times \mathbb{R}^n \to \mathbb{R}$ is uniformly continuous, then $h(t,\cdot)$ converges uniformly to $h(0,\cdot)$ as $t \to 0^+$ proves our claim.

Our claim implies

$$\lim_{t\to 0} \| P_t f - f \|_{L^1} = \int_{\mathbb{R}^n} \lim_{t\to 0} |P_t f(x) - f(x)|dx = \int_K \lim_{t\to 0} |P_t f(x) - f(x)|dx = 0.$$

Since P is uniformly bounded, by Banach-Steinhaus, $P_t f \to f$ in L^1 for any $f \in L^1(\mathbb{R}^n)$ (the closure of $C_c^1(\mathbb{R}^n)$ in the L^1-norm). This means P is strongly continuous. □

As a consequence, we know that P has an infinitesimal generator A_{FP} by Hille-Yosida. To identify A_{FP}, we will use the duality between the Frobenius-Perron and Koopman operators, first showing that U has an infinitesimal generator A_K and identifying it. To this end, let $t \in \mathbb{R}_+$, $x \in \mathbb{R}^n$, and suppose $f \in C_c^1(\mathbb{R}^n)$; then by the Mean Value Theorem, definition of U, and definition of solution semiflow,

$$\frac{U_t f(x) - f(x)}{t} = \frac{f(S(t,x)) - f(x)}{t} =$$

$$\nabla f(S(ct,x)) \cdot \frac{\partial}{\partial t} S(ct,x) = b(S(ct,x)) \cdot \nabla f(S(ct,x)),$$

for some $0 \leq c \leq 1$. Then

$$\lim_{t \to 0^+} \frac{U_t f(x) - f(x)}{t} = \lim_{t \to 0^+} b(S(ct,x)) \cdot \nabla f(S(ct,x)) = b(x) \cdot (\nabla f(x)).$$

By a similar argument as in the proof of Lemma 3.2, the limit is uniform in x. Thus, for at least $f \in C_c^1(\mathbb{R}^n)$, $\frac{U_t f - f}{t}$ converges in L^∞; in particular, $U_t f$ converges to f in L^∞ for all $f \in C_c^1(\mathbb{R}^n)$ (differentiability implies continuity). Since U is uniformly bounded, by the Banach-Steinhaus theorem, $U_t f \to f$ in L^∞ for all $f \in C_0$ (the closure of $C_c^1(\mathbb{R}^n)$ in the L^∞-norm). Further, $U_t(C_0) \subset C_0$, so U restricts to a strongly continuous semigroup on C_0. So, by Hille-Yosida, $U|_{C_0}$ has an infinitesimal generator A_K; our calculation shows $C_c^1(\mathbb{R}^n) \subset D(A_K)$ and that, for all $f \in C_c^1(\mathbb{R}^n)$,

$$A_K f = b \cdot (\nabla f). \tag{3.7}$$

We now use the duality between the Frobenius-Perron operators and the Koopman operators to identify A_{FP}. Note that P is not really the dual of $U|_{C_0}$; also, we would need reflexivity to insure that a strongly continuous contraction semigroup T has a strongly continuous dual contraction semigroup T^*, and that the dual of the generator of T is really the generator of T^* [17, Theorem 3.7.1]. However, let $g \in C_c^1(\mathbb{R}^n)$ and let $f \in D(A_{FP})$ be continuously differentiable. Then for any $t \geq 0$ we have

$$< P_t f, g >_{L^1(\mathbb{R}^n),L^\infty(\mathbb{R}^n)} = < f, U_t g >_{L^1(\mathbb{R}^n),L^\infty(\mathbb{R}^n)},$$

and we can subtract $< f, g >_{L^1(\mathbb{R}^n),L^\infty(\mathbb{R}^n)}$ from both sides and divide by t:

$$< \frac{P_t f - f}{t}, g >_{L^1(\mathbb{R}^n),L^\infty(\mathbb{R}^n)} = < f, \frac{U_t g - g}{t} >_{L^1(\mathbb{R}^n),L^\infty(\mathbb{R}^n)} .$$

We know that the limit as $t \to 0^+$ exists on both sides; on the right hand side, take this limit and use (3.7) and integration by parts:

$$< f, A_K g >_{L^1(\mathbb{R}^n),L^\infty(\mathbb{R}^n)} = \int f(A_K g) = \int f[b \cdot (\nabla g)]$$

$$= -\int g[\nabla \cdot (bf)]dx = < -\nabla \cdot (bf), g >_{L^1(\mathbb{R}^n),L^\infty(\mathbb{R}^n)} .$$

The above calculation identifies A_{FP}, which we have already shown to exist. In fact, for all $f \in D(A_{FP}) \cap C^1(\mathbb{R}^n)$,

$$A_{FP} f = -\nabla \cdot (bf).$$

Thus, we have proved

Theorem 3.1. *The infinitesimal generator A_K of the Koopman semigroup (restricted to C_0) is given by $A_K f = b \cdot \nabla f$ for $f \in C_c^1(\mathbb{R}^n)$. The infinitesimal generator A_{FP} of the Frobenius-Perron semigroup is given by $A_{FP} g = -\nabla \cdot (bg)$ for continuously differentiable $g \in D(A_{FP})$.*

Consider what happens in the case where we have a deterministic differential equation with global solution flow $S = \{S_t\}_{t \geq 0}$ with a "noisy" initial value, that is, an initial value that is a nondegenerate random variable, say, $X_0 = X^0$ a.s. Then we have the initial value problem

$$dX_t = b(X_t)dt, \tag{3.8}$$

$$X_0 = X^0 \quad \text{a.s.}$$

Lemma 3.3. *$X := (X_t)_{t \in \mathbb{R}_+}$ defined by $X_t := S_t \circ X_0$ for all $t \in \mathbb{R}_+$ solves (3.8).*

Proof. The proof is easy. Obviously, the initial condition is satisfied, and further,

$$\frac{\partial}{\partial t} X(t, \omega) = \frac{\partial}{\partial t} S(t, X^0(\omega)) = b(S(t, X^0(\omega))) = b(X_t(\omega)),$$

for any $\omega \in \Omega$. □

Remark: As a result of the lemma, we get the following useful equations:

$$P^{X_t}(B) = P^{S_t \circ X_0}(B) = (P^{X_0})^{S_t}(B) = P^{X_0}(S_t^{-1}(B)),$$

for all $t \in \mathbb{R}_+$ and $B \in \mathcal{B}^n$. Extending this to finite-dimensional distributions, we get

$$P^{(X_{t_1}, X_{t_2}, \cdots, X_{t_k})}(B_1 \times B_2 \times \cdots \times B_k) = P^{X_0}(\bigcap_{i=1}^{k} S_{t_i}^{-1}(B_i)), \tag{3.9}$$

for all $t_1, t_2, \cdots, t_k \in \mathbb{R}_+$ and $B \in \mathcal{B}^n$.

The Frobenius-Perron semigroup P gives us a new way of understanding (3.8) if X^0 has a density. Let X^0 have density g, so $P^{X^0} = \lambda_g$. Recall that, for any $t \in \mathbb{R}_+$, we denote the distribution of X_t by P^{X_t} (so $P^{X_0} = P^{X^0}$). Then P^{X_t} also has a density (since S_t is a diffeomorphism for each t), and $P^{X_t} = \lambda_{P_t g}$ since for any $A \in \mathcal{B}^n$,

$$P^{X_t}(A) = P^{(S_t \circ X_0)}(A) = P^{X_0}(S_t^{-1}(A)) = \int_{S_t^{-1}(A)} g(x)dx = \int_A P_t g(x)dx.$$

But the strong continuity of P allows us to use Theorem 1.2; we may set up the Cauchy problem

$$u_t = -\nabla \cdot (bu) = A_{FP} u \tag{3.10}$$

$$u(0, x) = g(x),$$

where g is the density of X^0. Solving (3.10) gives $u(t, \cdot) = P_t g(\cdot)$, the density of P^{X_t}, for any $t \in \mathbb{R}_+$. We call (3.10) the Liouville equation; one can interpret this physically as a conservation of mass equation (where b is a velocity field for a fluid with density u).

In summary, we have following lemma:

Lemma 3.4. *If X solves (3.8) and $P^{X_0} = \lambda_g$, then $P^{X_t} = \lambda_{P_t g} = \lambda_{u(t,\cdot)}$ for all $t \geq 0$, where u is the solution of (3.10).*

Probably the easiest example is the scalar "transport" equation $\dot{x} = b$, for a positive constant b, with initial condition $x(0) = x_0$ whose solution is $x(t) = S_t(x_0) := bt + x_0$. If there were an entire family of degenerate initial conditions (occurring with some probability), they would all be subject to the "transporting" motion $x \mapsto x + bt$.

Now, add a noisy initial condition so the transport equation becomes $dX_t = b\ dt$ with initial condition $X_0 = X^0$ having density g, where b is a positive constant. Then we can use the Frobenius-Perron semigroup to study the fluctuation of g via the equation $\frac{\partial}{\partial t}u(t,x) = A_{FP}u := -\nabla \cdot (bu)$ with initial condition $u(0,x) = g(x)$, which has solution $P_t g(x) = g(x - bt)$. This makes sense; imagine a process heavily concentrated at $x_0 \in \mathbb{R}$ initially, so that $g(x_0)$ is a "spike". Then the process should be heavily concentrated at $x_0 + bt$ after some time t, so the density g at this time should be "spiking" at $g(x_0 - bt)$.

For the weak version of (3.8), we would be given the initial condition $P^{X_0} = \mu$ instead of $X_0 = X^0$ a.s.; let us point out another way to determine P^{X_t}, for $t \geq 0$, given $P^{X_0} = \mu$, a way that works whether μ has a density or not.

Lemma 3.5. *Suppose, for every $x \in \mathbb{R}^n$, that $\dot{y} = b(y), y(0) = x$ has a global forward-in-time solution X^x with distribution $P^x := P^{X^x}$. Given a probability measure μ on $\mathcal{B}^n$, define P^μ by*

$$P^\mu(A) := \int_{x \in \mathbb{R}^n} P^x(A)\mu(dx),$$

for any $A \in \mathcal{B}(C)$. Then a stochastic process $X = (X_t)_{t \in \mathbb{R}_+}$ is a solution to the initial value problem

$$dX_t = b(X_t)dt, \tag{3.11}$$

$$P^{X_0} = \mu,$$

iff $P^X = P^\mu$.

Remark 1: P^μ is well-defined ($\dot{y} = b(y)$ determines a trivial convolution semigroup of measures; see [9, Lemma 8.7] or the discussion on pp. 9-10 in the section "Kernels and Semigroups of Kernels").

Remark 2: Clearly, $X_t^x = S(t,x)$ a.s. by Lemma 7, where S is the solution semiflow generated by $\dot{y} = b(y)$, and so, $P^{X_t^x} = \epsilon_{S(x,t)}$ for all $t \in \mathbb{R}_+$, $x \in \mathbb{R}^n$. Further, $(X^x)_{x\in\mathbb{R}^n}$ is a family of strong solutions to $\dot{y} = b(y)$ parameterized by $x \in \mathbb{R}^n$ (in the initial condition $y(0) = x$) such that each process X^x may live on a different probability space $(\Omega_x, \mathcal{F}_x, P_x)$ for each x; however, this doesn't matter, as all we care about is P^x.

Proof. Let $X := (X_t)_{t\geq 0}$ be a stochastic process on a probability space $(\Omega, \mathcal{F}, P)$. We want to show that X solves (3.11) iff $P^X = P^\mu$. To this end, suppose that X solves (3.11); we must show that $P^X = P^\mu$. Since X solves (3.11), $X = (S_t \circ X_0)_{t\geq 0}$ by Lemma 3.3 and $P^{X_0} = \mu$. Now let $k \in \mathbb{N}$, let $t_1, t_2, \cdots, t_k \in \mathbb{R}_+$, let $B_1, B_2, \cdots, B_k \in \mathcal{B}^n$, and call $\pi_{t_1,t_2,\cdots,t_k}$ the mapping from $C := C(\mathbb{R}_+, \mathbb{R}^n)$ into $\mathbb{R}^{n\times k}$ given by $\omega \mapsto (\omega(t_1), \omega(t_2), \cdots, \omega(t_k))$ Then, by definition of P^μ,

$$P^\mu(\pi^{-1}_{t_1,t_2,\cdots,t_k}(B_1 \times B_2 \times \cdots \times B_k))$$

$$= \int_{\mathbb{R}^n} P^{X^x}(\pi^{-1}_{t_1,t_2,\cdots,t_k}(B_1 \times B_2 \times \cdots \times B_k))\mu(dx)$$

$$= \int_{\mathbb{R}^n} P^{(X^x_{t_1}, X^x_{t_2}, \cdots, X^x_{t_n})}(B_1 \times B_2 \times \cdots \times B_k)\mu(dx).$$

Now, since X^x solves (3.8) and since $X_0^x = x$ a.s., we may apply (3.9) twice to see that the above equals

$$\int 1_{(S_{t_1}^{-1}(B_1)\cap S_{t_2}^{-1}(B_2)\cap\cdots\cap S_{t_k}^{-1}(B_k))}(x)\mu(dx)$$

$$= \mu(S_{t_1}^{-1}(B_1)\cap S_{t_2}^{-1}(B_2)\cap\cdots\cap S_{t_k}^{-1}(B_k))$$

$$= P^{(X_{t_1},X_{t_2},\cdots,X_{t_k})}(B_1\times B_2\times\cdots\times B_k),$$

and so $P^\mu = P^X$.

For the other implication, suppose that $P^X = P^\mu$. Then we must show that X is a solution to (3.11). To see this, observe that X satisfies the initial condition, since

$$P^{X_0}(B) = \mu(B),$$

for any $B \in \mathcal{B}^n$. It only remains to show that $(X_t)_{t\in R_+}$ and $(S_t \circ X_0)_{t\in\mathbb{R}_+}$ have the same joint distributions, which follows similarly to the above:

$$P^{(X_{t_1},X_{t_2},\cdots,X_{t_k})}(B_1\times B_2\times\cdots\times B_k)$$

$$= P^\mu(B_1\times B_2\times\cdots\times B_k)$$

$$= \mu(S_{t_1}^{-1}(B_1)\cap S_{t_2}^{-1}(B_2)\cap\cdots\cap S_{t_k}^{-1}(B_k))$$

$$= P^{(S_{t_1}\circ X_0,S_{t_2}\circ X_0,\cdots,S_{t_k}\circ X_0)}(B_1\times B_2\times\cdots\times B_k),$$

thus X is a solution and the lemma holds.

□

To interpret this result, fix x_1, x_2 in $\mathbb{R}^n$ and $\alpha \in [0,1]$, and call $\mu_\alpha = \alpha\epsilon_{x_1} + (1-\alpha)\epsilon_{x_2}$. Let X^{x_1} and X^{x_2} denote the respective solutions to (3.8) with initial conditions $X_0 = x_1$ a.s. and $X_0 = x_2$ a.s. Then if X solves (3.11) with initial condition $P^{X_0} = \mu_\alpha$, X must have distribution P^{μ_α}, where

$$P^{\mu_\alpha} = \alpha P^{X^{x_1}} + (1-\alpha)P^{X^{x_2}}.$$

So, if X is a strong solution to (3.8) with initial condition $X_0 = X^0$, where X^0 is a random variable equal to x_1 with probability α and x_2 with probability $1-\alpha$, then X is a modification of the process $\tilde{X}$, where $\tilde{X} = X^{x_1}$ with probability α and $\tilde{X} = X^{x_2}$ with probability $1-\alpha$. In this way, one may interpret the action of the above lemma as a kind of "stochastic superposition" (not the usual "superposition principle," which says that the linear combination of solutions is also a solution, which we cannot expect unless we assume b is linear). More profoundly, this extends to even nonzero σ, which means it suffices to examine degenerate initial conditions for (weak) stochastic differential equations. We extend the above ideas to the stochastic case in the next section, emphasizing the use of the Frobenius-Perron semigroup.

3.3 Koopman and Frobenius-Perron Operators: The Stochastic Case

We have studied $(0, b)$ with a degenerate initial distribution, and also for a noisy initial condition which is the nondegenerate distribution of an initial random variable (with a density). We want σ to be non-zero now, so let us extend the notions of Koopman operator and Frobenius-Perron operator to the stochastic case and then derive extended versions of A_K and A_{FP}. As before we derive A_K and exploit "duality" to obtain A_{FP}.

As in the previous section, it suffices to study the degenerate solutions by integrating P^x over x with respect to a given nondegenerate initial distribution μ; the proof involves deep probabilistic concepts (see [9, Theorem 21.10] and the preceding theorems ibidem) so we will simply state the result.

Lemma 3.6. *Suppose, for every $x \in \mathbb{R}^n$, that $dX_t = b(X_t)dt + \sigma(X_t)dB_t$, $X_0 = x$ a.s. has a global forward-in-time solution $X^x = (X_t^x)_{t\in\mathbb{R}_+}$ with distribution $P^x := P^{X^x}$. Given a probability measure μ on $\mathcal{B}^n$, define P^μ by*

$$P^\mu(A) := \int_{x\in\mathbb{R}^n} P^x(A)\mu(dx),$$

for any $A \in \mathcal{B}(C)$. Then a stochastic process $X = (X_t)_{t\in\mathbb{R}_+}$ is a solution to the initial value problem

$$dX_t = b(X_t)dt + \sigma(X_t)dB_t, \tag{3.12}$$

$$P^{X_0} = \mu,$$

iff $P^X = P^\mu$.

Under the assumptions of Lemma 3.6, let $X^x := (X_t^x)_{t\geq 0}$ denote the solution process to (σ, b) with initial condition $X_0 = x$ a.s. and let P^x be the distribution of X^x. Define E^x to be the expectation with respect to P^x. We can now define the stochastic Koopman operator; as before, we want something like "$U_t f = f \circ S_t$" for a "stochastic solution semiflow" (we can't backsolve; the noise in the flow causes a "diffusion") S (which acts on random variables) and we also want U to map $L^\infty(\mathbb{R}^n)$ into $L^\infty(\mathbb{R}^n)$. But something must give; we are dealing with a space of $\mathbb{R}^n$-valued random variables rather than $\mathbb{R}^n$. So, let $f \in L^\infty(\mathbb{R}^n)$; then for fixed $t \in \mathbb{R}_+$ and fixed $x \in \mathbb{R}^n$, $f \circ (X_t^x)$ is a bounded and Borel-measurable function that maps $C(\mathbb{R}_+, \mathbb{R}^n)$ into $\mathbb{R}$, and so $E^x(f \circ X_t^x)$ makes sense. This leads us to the following definition:

Definition 3.8. *Let $f \in L^\infty(\mathbb{R}^n)$ and $(X_t^x) := (X_t^x)_{t\geq 0, x\in\mathbb{R}^n}$ be the family such that, for every fixed x, (X_t^x) is the canonical realization of the solution to (σ, b) with initial condition $X_0 = x$ a.s. Define for all $t \in \mathbb{R}_+$ the stochastic Koopman operator $U_t : L^\infty(\mathbb{R}^n) \to L^\infty(\mathbb{R}^n)$ by*

$$U_t f(x) := E^x(f \circ X_t^x),$$

for all $f \in L^\infty(\mathbb{R}^n)$ and $x \in \mathbb{R}^n$.

Let us comment on this definition; first observe that the canonical realization is necessary; P^x (and hence E^x) lives over $\mathcal{B}(C)$. Next, note that it is consistent with the deterministic Koopman operator as $X_t = S_t \circ X_0 = S_t(x)$ when $X_0 = x$ a.s. for a global solution flow S generated by $\dot{y} = b(y)$, so $U_t f(x)$ reduces to $f(S_t(x))$. Next, it is obvious that $U_t f \in L^\infty(\mathbb{R}^n)$ for all $t \in \mathbb{R}_+$ and $f \in L^\infty(\mathbb{R}^n)$, and that U_t is nonexpansive for all

$t \in R_+$. In fact, $U := \{U_t\}_{t\geq 0}$ restricts to a strongly continuous semigroup on C_0 (see [9, Theorem 21.11]), so it has an infinitesimal generator. We emulate the argument in the deterministic case to identify it; let $n = 1$ for notational ease (the argument for any n is similar). Let $f \in C_c^2(\mathbb{R})$ and recall that if X^x solves (σ, b) with initial condition $X_0 = x$ a.s., then X^x and $f(X^x)$ are continuous martingales (which means we may apply Ito's formula). For the remainder of this argument, we call $X_t := X_t^x$ for convenience. Then we have $X_t = x + \int_0^t b(X_s)ds + \int_0^t \sigma(X_s)dB_s$, so apply Ito's formula to $f(X_t)$:

$$f(X_t) = f(x) + \int_0^t [b(X_s)f'(X_s) + \frac{1}{2}\sigma^2(X_s)f''(X_s)dt] + \int_0^t \sigma(X_s)f'(X_s)dB_s.$$

Taking the expected value with respect to P^x of both sides, we get

$$E^x f(X_t) = f(x) + E^x\bigg(\int_0^t b(X_s)f'(X_s) \tag{3.13}$$

$$+\frac{1}{2}\sigma^2(X_s)f''(X_s)dt\bigg) + E^x(\int_0^t \sigma(X_s)f'(X_s)dB_s).$$

By basic properties of Ito integrals, $E^x(\int_0^t \sigma(X_s)f'(X_s)dB_s) = 0$, so (3.13) becomes

$$E^x f(X_t) = f(x) + E^x(\int_0^t b(X_s)f'(X_s) + \frac{1}{2}\sigma^2(X_s)f''(X_s)dt).$$

Now, by definition of infinitesimal generator,

$$\begin{aligned} Af(x) = \lim_{t\to 0} \frac{E^x f(X_t) - f(x)}{t} & \\ &= \lim_{t\to 0} \frac{E^x(\int_0^t [b(X_s)f'(X_s) + \frac{1}{2}\sigma^2(X_s)f''(X_s)]dt)}{t} \\ &= E^x(b(X_0)f'(X_0) + \frac{1}{2}\sigma^2(X_0)f''(X_0)) \\ &= b(x)f'(x) + \frac{1}{2}\sigma^2(x)f''(x). \end{aligned}$$

Thus we have the characterization for the infinitesimal generator of the stochastic Koopman operator:

$$A_K f = bf' + \frac{1}{2}\sigma^2 f'', \tag{3.14}$$

for any $f \in C_c^2(\mathbb{R})$, or, for any dimension n, for $0 \leq i, j \leq n$,

$$A_K f = \sum_{i=1}^{n} b_i \frac{\partial f}{\partial x_i} + \frac{1}{2}\sum_{i,j=1}^{n} a_{ij}\frac{\partial^2 f}{\partial x_i \partial x_j}, \tag{3.15}$$

for any $f \in C_c^2(\mathbb{R}^n)$, where $(a_{ij}) = \sum_{k=1}^{d} \sigma_{ik}\sigma_{jk}$ (the Brownian motion is d-dimensional, for $1 \leq d \leq n$). Note that if the noise were zero, the generator would correspond to the deterministic case, as expected.

Next, we obtain the infinitesimal generator of Frobenius-Perron operator associated to the "stochastic solution semiflow" S induced by the solution of (σ, b) in the case that S is nonsingular. We must impose here that $\frac{\partial b}{\partial x}$, $\frac{\partial \sigma}{\partial x}$, and $\frac{\partial^2 \sigma}{\partial x^2}$ exist and are bounded.

Naively, if something like "$A_{FP}^* = A_K$" holds then from (3.14) and integration by parts we would get

$$A_{FP}f = -(bf)' + \frac{1}{2}(\sigma^2 f)'', \tag{3.16}$$

or for any dimension n,

$$A_{FP}f = -\sum_{i=1}^{n} \frac{\partial(b_i f)}{\partial x_i} + \frac{1}{2}\sum_{i,j=1}^{n} \frac{\partial^2(a_{ij}f)}{\partial x_i \partial x_j}. \tag{3.17}$$

In fact, this is the case; for a "from scratch" proof, see [11, Theorem 11.6.1].

Now, let X solve (σ, b) with initial condition $X_0 = X^0$ a.s. and suppose X_0 has density g. Then we can set up the problem

$$\frac{\partial u}{\partial t} = A_{FP}u, \tag{3.18}$$

$$u(0, x) = g(x),$$

where the A_{FP} is as in (3.17) and the solution $u(t, \cdot)$ to (3.18) is the density of X_t for every t; $\frac{\partial u}{\partial t} = A_{FP}u$ is called the Fokker-Planck equation. We are interested in finding a fundamental solution to (3.18); we digress slightly to give some necessary definitions and notation that leads to one result that guarantees existence/uniqueness under some technical conditions.

First, let us rewrite $A_{FP}u$ in nondivergence form:

$$A_{FP}u = -\sum_{i=1}^{n} \frac{\partial(b_i(x)u)}{\partial x_i} + \frac{1}{2}\sum_{i,j=1}^{n} \frac{\partial^2(a_{ij}(x)u)}{\partial x_i \partial x_j} \tag{3.19}$$

$$= \tilde{c}(x)u + \sum_{i=1}^{n} \tilde{b}_i(x)\frac{\partial u}{\partial x_i} + \frac{1}{2}\sum_{i,j=1}^{n} a_{ij}\frac{\partial^2 u}{\partial x_i \partial x_j},$$

where

$$\tilde{b}_i(x) = -b_i(x) + \sum_{i,j=1}^{n} \frac{\partial a_{ij}(x)}{\partial x_j}$$

and

$$\tilde{c}(x) = \frac{1}{2}\sum_{i,j=1}^{n} \frac{\partial^2 a_{ij}(x)}{\partial x_i \partial x_j} - \sum_{i=1}^{n} \frac{\partial b_i(x)}{\partial x_i}.$$

Of course, the coefficients must be sufficiently smooth for the above to make sense; we also want them to satisfy growth conditions, namely, that there is a positive constant M such that

$$|a_{ij}(x)| \leq M, |\tilde{b}_i(x)| \leq M(1+|x|), |\tilde{c}(x)| \leq M(1+|x|^2). \tag{3.20}$$

We know $a_{ij} = a_{ji}$, so given any $\lambda = (\lambda_1, \lambda_2, \cdots, \lambda_n) \in \mathbb{R}^n$, we at least know $\sum_{i,j=1}^{n} a_{ij}\lambda_i\lambda_j = \sum_{k=1}^{n}(\sum_{i=1} \sigma_{ik}(x)\lambda_i)^2 \geq 0$. We would like strict inequality, so let us assume that the uniform parabolicity property holds, that is, that there is a constant $\rho > 0$ such that

$$\sum_{i,j=1}^{n} a_{ij}(x)\lambda_i\lambda_j \geq \rho\sum_{i=1}^{n} \lambda_i^2, \tag{3.21}$$

for any $x \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}^n$.

We condense the above into the following definition:

Definition 3.9. *Given (3.18), we say a_{ij} and b_i are Cauchy-regular if they are C^4 functions such that the corresponding a_{ij}, $\tilde{b}_i$ and $\tilde{c}$ of (3.19) satisfy (3.21) and (3.20).*

Now we recall the definition of a classical solution.

Definition 3.10. *Let $f \in C(\mathbb{R}^n)$. We say $u : \mathbb{R}_+ \times \mathbb{R}^n \to \mathbb{R}^n$ is a classical solution of (3.18) if*

i) for all $T > 0$ there are positive constants c, α such that $|u(t,x)| \leq ce^{\alpha x^2}$ for all $0 < t \leq T$, $x \in \mathbb{R}^n$,

ii) $u_t, u_{x_i}, u_{x_i,x_j}$ are continuous for all $1 \leq i, j, \leq n$ and u satisfies

$$u_t = \tilde{c}(x)u + \sum_{i=1}^{n} \tilde{b}_i(x)\frac{\partial u}{\partial x_i} + \frac{1}{2}\sum_{i,j=1}^{n} a_{ij}\frac{\partial^2 u}{\partial x_i \partial x_j},$$

for all $t > 0$ and $x \in \mathbb{R}^n$, and

iii) $\lim_{t\to 0} u(t,x) = f(x)$.

We are now able to state the desired existence/uniqueness theorem:

Theorem 3.2. *Given (3.18), let a_{ij}, b_i be Cauchy-regular and let $f \in C(\mathbb{R}^n)$ satisfy $|f(x)| \leq ce^{\alpha x^2}$ with positive constants c, α. Then there is a unique classical solution to (3.18) given by $u(t,x) = \int \Gamma(t,x,y)f(y)dy$, where the fundamental solution (or kernel) $\Gamma(t,x,y)$ is defined for all $t > 0$, $x, y \in \mathbb{R}^n$, is continuous and differentiable with respect to t, twice differentiable with respect to x_i for all $1 \leq i \leq n$, and satisfies the equation*

$$u_t = \tilde{c}(x)u + \sum_{i=1}^{n} \tilde{b}_i(x)\frac{\partial u}{\partial x_i} + \frac{1}{2}\sum_{i,j=1}^{n} a_{ij}\frac{\partial^2 u}{\partial x_i \partial x_j}$$

as a function of t and x for every fixed y.

Our slight digression concludes with at least one condition under which a fundamental solution exists. Now, if we are able to find a fundamental solution $\Gamma(t, x, y)$ to the Fokker-Planck equation then given any initial condition $u(0, x) = g(x)$, where $g \in L^1(\mathbb{R}^n)$, we can define a family of operators $\{P_t\}_{t\geq 0}$ by

$$u(t, x) = P_t(g(x)) = \int_{\mathbb{R}^d} \Gamma(t, x, y)g(y)dy, \tag{3.22}$$

and u is often called a generalized solution in this case (of course, g has to be continuous in order for u to be a classical solution).

Definition 3.11. *We call $\{P_t\}_{t\geq 0}$ a stochastic semigroup if $\{P_t\}_{t\geq 0}$ is Markovian semigroup of linear operators (on $L^1(\mathbb{R}^n)$) that is monotone ($P_t f \geq 0$ when $f \geq 0$, for all $t \in \mathbb{R}_+$) and norm-preserving ($\|P_t f\| = \|f\|$ when $f \geq 0$, for all $t \in \mathbb{R}_+$).*

The proof of the next theorem can be found in ([11, pp. 369-370]).

Theorem 3.3. *$\{P_t\}_{t\geq 0}$ as defined in (3.22) is a stochastic semigroup.*

This theorem justifies the following definition:

Definition 3.12. *We call $P := \{P_t\}_{t\geq 0}$ as defined in (3.22) the stochastic Frobenius-Perron semigroup.*

Let us now consider the simple example

$$dX_t = dB_t, \tag{3.23}$$

$$X_0 = X^0 \ \ a.s.,$$

where X^0 has density g. Then the solution is a Brownian motion, and (3.18) becomes the heat equation

$$u_t = \frac{1}{2}\Delta u,$$

$$u(0,x) = g(x),$$

which has solution

$$u(t,x) = (\frac{1}{2\pi t})^{\frac{d}{2}} \int_{\mathbb{R}^d} e^{-\frac{|x-y|^2}{2t}} g(y)dy, \tag{3.24}$$

for $x \in \mathbb{R}^d, t \geq 0$. Notice that the fundamental solution

$$(\frac{1}{2\pi t})^{\frac{d}{2}} e^{-\frac{|x-y|^2}{2t}}$$

is the density of a Brownian motion, as we expect.

One way to think about what happens is that, for a noiseless stochastic differential equation with degenerate initial condition, we have a point moving through space in time governed by a flow (in essence, an ordinary differential equation). If the initial condition is nondegenerate with a density, we may understand how the family of points evolves as a density via the partial differential equation generated by the Frobenius-Perron operator.

Now, if a stochastic differential equation has a degenerate initial condition, we still have a point moving through space in time governed by a flow, but there is noise and we cannot actually tell where that point is; we are fluctuating random variables or measures. If the measures are absolutely continuous, we may instead fluctuate densities just as in the

previous case, which means that "deterministic partial differential equations have the same complexity as stochastic differential equations with degenerate initial conditions."

Another interpretation for the latter case is that a point moves through space governed by a Brownian motion whose "expected flow" is described by b and whose "spread" or "intensity" is described by σ. For example, in (3.23), the flow is trivial, so we expect that the point stays where it started in space, but as time goes the noise may move it away. With the above interpretation, we see now that there is no difference between $(\sigma, b) := (1, 0)$ with degenerate initial condition $X_0 = x$ and $(\sigma, b) := (0, b_L)$ with nondegenerate initial condition having a Lebesgue density g, where b_L can be derived from the Liouville equation.

So how much more complicated is the "mixed" case where neither σ nor b are zero? We can actually remove b from our consideration; this result is called the "transformation of drift" formula (so-called because b is often referred to as the "drift" term), which in our situation can be stated as follows (see [5, p. 43]):

Given any $x \in \mathbb{R}^n$, let X^x solve (σ, b) with initial condition $X_0^x = x$ a.s. Assume $\sigma : \mathbb{R}^n \to \mathbb{R}^n \times \mathbb{R}^n$ and $\sigma(y)$ has positive eigenvalues for every y. Further, let $f : \mathbb{R}^n \to \mathbb{R}^n$ and suppose Y_t^x solves $(\sigma, b + \sigma f)$. Then $P^{X_t^x}$ and $P^{Y_t^x}$ are absolutely continuous with respect to each other and

$$dP^{Y_t^x} = \exp[\int_0^t f(X_s^x) dB_s - \frac{1}{2}\int_0^t |f(X_s^x)|^2 dt] dP^{X_t^x}. \tag{3.25}$$

In particular, we could pick f such that $\sigma f = -b$, and obtain a relationship between (σ, b) and $(\sigma, 0)$; we have already realized how $(\sigma, 0)$ relates to a deterministic partial differential equation (as we did in the study of (3.23)). So, in theory, one can describe the

dynamical systems aspects of (σ, b) in general by tracing back to $(\sigma, 0)$ or $(0, b)$ (although this may be quite unwieldy).

Now that we understand dynamical systems in a stochastic setting, we move to the notions of stability in a stochastic setting, defining what the various notions of "stochastic stability" are as well as emulating Liapunov theory to demonstrate stability/instability of solutions to stochastic differential equations.

3.4 Liapunov Stability

We begin by recalling some notation and some basic notions of stability of deterministic dynamical systems.

As we discussed in section 3.1, "Stochastic Dynamical Systems," we let $b \in C^1(\mathbb{R}^n, \mathbb{R}^n)$ and consider the system $\dot{u} = b(u)$. Given any initial value $x \in \mathbb{R}^n$ there exists a largest open interval of existence $I_x \subset \mathbb{R}$ containing 0 such that the system $\dot{u} = b(u)$ has a unique solution $u^x \in C^1(I_x, \mathbb{R}^n)$ with $u^x(0) = x$. The system $\dot{u} = b(u)$ generates a local solution flow $S : D(S) \subset \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n$ with $D(S) := \{(t, x) \in \mathbb{R} \times \mathbb{R}^n | t \in I_x\}$ where $S(t, x) := u^x(t)$ for all $(t, x) \in D(S)$; we know $D(S)$ is open, S is $C^1(D(S), \mathbb{R}^n)$, and S is satisfies the group property.

In what follows, we assume that S is a global solution flow.

Definition 3.13. *We say $\bar{x}$ is an equilibrium point of S if $S(t, \bar{x}) = \bar{x}$ for every $t \in \mathbb{R}$.*

Observe that $\bar{x}$ is an equilibrium point of S iff $b(\bar{x}) = 0$.

Definition 3.14. *An equilibrium point $\bar{x}$ of S is called stable if for any $\epsilon > 0$, there is $\delta(\epsilon) > 0$ such that whenever $\|x - \bar{x}\| < \delta$, it follows that $\|S(t, x) - \bar{x}\| < \epsilon$ for all $t \geq 0$. An equilibrium point that is not stable is called unstable. An equilibrium point of a system is asymptotically stable if it is stable and, in addition, there is $r > 0$ such that $\lim_{t\to\infty} S(t, x) = \bar{x}$ for all x such that $\|x - \bar{x}\| < r$.*

We now recall the principle of linearized stability, which in essence extracts information about the stability of the nonlinear system from the stability of the linearized system. More specifically, for an equilibrium point $\bar{u}$, we linearize b at $\bar{u}$ so that our system becomes $\dot{v} = Db(\bar{u})v$, where $v = u - \bar{u}$ and $Db(\bar{u})$ is the Jacobian matrix. It can be shown [8,

Theorem 9.5 and Theorem 9.7] that if $Db(\bar{u})$ has only eigenvalues with negative real parts, then $\bar{u}$ is asymptotically stable, while if any eigenvalue has positive real part, then $\bar{u}$ is unstable (for eigenvalues with real part 0, the linearized system is insufficient to determine stability).

Assuming that $b(0) = 0$, we are interested in the stability of the trivial solution $u := 0$; we use Liapunov theory in this situation.

Definition 3.15. *We say a C^1-function $V : D(V) \subset \mathbb{R}^n \to \mathbb{R}$ is positive definite if $D(V)$ is open and contains the origin, if $V(0) = 0$ and if $V(x) > 0$ for all non-zero x. If $-V$ is positive definite, we call V negative definite. Define the orbital derivative of V to be $A_K V = (b \cdot \nabla)V = \sum_{i=1}^n b_i \frac{\partial V}{\partial x_i}$. We call a positive definite V (strictly) Liapunov if $A_K V(x) \leq (<)0$ for all nonzero x.*

The utility of Liapunov functions is illustrated in the following theorem, which is proven e.g. in [8, Theorem 9.12].

Theorem 3.4. *If 0 is an equilibrium point of $\dot{u} = b(u)$, and if there exists a (strictly) Liapunov function V, then 0 is (asymptotically) stable. Further, 0 is unstable if $A_K V > 0$.*

Moving to the stochastic case, we generalize the concepts of stability, orbital derivative, Liapunov function, and the principle of linearized stability. Stability and orbital derivative are fairly straight forward to generalize, and Liapunov functions are only a little trickier, but unfortunately, the principle of linearized stability is quite difficult to generalize. Recall that X^x denotes the solution to (σ, b) with degenerate initial condition $X_0 = x$ a.s.; assume global solvability, that is, assume X^x exists for every $x \in \mathbb{R}^n$. Throughout, assume that $b(0) = 0$ and $\sigma(0) = 0$, so that (σ, b) admits that trivial solution $X = 0$.

Definition 3.16. *If for all $\epsilon > 0$, we have*

$$\lim_{x \to 0} P(\sup_{t \geq 0} |X_t^x| > \epsilon) = 0,$$

then we say the trivial solution $X = 0$ is stable in probability.

In essence, this means that as x goes to 0, the probability that a path starting at x will remain in an arbitrarily prescribed neighborhood of 0 is 1. This is quite similar to the deterministic version of stability, except now when x is close to 0, the probability that X^x is also close to zero is close to 1.

Definition 3.17. *If $X = 0$ is stable in probability and, for every x,*

$$\lim_{x \to 0} P(\lim_{t \to \infty} |X_t^x| = 0) = 1,$$

we say $X = 0$ is asymptotically stable in probability.

Basically, this means that as x goes to 0, the probability that a path starting at x will eventually approach 0 as time goes to infinity is 1.

Definition 3.18. *Let (σ, b) admit a trivial solution $X^0 = 0$. If X^0 is stable in probability and, for every x,*

$$P(\lim_{t \to \infty} X_t^x = 0) = 1$$

we say X^0 is asymptotically stable in the large.

Asymptotic stability in the large is the most powerful notion of stability, since the probability that any path (no matter where it starts) goes to 0 as time goes to infinity is 1.

If we are to generalize the Liapunov stability theory to the above concepts, we would need to study the sign of the "stochastic orbital derivative"; to see what the "stochastic orbital derivative" is, we do a little reverse engineering. Notice that the deterministic orbital derivative takes the form of the generator of the deterministic Koopman semigroup A_K, so analogously, it makes sense to think that the "stochastic orbital derivative" should take the form of the generator of the stochastic Koopman semigroup. This formally justifies the following definition.

Definition 3.19. *For V in $C^2(\mathbb{R}^n)$, we define the stochastic orbital derivative of V to be*

$$A_K V = \sum_{i=1}^{n} b_i \frac{\partial V}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^{n} a_{ij} \frac{\partial^2 V}{\partial x_i \partial x_j},$$

where as before, $A := (a_{ij}) = (\sum_{k=1}^{n} \sigma_{ik}\sigma_{jk})$.

We remark that the notation "A_K" as well as the stochastic generalization of orbital derivative are consistent; they reduce to the deterministic case when σ is 0.

Now we can generalize the Liapunov theory, which parallels the deterministic case quite similarly; we remark up front that we are presenting a brief summary with some simplifying conditions and we are only operating in the time-homogeneous case, and that there are plenty of weaker assumptions and technical details behind what follows (the reader is invited to check [7, Chapter 5] for more).

Definition 3.20. *Let $V : D(V) \subset \mathbb{R}^n \to \mathbb{R}$, where $D(V)$ is open and contains the origin, $V(0) = 0$, and $V(x) > 0$ for all non-zero x. Further, let $V \in C^2(D(V) \setminus \{0\})$. We say V is a (strict) stochastic Liapunov function if $A_K V(x) \leq (<) 0$ for all nonzero x.*

Theorem 3.5. *If V is a stochastic Liapunov function then $X = 0$ is stable in probability. Further, if the matrix A has positive eigenvalues, then $X = 0$ is stable in probability iff it is asymptotically stable in probability.*

The proof of this theorem can be found in [7, pp. 164,168].

Asymptotic stability in the large is almost "too nice" for practical purposes; still, there are several conditions that are sufficient to guarantee it. One not surprising condition is that $X = 0$ is asymptotically stable in the large if $X = 0$ is stable in probability and recurrent to the domain $|x| < \epsilon$ for all $\epsilon > 0$ (a process Y is recurrent to A if $\sup\{t \geq 0 : P(Y_t \in A) = 1\} = \infty$, else it is transient). There are stricter conditions which can be imposed on V which are of little interest to us; see ([7, Theorem 4.4, Theorem 4.5]) for those details.

As far as instability goes, things are usually a little trickier. Intuitively, systems that are stable without noise may become unstable with the addition of noise. Much less intuitively, an unstable system can be stabilized by the addition of noise! We shall soon see examples of these situations, but for now, we state one sufficient condition for instability.

Theorem 3.6. *Let V be a stochastic Liapunov function with the exception that $D(V)$ may not contain zero, let $\lim_{x\to 0} V(x) = \infty$, and call $U_r = \{x \in D(V) \mid |x| < r\}$ for $r > 0$. If A has positive eigenvalues, then $X = 0$ is unstable in probability, and further, $P(\sup_{t>0} |X_t^x| < r) = 0$ for all $x \in U_r$.*

Contrast this to the deterministic case, and notice that $A_K V$ does not change sign but V is now "inversely positive definite," which makes the above believable.

Let us now look at some examples; of course, there is little to do with the trivial solutions to the transport equation or the Langevin equation, so let us move the next most complicated example.

Example 3.1.

Reconsider the one-dimensional equation $dX_t = bX_t dt + \sigma X_t dB_t$, where b, σ are positive constants, with initial condition $X_0 = x$ a.s. We have already solved this explicitly, and we know its solution is

$$X_t = xe^{(b-\frac{1}{2}\sigma^2)t+\sigma B_t}.$$

We can see that when $2b < \sigma^2$, the expected value of the solution decays to 0 as time goes to infinity, so we expect that the condition $2b < \sigma^2$ insures the zero solution $X = 0$ is stable; let us use the Liapunov theory to verify this. Pick $V(x) = |x|^{1-\frac{2b}{\sigma^2}}$; V is positive-definite and twice-continuously differentiable (except at 0) so we may examine $A_K V$ for nonzero x:

$$A_K V(x) = bxV'(x) + \frac{1}{2}\sigma^2 x^2 V''(x),$$

which is the same as

$$A_K V(x) = bx(1 - \frac{2b}{\sigma^2})|x|^{-\frac{2b}{\sigma^2}} + \frac{1}{2}\sigma^2 x^2 (1 - \frac{2b}{\sigma^2})(-\frac{2b}{\sigma^2})|x|^{-\frac{2b}{\sigma^2}-1}.$$

With a bit of algebra, it is clear that $A_K V < 0$ when $2b < \sigma^2$. Thus, $X = 0$ is asymptotically stable in probability.

Computationally, this example is quite simple, and interpreting stability in this case as an "extinct population" is reasonable. However, the results may cause the reader difficulty

when it comes to a physical interpretation. Notice that if there is no noise, we have $\dot{x} = bx$, where $b > 0$; clearly this has an unstable trivial solution, so in this case, adding "enough" noise actually stabilizes the trivial solution. This does not jibe with our physical intuition, so for consistency's sake the condition $2b < \sigma^2$ as above is deemed "physically unfeasible" (for a further discussion of this, see [7, 173-176]).

Remark: The discussion in ([7, 173-176]) will appeal to readers interested in a contrast of the Ito (left-endpoint) interpretation and the Stratonovich (midpoint) interpretation of the stochastic integral. It turns out that, under the Stratonovich interpretation, the sign of b alone determines the stability of the trivial solution.

Along these lines, if "not enough" noise is added, or really, "not enough physically feasible" noise is added, then the trivial solution should remain unstable; this is indeed the case when $2b > \sigma^2$. To see this, select $V(x) = -\ln|x|$. Then all the conditions to determine instability are satisfied, since for non-zero x,

$$A_K V = -b + \frac{1}{2}\sigma \leq 0.$$

Of course, if b is negative, the trivial solution is stable no matter what σ is.

It is intuitive to think that any stable system will become unstable with the addition of enough noise, but in fact it depends upon the dimension of the space. We can mimic the above argument in a fairly general setting: suppose we have a system of n equations, each equation of which has a stable trivial solution. Now add noise to it so our system becomes

$$dX_t = b(X_t)dt + \sigma X_t dB_t,$$

for $\sigma > 0$ a constant. Then picking $V(x) = -\ln(|x|^2)$, we see after several steps of calculus that

$$A_K V(x) = -\frac{2x \cdot b(x)}{|x|^2} - \sigma^2(n-2).$$

We satisfy the hypotheses of Theorem 3.6 when $n > 2$, as we can pick σ large enough to destroy the stability of the trivial solution of the original system. Notice that if $n = 2$ and we take $b^{(i)}(X_t) := b_i X_t$ for $i = 1, 2$, where b_i are negative constants for $i = 1, 2$, the asymptotic stability of the system cannot be destroyed by arbitrarily large noise; let σ be any constant. Then there is a sufficiently small positive constant $a := a(\sigma)$ such that taking $V(x) = |x|^a$ yields

$$A_K V(x) = a|x|^{a-2}(b_1 x_1^2 + b_2 x_2^2 + \frac{a\sigma^2|x|^2}{2}) < 0.$$

This means the trivial solution of the system is still asymptotically stable (in fact, asymptotically stable in the large).

Let us move to the situation where the trivial solution is stable, but not asymptotically stable. In this case, stability may be so delicate that even the slightest of noise ruins it; this is exhibited in the next example.

Example 3.2.

Consider the system

$$dX^1 = X^2 dt + \sigma(X) dB_t^1,$$

$$dX^2 = -X^1 dt + \sigma(X) dB_t^2,$$

where $X = (X^1, X^2)$ and B^1, B^2 are independent Brownian motions. In the deterministic case, we have a stable equilibrium at zero that is not asymptotically stable. Pick $V(x) = -\ln(|x|^2)$ for $x = (x_1, x_2)$; similarly to the above example this satisfies all the necessary requirements to test for instability, and we see

$$A_K V(x) = x_2 \frac{\partial V(x)}{\partial x_1} - x_1 \frac{\partial V(x)}{\partial x_2} + \frac{1}{2}\sigma^2(x)[\frac{\partial^2 V(x)}{\partial x_1^2} + \frac{\partial^2 V(x)}{\partial x_2^2}].$$

With a bit of calculation we see $A_K V(x) = 0$ whenever $\sigma(x)$ is nonzero for x nonzero, which means we have instability for arbitrarily small positive noise.

So we have seen simple examples where

i) instability becomes stability with enough noise (although this is not "physically feasible"),

ii) stability is not affected by (arbitrarily large) noise, and

iii) stability is destroyed by (arbitrarily small) noise,

which shows the complicated and interesting nature of stochastic stability.

Now we briefly discuss the principle of linearized stability; with the above in mind it should not be surprising that there are quite a lot of difficulties with extracting information about the full system from the linear approximation. So, what can we say about the full system if we know how its linearization acts? For one thing, the full system is stable if the linearized system has constant coefficients and is asymptotically stable. One needs some other concepts like "exponential stability" to say more; interested readers may want to start with [7, Chapter 7]. From this point we abandon Liapunov theory in favor of the "density fluctuation" type of stability theory.

3.5 Markov Semigroup Stability

Of more practical importance to us is the use of Frobenius-Perron operators and the Fokker-Planck equations when dealing with stability of solutions to stochastic differential equations.

Let $(X, \mathcal{A}, \mu)$ be a measure space, let $P := \{P_t\}_{t \geq 0}$ be a stochastic semigroup, and call $D := \{f \in L^1(X) \mid \|f\| = 1, f \geq 0\}$ the set of densities.

Definition 3.21. *We say $f^* \in D$ is an invariant density for P (also called a stationary density) if $P_t f^* = f^*$ for all $t \geq 0$.*

When P is obvious, we may just say that f^* is an invariant density.

Definition 3.22. *We say P is asymptotically stable if P has a unique invariant density f^*, and if, for all $f \in D$,*

$$\lim_{t \to \infty} \|P_t f - f^*\|_{L^1(X)} = 0.$$

The analog to instability is called sweeping.

Definition 3.23. *We say that P is sweeping with respect to a set $A \in \mathcal{A}$ if, for all $f \in D$,*

$$\lim_{t \to \infty} \int_A P_t f(x) dx = 0.$$

Given some σ-algebra $\mathcal{F} \subset \mathcal{A}$ of X, if P is sweeping for all $A \in \mathcal{F}$, then we say it is sweeping with respect to $\mathcal{F}$.

When the context is clear, we usually just say that a semigroup is sweeping.

Of particular interest are stochastic semigroups that are kernel operators (when $(X, \mathcal{A}, \mu) := (\mathbb{R}^n, \mathcal{B}^n, \lambda^n)$).

Definition 3.24. *We say P is a stochastic semigroup of kernel operators (on $\mathbb{R}^n$) if for any $x \in \mathbb{R}^n$, $t \in \mathbb{R}_+$, and $f \in D$,*

$$P_t f(x) = \int_{\mathbb{R}^n} K(t, x, y) f(y) dy,$$

where $K := K(t, x, y) : \mathbb{R}_+ \times \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_+$ is a (stochastic) kernel, in the sense that

$$\int_{\mathbb{R}^n} P_t f(x) dx = 1.$$

Stochastic semigroups of kernel operators will correspond to a semigroup of Frobenius-Perron operators associated to a Fokker-Planck equation having a fundamental solution; for the remainder of the section, let the hypotheses of Theorem 3.2 be satisfied (so a_{ij} and b_i are Cauchy-regular for (3.18)) and call $P := \{P_t\}_{t \geq 0}$ the stochastic Frobenius-Perron semigroup associated to (3.18).

We emulate the Liapunov-type stability theory by again appealing to A_K.

Definition 3.25. *Let $V \in C^2(\mathbb{R}^n)$ be nonnegative, let $\lim_{|x| \to \infty} V(x) = \infty$, and let there exist constants γ, δ such that $V(x), |\frac{\partial V(x)}{\partial x_i}|$, and $|\frac{\partial^2 V(x)}{\partial x_i \partial x_j}|$ are all bounded by $\gamma e^{\delta |x|}$, for $1 \leq i, j \leq n$. If in addition, there exist positive constants α and β such that V satisfies*

$$A_K V(x) \leq -\alpha V(x) + \beta,$$

then we call V Markovian-Liapunov (ML).

The next theorem is quite natural; a proof can be found in ([11, Theorem 11.9.1]).

Theorem 3.7. *P (associated to (3.18)) is asymptotically stable if there exists a ML function V.*

When P is asymptotically stable we can determine the invariant density u_*; since u_* does not change in time, then u_* is the unique density that satisfies the special case of (3.18):

$$\frac{1}{2}\sum_{i,j=1}^{d}\frac{\partial^2}{\partial x_i \partial x_j}(a_{ij}u_*) - \sum_{i=1}^{d}\frac{\partial}{\partial x_i}(b_i u_*) = 0.$$

Next we deal with the conditions under which P is sweeping; in this context it is understood that we are considering sweeping from the family of compact subsets $\mathcal{B}_c$ of $\mathbb{R}^n$. In other words, if for all $A \in \mathcal{B}_c$ and for all $f \in D$,

$$\lim_{t\to\infty}\int_A P_t f(x)dx = \lim_{t\to\infty}\int_A u(t,x)dx = 0,$$

then P is sweeping.

Definition 3.26. *Let $V \in C^2(\mathbb{R}^n)$ be positive and let there exist constants γ, δ such that $V(x), |\frac{\partial V(x)}{\partial x_i}|$, and $|\frac{\partial^2 V(x)}{\partial x_i \partial x_j}|$ are all bounded by $\gamma e^{\delta|x|}$. If in addition, there exists a positive constant α such that V satisfies*

$$A_K V(x) \leq -\alpha V(x), \tag{3.26}$$

then we call V a Bielecki function.

The proof of the next theorem can be found in [11, Theorem 11.11.1].

Theorem 3.8. *P (associated to (3.18)) is sweeping if there exists a Bielecki function V.*

Example 3.3.

One very simple example in one dimension is $(\sigma, -bx)$ with initial condition $X_0 = X^0$ a.s., where X^0 has density f and σ and b are positive constants. We have already explicitly solved this; recall that the solution is

$$X_t = e^{-bt}X^0 + \sigma \int_0^t e^{b(s-t)} dB_s.$$

Trying to use Liapunov theory as before proves fruitless, as the trivial solution would have to have $\sigma := 0$. However, we can see that the expected value of this process at any time t is $E(X_t) = e^{-bt}E(X^0)$, and that the variance $V(X_t)$ is $e^{-2bt}V(X^0) + \sigma^2 \int_0^t e^{2b(s-t)} ds$; if time goes to infinity then $V(X_t)$ goes to $\frac{\sigma^2}{2b}$ and $E(X_t)$ goes to 0. Thus we should see some kind of asymptotic stability with a limiting density exhibiting the same kind of variance; a natural guess is a Gaussian density centered at zero with variance $\frac{\sigma^2}{2b}$.

Pick $V(x) = x^2$; observe that V is ML since

$$A_K V(x) = \frac{1}{2}(\sigma^2)(2) + (-bx)(2x) \leq -\alpha x^2 + \beta$$

is satisfied when $\alpha := 2b$ and $\beta := \sigma^2$. Hence P is asymptotically stable and the limiting density satisfies $A_{FP}u_* = 0$, or

$$\frac{1}{2}(\sigma^2 u_*(x))'' - (-bxu_*(x))' = 0,$$

and this has solution

$$u_*(x) = \sqrt{\frac{b}{\pi\sigma^2}} e^{\frac{-bx^2}{\sigma^2}}.$$

Note that this is a normal density with expected value zero and variance $\frac{\sigma^2}{2b}$, which is consistent with our expectations.

Example 3.4.

To see how sweeping works, we study $dX_t = bX_t dt + \sigma dB_t$ with initial condition $X_0 = X^0$ a.s., where X^0 has density f and σ and b are positive constants. Pick $V(x) = e^{-kx^2}$, for some positive constant k. To see if V is a Bielecki function we need to find a positive α such that

$$\frac{1}{2}\sigma^2 e^{-kx^2}[(4k^2x^2) + (-2k)] + bxe^{-kx^2}(-2kx) \leq -\alpha V(x).$$

A bit of manipulation gives

$$2k((\sigma^2 k - b)x^2 - \sigma^2) \leq -\alpha,$$

and we satisfy this if we take $k := \frac{b}{\sigma^2}$ and $\alpha := b$. Thus the semigroup is sweeping.

Roughly speaking, sweeping and asymptotic stability are the only possibilities; this is the so-called Foguel alternative ([11, Theorem 11.12.1]):

Theorem 3.9. *Let the hypotheses of Theorem 3.2 be satisfied, and let P be the stochastic Frobenius-Perron semigroup associated to (3.18). Suppose all stationary nonnegative solutions to (3.18) take the form $cu_*(x)$, where $u_* > 0$ a.e. and c is a nonnegative constant, and call*

$$I := \int_{\mathbb{R}^n} u_*(x)dx. \tag{3.27}$$

If $I < \infty$, P is asymptotically stable; if $I = \infty$, P is sweeping.

This makes sense; some normalized version of u_* would be the exact limiting density, provided u_* had a finite integral.

We now give a template in one dimension of how to utilize the Foguel alternative. Consider

$$dX_t = b(X_t)dt + \sigma(X_t)dt,$$

where $a(x) = \sigma^2(x)$ and $b(x)$ are Cauchy-regular.

The Fokker-Planck equation takes the form

$$\frac{1}{2}(\sigma^2(x)u_*(x))'' - (b(x)u_*(x))' = 0,$$

or, writing $z(x) = \sigma^2(x)u_*(x)$,

$$\frac{dz}{dx} = \frac{2b(x)}{\sigma^2(x)}z + c_1,$$

for c_1 a constant. Then, if $e^{\int_0^x B(y)dy}$ makes sense, where $B(y) := \frac{2b(y)}{\sigma^2(y)}$, we get, for c_2 a constant,

$$z(x) = e^{\int_0^x B(y)dy}\left(c_2 + c_1\int_0^x e^{\int_0^y -B(z)dz}dy\right).$$

We only care about the a.s. positive stationary solutions for the application of the Foguel alternative, so it is enough to examine the sign of $c_2 + c_1\int_0^x(e^{\int_0^y -B(z)dz})dy$ for almost every x.

If we assume that $xb(x) \leq 0$ for all $|x| \geq r$, for r a positive constant (so $[-r, r]$ is not repelling for trajectories of $\dot{x} = b(x)$), then (according to Maple) $\int_0^x(e^{\int_0^y -B(z)dz})dy \to -\infty$ when $x \to -\infty$; this means z cannot be positive for every x unless $c_1 = 0$, and thus, the

stationary nonnegative solutions must take the form

$$u_*(x) = \frac{1}{\sigma^2(x)} c_2 e^{\int_0^x B(y)dy}.$$

We now need to check whether $\int_{\mathbb{R}} u_*(x)dx$ is finite or not, which is the same as observing that

$$I := \int_{-\infty}^{\infty} \frac{1}{\sigma^2(x)} e^{\int_0^x B(y)dy} \tag{3.28}$$

is finite or not. If $I < \infty$, P is asymptotically stable, and if $I = \infty$, P is sweeping. We now summarize these results:

Corollary 3.1. *Assume $a(x) = \sigma^2(x)$ and $b(x)$ are Cauchy-regular for $dX_t = b(X_t)dt + \sigma(X_t)dt$ and assume $xb(x) \leq 0$ for all $|x| \geq r$, for r a positive constant. Then if I in (3.28) is finite, P is asymptotically stable, and if I in (3.28) is infinite, P is sweeping.*

Example 3.5.

Let $\sigma(x) := \sigma$ be a nonzero constant and let $b(x) = -\frac{Kx}{1+x^2}$, for $K \geq 0$ constant. Then

$$B(x) = \frac{-1}{\sigma^2} \int_0^x \frac{2Ky}{1+y^2} dy = -K \ln(1+x^2),$$

and

$$u_*(x) = ce^{\frac{-K}{\sigma^2} \ln(1+x^2)} = \frac{C}{(1+x^2)^\kappa},$$

where $\kappa := \frac{-K}{\sigma^2}$. We see u_* is integrable iff $\frac{K}{\sigma^2} > \frac{1}{2}$, which implies P is asymptotically stable. Also, $0 \leq \frac{K}{\sigma^2} \leq \frac{1}{2}$ implies P is sweeping. In conclusion, the origin is attracting in

the deterministic case, but in the stochastic case, we can calculate the critical amount of noise needed to destroy the asymptotic stability.

Example 3.6.

Let b, σ be positive constants and reconsider the equation

$$dX_t = -bX_t dt + \sigma X_t dB_t,$$

with initial condition $X_0 = X^0$ a.s. (so $b(x) := -bx$ and $\sigma(x) := \sigma x$).

We have already solved this explicitly and observed that, for any degenerate initial condition $X_0 = x$ a.s., the solution will go to zero as time goes to infinity. We also used a stochastic Liapunov function to deduce asymptotic stability. Note that we cannot apply the template; the necessary prerequisites for the template are not satisfied since $a(x) = \sigma^2(x) = \sigma^2 x^2$ is not bounded by any constant M and hence, is not Cauchy-regular.

3.6 Long-time Behavior of a Stochastic Predator-prey Model

This is a summary of "Long-time behaviour of a stochastic prey-predator model" by Rudnicki [16].

We consider the system

$$dX_t = \sigma X_t dB_t + (\alpha X_t - \beta X_t Y_t - \mu X_t^2)dt, \tag{3.29}$$

$$dY_t = \rho Y_t dB_t + (-\gamma Y_t + \delta X_t Y_t - \nu Y_t^2)dt, \tag{3.30}$$

which is a stochastic Lotka-Volterra predator-prey model. In [4], the existence of a solution to (3.29, 3.30) is proven. We interpret the (positive) constant coefficients in the following way: α is the growth rate of the prey in the absence of predators, β is the "predation rate" that kills off the prey, and μ is inversely related to the "carrying capacity" of the prey, in that if the population grows too much, the environment cannot support further growth. We interpret γ as the decay rate of the predator in the absence of prey and δ as the predation rate that causes predator growth. We may also think of ν as the "reciprocal carrying capacity" of the predator. Further, we interpret σ, ρ as "noise terms" like disease or weather fluctuations that would interfere with an ideal model.

Suppose that $\sigma = \rho = 0$ in $(3.29, 3.30)$, so that we are in the deterministic case.

One can compute equilibrium points: $(0,0), (0, -\frac{\gamma}{\nu}), (\frac{\alpha}{\mu}, 0)$, and $(\bar{x}, \bar{y})$, where

$$\bar{x} = \frac{\alpha\nu + \gamma\beta}{\delta\beta + \mu\nu},$$

$$\bar{y} = \frac{\alpha\delta - \mu\gamma}{\delta\beta + \mu\nu}.$$

We observe that $(0,0)$ is unstable, $(0, -\frac{\gamma}{\nu})$ is biologically irrelevant, and $(\frac{\alpha}{\mu}, 0)$ yields 2 cases, namely, stability for $\mu\gamma > \alpha\delta$ and instability for $\mu\gamma < \alpha\delta$. Finally, $(\bar{x}, \bar{y})$ yields 3 cases, namely, it lies in the fourth quadrant and is biologically irrelevant for $\mu\gamma > \alpha\delta$, lies in the first quadrant and is asymptotically stable for $\mu\gamma < \alpha\delta$, and lies on the x-axis for $\mu\gamma = \alpha\delta$.

So how does this relate to the stochastic case? Let us for now sacrifice technicality for intuition, and examine the terms

$$c_1 = \alpha - \frac{\sigma^2}{2}, \qquad c_2 = \gamma + \frac{\rho^2}{2}.$$

These are the "stochastic versions" of α and γ, respectively (which make sense; if there are very large fluctuations in disease, weather, etc., then it could significantly affect birth/death rates). Then conditions like "$\mu\gamma < (>)\alpha\delta$" become "$\mu c_2 < (>) c_1\delta$." We get something analogous in Rudnicki's Theorem 1, namely, if $c_1 < 0$, then the prey die, and so do the predators. If $c_1 > 0$, if we have "$\mu c_2 > c_1\delta$", the predators growth will be negative, and eventually, the predators die out; if we have "$\mu c_2 < \alpha c_1$", then we obtain a "nice" result, that somehow the system reaches a desired level of stability. One can see how large noise in c_1 could reduce the prey's birth rate to below zero, and hence, cause extinction. Without this noise term or predators, the population would converge to a positive equilibrium, but with the noise term, "bad" environmental fluctuations cause extinction (even with no predators!). Similarly, the predators can die if ρ is too large, no matter how the prey acts. The effects of the incorporation of the noise term is in essence a decrease in the prey's birth

rate and an increase in the predator's death rate. This is arguably a sensible refinement, as it is a little idealistic to think that very small populations will always survive; one must expect some role to be played by the unpredictability of nature.

So, equipped with the basic idea, we proceed to make more precise the above by formally stating Rudnicki's main theorem and outlining the strategy of the proof. First, transform (3.29, 3.30) by calling $X_t = e^{\xi_t}$ and $Y_t = e^{\eta_t}$, so we arrive at the main system

$$d\xi_t = \sigma dB_t + (\alpha - \frac{\sigma^2}{2} - \mu e^{\xi_t} - \beta e^{\eta_t})dt, \tag{3.31}$$

$$d\eta_t = \rho dB_t + (-\gamma - \frac{\rho^2}{2} + \delta e^{\xi_t} - \nu e^{\eta_t})dt. \tag{3.32}$$

Let the solution process (ξ_t, η_t) be such that the distribution of the initial value (ξ_0, η_0) is absolutely continuous with density $v(x, y)$. Then (ξ_t, η_t) has density $u(x, y, t)$, where u satisfies the Fokker-Planck equation:

$$\frac{\partial u}{\partial t} = \frac{1}{2}\sigma^2\frac{\partial^2 u}{\partial x^2} + \sigma\rho\frac{\partial^2 u}{\partial x \partial y} + \frac{1}{2}\rho^2\frac{\partial^2 u}{\partial y^2} - \frac{\partial(f_1(x,y)u)}{\partial x} - \frac{\partial(f_2(x,y)u)}{\partial y}, \tag{3.33}$$

where $f_1(x, y) = c_1 - \mu e^x - \beta e^y$, $f_2(x, y) = -c_2 + \delta e^x - \nu e^y$, and where $c_1 = \alpha - \frac{1}{2}\sigma^2$, $c_2 = \gamma + \frac{1}{2}\rho^2 > 0$.

To verify this, it must be shown that the transition probability function for (ξ_t, η_t), which we call $\mathcal{P}(t, x, y, A)$, is absolutely continuous with respect to Lebesgue measure for each (x, y) and $t > 0$. This means that the distribution of any solution is absolutely continuous and has density u satisfying (3.33). This allows us to proceed by studying "fluctuation

of densities", using advanced techniques based on the section on Markov semigroup stability (see [14] and [15]). We now state the paper's main theorem (Theorem 1):

Let (ξ_t, η_t) solve (3.31,3.32). Then for all $t > 0$ the distribution of (ξ_t, η_t) has a density $u(t, x, y)$ satisfying (3.33).

1) If $c_1 > 0$ and $\mu c_2 < \delta c_1$, then there is a unique density u_* which is an asymptotically stable stationary solution of (3.33). This means that, no matter what the initial distribution of (ξ_0, η_0) is, (ξ_t, η_t) converges in distribution to a random variable with density u_*.

2) If $c_1 > 0$ and $\mu c_2 > \delta c_1$, then $\lim_{t\to\infty} \eta_t = -\infty$ a.s. and the distribution of ξ_t converges weakly to the measure with density $f_*(x) = C \exp(\frac{2c_1 x}{\sigma^2} - (\frac{2\mu}{\sigma^2})e^x)$.

3) If $c_1 < 0$, then ξ_t and η_t go to $-\infty$ a.s. as t goes to ∞.

We outline the proof of this theorem by lemmas, introducing notation as necessary: Call $P_t v(x, y) = u(t, x, y)$. Then$\{P_t\}$ is a Markov semigroup corresponding to (3.33) (write (3.33) as $\frac{\partial u}{\partial t} = \mathcal{A}u$. Then $\mathcal{A}$ is the infinitesimal generator of $\{P_t\}$).

Lemma 1: $\{P_t\}_{t\geq 0}$ is an integral Markov semigroup with a continuous kernel k.

In fact, $k = k(t, x, y; x_0, y_0) \in C^\infty(\mathbb{R}^+ \times \mathbb{R}^2 \times \mathbb{R}^2)$ is the density of $\mathcal{P}(t, x_0, y_0, \cdot)$, so that

$$P_t v(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} k(t, x, y; \xi, \eta) v(\xi, \eta) d\xi d\eta \tag{3.34}$$

is the integral representation of $\{P_t\}$. The Hormander condition is verified to prove that a density exists.

We will need that k is positive to apply some "Foguel alternative type" results; the basic idea is to find some set that is an attractor and realize that k is positive on that set (which is all that is needed). To this end, a method based on support theorems is introduced, and we get

Lemma 2: For each $(x_0, y_0) \in E$ and for almost every $(x, y) \in E$, there exists $T > 0$ such that $k(T, x, y; x_0, y_0) > 0$, where

i) $E = \mathbb{R}^2$ if $\sigma > \rho$ or $\beta\rho \geq \nu\sigma$,

ii) $E = E(M_0) = \{(x, y) | y < (\frac{\rho}{\sigma})x + M_0\}$, where M_0 is the smallest number such that $(f_1, f_2) \cdot [\rho, \sigma] \geq 0$ for $(x, y) \notin E(M_0)$, if $\sigma \geq \rho$ and $\beta\rho < \nu\sigma$.

So, in the case of i) the invariant density u_* is positive everywhere, while in the case of ii) we have a smaller support. If i) we can use the following result:

If an integral Markov semigroup has only one invariant density that is a.e. positive, then the semigroup is asymptotically stable. Also, if there is no invariant density, the semigroup is sweeping from compact sets (or simply "sweeping").

However, if ii) holds, the situation is more delicate, and we must insure that, a.e., for any $t > 0, f \in D$,

$$\int_0^\infty P_t f dt > 0 \tag{3.35}$$

in order to yield that the (integral Markov) semigroup is either asymptotically stable or sweeping (also called the Foguel alternative). In fact, in the case of ii) one can show

Lemma 3: In the situation of Lemma 2 ii),

$$\lim_{t \to \infty} \int \int_E P_t f(x, y) dx dy = 1. \tag{3.36}$$

Now we have

Lemma 4: $\{P_t\}$ is either sweeping or asymptotically stable.

Of course, one would like to know which one is happening, so naturally the next result is

Lemma 5: If $c_1 > 0$ and $\mu c_2 < \delta c_1$ then $\{P_t\}$ is asymptotically stable.

The proof of this lemma relies upon the construction of a Khasminskii function, the existence of which precludes sweeping. This yields Theorem 1 i).

For Theorem 1 ii) and iii), recall that, for equation (σ, b) and its solution X_t, if we define

$$s(x) = \int_0^x \exp(-\int_0^y \frac{2b(r)}{\sigma^2(r)})drdy, \tag{3.37}$$

then $s(-\infty) > -\infty$ and $s(\infty) = \infty$ implies $\lim_{t\to\infty} X_t = -\infty$. From this fact (and a bit of ergodic theory) it is simple to derive Lemmas 6 and 7, which are Theorem 1 iii) and ii), respectively.

Bibliography

[1] H. Amann, Ordinary Differential Equations, de Gruyter, Berlin & New York, 1990.

[2] L. Arnold, Random Dynamical Systems, Springer, Berlin & New York, 1998

[3] H. Bauer, Probability Theory, de Gruyter, Berlin & New York, 1996.

[4] Chessa, S. and Fujita Y. H. (2002). The stochastic equation of predator-prey population dynamics. Boll. Unione Mat. Ital. Sez. B. Artic. Ric. Mat. 5, 789–804

[5] Friedlin, M. and Wentzell, A. Random Perturbations of Dynamical Systems, Springer, New York, 1988.

[6] T. Gard, Introduction to Stochastic Differential Equations, Marcel Dekker Inc., New York, 1988.

[7] R.Z. Hasminskii, Stochastic Stability of Differential Equations, Alphen aan den Rijn, Netherlands, 1980.

[8] H. Kocak and J. Hale, Dynamics and Bifurcations, Springer-Verlag, New York, 1991.

[9] O. Kallenberg, Foundations of Modern Probability, Springer-Verlag, New York, 2002.

[10] I. Karatzas, and S. Shreve, Brownian Motion and Stochastic Calculus (Second edition), Springer-Verlag, Berlin & New York, 1991.

[11] A. Lasota, and M. Mackey, Chaos, Fractals and Noise, Springer-Verlag, New York, 1991.

[12] B. Oksendal, Stochastic Differential Equations (Second Edition), Springer-Verlag, Berlin & New York, 1989.

[13] S. Saperstone, Semidynamical Systems in Infinite Dimensional Spaces, Springer-Verlag, New York, 1981.

[14] K. Pichr and R. Rudnicki (2000). Continuous Markov semigroups and stability of transport equations. J. Math. Anal. Appl. 249 (2000), pp. 668685.

[15] R. Rudnicki (1995). On asymptotic stability and sweeping for Markov operators. Bull. Polish Acad.: Math. 43 (1995), pp. 245262.

[16] Rudnicki, R. (2003). Long-time behaviour of a stochastic prey-predator model. Stoch. Process. Appl. 108, 93–107.

[17] I. Vrabie, C_0-Semigroups and Applications, Elsevier, Boston, 2003.