Ergodic Properties of Markov Processes

Luc Rey-Bellet

Department of Mathematics and Statistics, University of Massachusetts,
Amherst, MA 01003, USA
e-mail: lr7q@math.umass.edu

1 Introduction

In these notes we discuss Markov processes, in particular stochastic differential equations (SDE) and develop some tools to analyze their long-time behavior. There are several ways to analyze such properties, and our point of view will be to use systematically Liapunov functions which allow a nice characterization of the ergodic properties. In this we follow, at least in spirit, the excellent book of Meyn and Tweedie [7]. In general a Liapunov function W is a positive function which grows at infinity and satisfies an inequality involving the generator of the Markov process L: roughly speaking we have the implications (α and β are positive constants)

1. $LW \le \alpha + \beta W$ implies existence of solutions for all times.
2. $LW \le -\alpha$ implies the existence of an invariant measure.
3. $LW \le \alpha - \beta W$ implies exponential convergence to the invariant. measure

For (2) and (3), one should assume in addition, for example smoothness of the transition probabilities (i.e the semigroup e^{tL} is smoothing) and irreducibility of the process (ergodicity of the motion). The smoothing property for generator of SDE's is naturally linked with hypoellipticity of L and the irreducibility is naturally expressed in terms of control theory.

In sufficiently simple situations one might just guess a Liapunov function. For interesting problems, however, proving the existence of a Liapunov functions requires both a good guess and a quite substantial understanding of the dynamics. In these notes we will discuss simple examples only and in the companion lecture [11] we will apply these techniques to a model of heat conduction in anharmonic lattices. A simple set of equations that the reader should keep in mind here are the Langevin equations

$$\begin{aligned} dq &= pdt\,, \\ dp &= (-\nabla V(q) - \lambda p)dt + \sqrt{2\lambda T} dB_t\,, \end{aligned}$$

where, $p, q \in \mathbf{R}^n$, $V(q)$ is a smooth potential growing at infinity, and B_t is Brownian motion. This equation is a model a particle with Hamiltonian $p^2/2 + V(q)$ in contact with a thermal reservoir at temperature T. In our lectures on open classical systems [11] we will show how to derive similar and more general equations from Hamiltonian dynamics. This simple model already has the feature that the noise is degenerate by which we mean that the noise is acting only on the p variable. Degeneracy (usually even worse than in these equations) is the rule and not the exception in mechanical systems interacting with reservoirs.

The notes served as a crash course in stochastic differential equations for an audience consisting mostly of mathematical physicists. Our goal was to provide the reader with a short guide to the theory of stochastic differential equations with an emphasis long-time (ergodic) properties. Some proofs are given here, which will, we hope, give a flavor of the subject, but many important results are simply mentioned without proof.

Our list of references is brief and does not do justice to the very large body of literature on the subject, but simply reflects some ideas we have tried to conveyed in these lectures. For Brownian motion, stochastic calculus and Markov processes we recommend the book of Oksendal [10], Kunita [15], Karatzas and Shreve [3] and the lecture notes of Varadhan [13, 14]. For Liapunov function we recommend the books of Has'minskii [2] and Meyn and Tweedie [7]. For hypoellipticity and control theory we recommend the articles of Kliemann [4], Kunita [6], Norris [8], and Stroock and Varadhan [12] and the book of Hörmander [1].

2 Stochastic Processes

A *stochastic process* is a parametrized collection of random variables

$$\{x_t(\omega)\}_{t \in T} \tag{1}$$

defined on a probability space $(\tilde{\Omega}, \mathcal{B}, \mathbf{P})$. In these notes we will take $T = \mathbf{R}^+$ or $T = \mathbf{R}$. To fix the ideas we will assume that x_t takes value in $X = \mathbf{R}^n$ equipped with the Borel σ-algebra, but much of what we will say has a straightforward generalization to more general state space. For a fixed $\omega \in \tilde{\Omega}$ the map

$$t \mapsto x_t(\omega) \tag{2}$$

is a *path* or a *realization* of the stochastic process, i.e. a random function from T into $\mathbf{R}^n$. For fixed $t \in T$

$$\omega \mapsto x_t(\omega) \tag{3}$$

is a random variable ("the state of the system at time t"). We can also think of $x_t(\omega)$ as a function of two variables (t, ω) and it is natural to assume that $x_t(\omega)$ is jointly measurable in (t, ω). We may identify each ω with the corresponding path $t \mapsto x_t(\omega)$ and so we can always think of $\tilde{\Omega}$ as a subset of the set $\Omega = (\mathbf{R}^n)^T$ of all functions from T into $\mathbf{R}^n$. The σ-algebra $\mathcal{B}$ will then contain the σ-algebra $\mathcal{F}$ generated by sets of the form

$$\{\omega\,;\, x_{t_1}(\omega) \in F_1, \cdots, x_{t_n}(\omega) \in F_n\}\ , \tag{4}$$

where F_i are Borel sets of $\mathbf{R}^n$. The σ-algebra $\mathcal{F}$ is simply the Borel σ-algebra on Ω equipped with the product topology. From now on we take the point of view that a stochastic process is a probability measure on the measurable (function) space $(\Omega, \mathcal{F})$.

One can seldom describe explicitly the full probability measure describing a stochastic process. Usually one gives the *finite-dimensional* distributions of the process x_t which are probability measures $\mu_{t_1, \cdots, t_k}$ on $\mathbf{R}^{nk}$ defined by

$$\mu_{t_1, \cdots, t_k}(F_1 \times \cdots \times F_k) \;=\; \mathbf{P}\{x_{t_1} \in F_1\,, \cdots, x_{t_k} \in F_k\}\ , \tag{5}$$

where $t_1, \cdots, t_k \in T$ and the F_i are Borel sets of $\mathbf{R}^n$.

A useful fact, known as Kolmogorov Consistency Theorem, allows us to construct a stochastic process given a family of compatible finite-dimensional distributions.

Theorem 2.1. (Kolmogorov Consistency Theorem) *For $t_1, \cdots, t_k \in T$ and $k \in \mathbf{N}$ let $\mu_{t_1, \cdots, t_k}$ be probability measures on $\mathbf{R}^{nk}$ such that*

1. *For all permutations σ of $\{1, \cdots, k\}$*

$$\mu_{t_{\sigma(1)}, \cdots, t_{\sigma(k)}}(F_1 \times \cdots \times F_k) \;=\; \mu_{t_1, \cdots, t_k}(F_{\sigma^{-1}(1)} \times \cdots \times F_{\sigma^{-1}(k)})\,. \tag{6}$$

2. *For all $m \in \mathbf{N}$*

$$\mu_{t_1, \cdots, t_k}(F_1 \times \cdots \times F_k) \;=\; \mu_{t_1, \cdots, t_{k+m}}(F_1 \times \cdots \times F_k \times \mathbf{R}^n \times \cdots \times \mathbf{R}^n)\,. \tag{7}$$

Then there exists a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ and a stochastic process x_t on Ω such that

$$\mu_{t_1, \cdots, t_k}(F_1 \times \cdots \times F_k) \;=\; P\{x_{t_1} \in F_1\,, \cdots, x_{t_k} \in F_k\}\ , \tag{8}$$

for all $t_i \in T$ and all Borel sets $F_i \subset \mathbf{R}^n$.

3 Markov Processes and Ergodic Theory

3.1 Transition probabilities and generators

A Markov process is a stochastic process which satisfies the condition that the future depends only on the present and not on the past, i.e., for any $s_1 \le \cdots \le s_k \le t$ and any measurable sets $F_1, \cdots, F_k$, and F

$$\mathbf{P}\{x_t(\omega) \in F | x_{s_1}(\omega) \in F_1, \cdots, x_{s_k}(\omega) \in F_k\} = \mathbf{P}\{x_t(\omega) \in F | x_{s_k}(\omega) \in F_k\}\,. \tag{9}$$

More formally let $\mathcal{F}_t^s$ be the subalgebra of $\mathcal{F}$ generated by all events of the form $\{x_u(\omega) \in F\}$ where F is a Borel set and $s \le u \le t$. A stochastic process x_t is a *Markov process* if for all Borel sets F, and all $0 \le s \le t$ we have almost surely

$$\mathbf{P}\{x_t(\omega) \in F \,|\, \mathcal{F}_s^0\} \,=\, \mathbf{P}\{x_t(\omega) \in F \,|\, \mathcal{F}_s^s\} \,=\, \mathbf{P}\{x_t(\omega) \in F \,|\, x(s,\omega)\}\,. \tag{10}$$

We will use later an equivalent way of describing the Markov property. Let us consider 3 subsequent times $t_1 < t_2 < t_3$. The Markov property means that for any g bounded measurable

$$E[g(x_{t_3}) | \mathcal{F}_{t_2}^{t_2} \times \mathcal{F}_{t_1}^{t_1}] \,=\, E[g(x_{t_3}) | \mathcal{F}_{t_2}^{t_2}]\,. \tag{11}$$

The time reversed Markov property that for any bounded measurable function f

$$E[f(x_{t_1}) | \mathcal{F}_{t_3}^{t_3} \times \mathcal{F}_{t_2}^{t_2}] \,=\, E[f(x_{t_1}) | \mathcal{F}_{t_2}^{t_2}]\,, \tag{12}$$

which says that the past depends only on the present and not on the future. These two properties are in fact equivalent, since we will show that they are both equivalent to the symmetric condition

$$E[g(x_{t_3}) f(x_{t_1}) | \mathcal{F}_{t_2}^{t_2}] \,=\, E[g(x_{t_3}) | \mathcal{F}_{t_2}^{t_2}] E[f(x_{t_1}) \mathcal{F}_{t_2}^{t_2}]\,, \tag{13}$$

which asserts that given the present, past and future are conditionally independent. By symmetry it is enough to prove

Lemma 3.1. *The relations (11) and (13) are equivalent.*

Proof. Let us fix f and g and let us set $x_{t_i} = x_i$ and $\mathcal{F}_{t_i}^{t^i} \equiv \mathcal{F}_i$, for $i = 1, 2, 3$. Let us assume that Eq. (11) holds and denote by $\hat{g}(x_2)$ the common value of (11). Then we have

$$\begin{aligned} \mathbf{E}[g(x_3) f(x_1) | \mathcal{F}_2] \,&=\, \mathbf{E}\left[\mathbf{E}[g(x_3) f(x_1) | \mathcal{F}_2 \times \mathcal{F}_1] \,|\, \mathcal{F}_2\right] \\ &=\, \mathbf{E}\left[f(x_1) \mathbf{E}[g(x_3) | \mathcal{F}_2 \times \mathcal{F}_1] \,|\, \mathcal{F}_2\right] \,=\, \mathbf{E}\left[f(x_1) \hat{g}(x_2) \,|\, \mathcal{F}_2\right] \\ &=\, \mathbf{E}\left[f(x_1) \,|\, \mathcal{F}_2\right] \hat{g}(x_2) \,=\, \mathbf{E}\left[f(x_1) \,|\, \mathcal{F}_2\right] \mathbf{E}[g(x_3) | \mathcal{F}_2]\,, \end{aligned} \tag{14}$$

which is Eq. (13). Conversely let us assume that Eq. (13) holds and let us denote by $\overline{g}(x_1, x_2)$ and by $\hat{g}(x_2)$ the left side and the right side of (11). Let $h(x_2)$ be any bounded measurable function. We have

$$\mathbf{E}\left[f(x_1)h(x_2)\overline{g}(x_1,x_2)\right] \;=\; \mathbf{E}\left[f(x_1)h(x_2)\mathbf{E}[g(x_3)|\mathcal{F}_2 \times \mathcal{F}_1]\right]$$

$$\begin{aligned} &= \mathbf{E}\left[f(x_1)h(x_2)g(x_3)\right] \;=\; \mathbf{E}\left[h(x_2)\mathbf{E}[f(x_1)g(x_3)\,|\,\mathcal{F}_2]\right] \\ &= \mathbf{E}\left[h(x_2)\left(\mathbf{E}[g(x_3)\,|\,\mathcal{F}_2]\right)\left(\mathbf{E}[f(x_1)\,|\,\mathcal{F}_2]\right)\right] \\ &= \mathbf{E}\left[h(x_2)\hat{g}(x_2)\mathbf{E}[f(x_1)\,|\,\mathcal{F}_2]\right] \;=\; \mathbf{E}\left[f(x_1)h(x_2)\hat{g}(x_2)\right] . \end{aligned} \tag{15}$$

Since f and h are arbitrary this implies that $\overline{g}(x_1,x_2) = \hat{g}(x_2)$ a.s. $\square$

A natural way to construct a Markov process is via a *transition probability function*

$$P_t(x,F)\,, \qquad t \in T\,, \quad x \in \mathbf{R}^n\,, \quad F \text{ a Borel set}\,, \tag{16}$$

where $(t,x) \mapsto P_t(x,F)$ is a measurable function for any Borel set F and $F \mapsto P_t(x,F)$ is a probability measure on $\mathbf{R}^n$ for all (t,x). One defines

$$\mathbf{P}\{x_t(\omega) \in F\,|\,\mathcal{F}_s^0\} \;=\; P\{x_t(\omega) \in F\,|\,x_s(\omega)\} \;=\; P_{t-s}(x_s(\omega),F)\,. \tag{17}$$

The finite dimensional distribution for a *Markov process starting at x at time* 0 are then given by

$$\begin{aligned} \mathbf{P}\{x_{t_1} \in F\} &= P_{t_1}(x,F_1)\,, \\ \mathbf{P}\{x_{t_1} \in F_1, x_{t_2} \in F_2) &= \int_{F_1} P_{t_1}(x,dx_1)P_{t_2-t_1}(x_1,F_2)\,, \\ &\vdots \\ \mathbf{P}\{x_{t_1} \in F_1, \cdot\cdot, x_{t_k} \in F_k\} &= \int_{F_1}\cdots\int_{F_{k-1}} P_{t_1}(x,dx_1)\cdot\cdot P_{t_k-t_{k-1}}(x_{k-1},F_k)\,. \end{aligned} \tag{18}$$

By the Kolmogorov Consistency Theorem this defines a stochastic process x_t for which $\mathbf{P}\{x_0 = x\} = 1$. We denote $\mathbf{P}_x$ and $\mathbf{E}_x$ the corresponding probability distribution and expectation.

One can also give an *initial distribution* π, where π is a probability measure on $\mathbf{R}^n$ which describe the initial state of the system at $t = 0$. In this case the finite dimensional probability distributions have the form

$$\int_{\mathbf{R}^n}\int_{F_1}\cdots\int_{F_{k-1}} \pi(dx)P_t(x,dx_1)P_{t_2-t_1}(x_1,dx_2)\cdots P_{t_k-t_{k-1}}(x_{k-1},F_k)\,, \tag{19}$$

and we denote $\mathbf{P}_\pi$ and $\mathbf{E}_\pi$ the corresponding probability distribution expectation.

Remark 3.2. We have considered here only time homogeneous process, i.e., processes for which $\mathbf{P}_x\{x_t(\omega) \in F\,|\,x_s(\omega)\}$ depends only on $t-s$. This can generalized this by considering transition functions $P(t,s,x,A)$.

The following property is a immediate consequence of the fact that the future depends only on the present and not on the past.

Lemma 3.3. (Chapman-Kolmogorov equation) *For* $0 \le s \le t$ *we have*

$$P_t(x, A) \;=\; \int_{\mathbf{R}^n} P_s(x, dy) P_{t-s}(y, A)\,. \tag{20}$$

Proof. : We have

$$\begin{aligned} P_t(x, A) \;&=\; P\{x_0 = x, x_t \in A\} \;=\; P\{x_0 = x, x_s \in \mathbf{R}^n\,,\, x_t \in A\} \\ &=\; \int_{\mathbf{R}^n} P_s(x, dy) P_{t-s}(y, A)\,. \end{aligned} \tag{21}$$

□

For a measurable function $f(x)$, $x \in \mathbf{R}^n$, we have

$$\mathbf{E}_x[f(x_t)] \;=\; \int_{\mathbf{R}^n} P_t(x, dy) f(y)\,. \tag{22}$$

and we can associate to a transition probability a linear operator acting on measurable function by

$$T_t f(x) \;=\; \int_{\mathbf{R}^n} P_t(x, dy) f(y) \;=\; E_x[f(x_t)]\,. \tag{23}$$

From the Chapman-Kolmogorov equation it follows immediately that T_t is a semi-group: for all $s, t \ge 0$ we have

$$T_{t+s} \;=\; T_t T_s\,. \tag{24}$$

We have also a dual semigroup acting on σ-finite measures on $\mathbf{R}^n$:

$$S_t \mu(A) \;=\; \int_{\mathbf{R}^n} \mu(dx) P_t(x, A)\,. \tag{25}$$

The semigroup T_t has the following properties which are easy to verify.

1. T_t preserves the constant, if $1(x)$ denotes the constant function then

$$T_t 1(x) \;=\; 1(x)\,. \tag{26}$$

2. T_t is positive in the sense that

$$T_t f(x) \;\ge\; 0 \quad \text{if} \quad f(x) \;\ge\; 0\,. \tag{27}$$

3. T_t is a contraction semigroup on $L^\infty(dx)$, the set of bounded measurable functions equipped with the sup-norm $\|\cdot\|_\infty$.

$$\begin{aligned} \|T_t f\|_\infty \;&=\; \sup_x \Big| \int_{\mathbf{R}^n} P_t(x, dy) f(y) \Big| \\ &\le\; \sup_y |f(y)| \sup_x \int)\mathbf{R}^n P_t(x, dy) \;=\; \|f\|_\infty\,. \end{aligned} \tag{28}$$

The spectral properties of the semigroup T_t are important to analyze the long-time (ergodic) properties of the Markov process x_t. In order to use method from functional analysis one needs to define these semigroups on function spaces which are more amenable to analysis than the space of measurable functions.

We say that the semigroup T_t is *weak-Feller* if it maps the set of bounded continuous function $\mathcal{C}^b(\mathbf{R}^n)$ into itself. If the transition probabilities $P_t(x, A)$ are stochastically continuous, i.e., if $\lim_{t\to 0} P_t(x, B_\epsilon(x)) = 1$ for any $\epsilon > 0$ ($B_\epsilon(x)$ is the ϵ-neighborhood of x) then it is not difficult to show that $\lim_{t\to 0} T_t F(x) = f(x)$ for any $f(x) \in \mathcal{C}^b(\mathbf{R}^n)$ (details are left to th reader) and then T_t is a contraction semigroup on $\mathcal{C}^b(\mathbf{R}^n)$.

We say that the semigroup T_t is *strong-Feller* if it maps bounded measurable function into continuous function. This reflects the fact that T^t has a "smoothing effect". A way to show the strong-Feller property is to establish that the transition probabilities $P_t(x, A)$ have a density

$$P_t(x, dy) \;=\; p_t(x, y)dy \,, \tag{29}$$

where $p_t(x, y)$ is a sufficiently regular (e.g. continuous or differentiable) function of x, y and maybe also of t. We will discuss some tools to prove such properties in Section 7.

If T_t is weak-feller we define the *generator* L of T_t by

$$Lf(x) \;=\; \lim_{t\to 0} \frac{T_t f(x) - f(x)}{t} \,. \tag{30}$$

The domain of definition of L is set of all f for which the limit (30) exists for all x.

3.2 Stationary Markov processes and Ergodic Theory

We say that a stochastic process is *stationary* if the finite dimensional distributions

$$\mathbf{P}\{x_{t_1+h} \in F_1, \cdots, x_{t_k+h} \in F_k\} \tag{31}$$

are independent of h, for all $t_1 < \cdots < t_k$ and all measurable F_i. If the process is Markovian with initial distribution $\pi(dx)$ then (take $k = 1$)

$$\int_{\mathbf{R}^n} \pi(dx) P_t(x, F) \;=\; S_t\pi(F) \tag{32}$$

must be independent of t for any measurable F, i.e., we must have

$$S_t\pi \;=\; \pi \,, \tag{33}$$

for all $t \geq 0$. The condition (33) alone implies stationarity since it implies that

$$\begin{aligned} &\mathbf{P}_\pi\{x_{t_1+h} \in F_1, \cdots, x_{t_k+h} \in F_k\} \\ &= \int_{\mathbf{R}^n} \int_{F_1} \cdots \int_{F_{k-1}} \pi(dx) P_{t_1+h}(x, dx_1) \cdots P_{t_k - t_{k-1}}(x_{k-1}, F_k) \,, \\ &= \int_{F_1} \cdots \int_{F_{k-1}} \pi(dx) P_{t_1}(x, dx_1) \cdots P_{t_k - t_{k-1}}(x_{k-1}, F_k) \,, \end{aligned} \tag{34}$$

which is independent of h.

Intuitively stationary distribution describe the long-time behavior of x_t. Indeed let us suppose that the distribution of x_t with initial distribution μ converges in some sense to a distribution $\gamma = \gamma_\mu$ (a priori γ may depend on the initial distribution μ), i.e.,

$$\lim_{t\to\infty} \mathbf{P}_\mu\{x_t \in F\} = \gamma_\mu(F)\,, \tag{35}$$

for all measurable F. Then we have, formally,

$$\begin{aligned}\gamma_\mu(F) &= \lim_{t\to\infty} \int_{\mathbf{R}^n} \mu(dx) P_t(x,F) \\ &= \lim_{t\to\infty} \int_{\mathbf{R}^n} \mu(dx) \int_{\mathbf{R}^n} P_{t-s}(x,dy) P_s(y,F) \\ &= \int \gamma_\mu(dy) \int P_s(y,F) = S_s\gamma_\mu(F)\,,\end{aligned} \tag{36}$$

i.e., γ_μ is a stationary distribution.

In order to make this more precise we recall some concepts and results from ergodic theory. Let $(X, \mathcal{F}, \mu)$ be a probability space and ϕ_t, $t \in \mathbf{R}$ a group of measurable transformations of X. We say that ϕ_t is *measure preserving* if $\mu(\phi_{-t}(A)) = \mu(A)$ for all $t \in \mathbf{R}$ and all $A \in \mathcal{F}$. We also say that μ is an invariant measure for ϕ_t. A basic result in ergodic theory is the pointwise Birkhoff ergodic theorem.

Theorem 3.4. (Birkhoff Ergodic Theorem) *Let ϕ_t be a group of measure preserving transformations of $(X, \mathcal{F}, \mu)$. Then for any $f \in L^1(\mu)$ the limit*

$$\lim_{t\to\infty} \frac{1}{t} \int_0^t f(\phi_s(x))\, ds = f^*(x) \tag{37}$$

exists μ-a.s. The limit $f^*(x)$ is ϕ_t invariant, $f(\phi^t(x)) = f(x)$ for all $t \in \mathbf{R}$, and $\int_X f\, d\mu = \int_X f^*\, d\mu$.

The group of transformation ϕ_t is said to be *ergodic* if $f^*(x)$ is constant μ-a.s. and in that case $f^*(x) = \int f\, d\mu$, μ-a.s. Ergodicity can be also expressed in terms of the σ-field of invariant subsets. Let $\mathcal{G} \subset \mathcal{F}$ be the σ-field given by $\mathcal{G} = \{A \in \mathcal{F} \,:\, \phi^{-t}(A) = A \text{ for all } t\}$. Then in Theorem 3.4 $f^*(x)$ is given by the conditional expectation

$$f^*(x) = E[f\,|\mathcal{G}]\,. \tag{38}$$

The ergodicity of ϕ_t is equivalent to the statement that $\mathcal{G}$ is the trivial σ-field, i.e., if $A \in \mathcal{G}$ then $\mu(A) = 0$ or 1.

Given a measurable group of transformation ϕ_t of a measurable space, let us denote by $\mathcal{M}$ the set of invariant measure. It is easy to see that $\mathcal{M}$ is a convex set and we have

Proposition 3.5. *The probability measure μ is an extreme point of $\mathcal{M}$ if and only if μ is ergodic.*

Proof. Let us suppose that μ is not extremal. Then there exists μ_1, $\mu_2 \in \mathcal{M}$ with $\mu_1 \neq \mu_2$ and $0 < a < 1$ such that $\mu = a\mu_1 + (1-a)\mu_2$. We claim that μ is not ergodic. It μ were ergodic then $\mu(A) = 0$ or 1 for all $A \in \mathcal{G}$. If $\mu(A) = 0$ or 1, then $\mu_1(A) = \mu_2(A) = 0$ or $\mu_1(A) = \mu_2(A) = 1$. Therefore μ_1 and μ_2 agree on the σ-field $\mathcal{G}$. Let now f be a bounded measurable function and let us consider the function

$$f^*(x) \;=\; \lim_{t\to\infty} \frac{1}{t} \int_0^t f(\phi_s x)\, ds\,, \tag{39}$$

which is defined on the set E where the limit exists. By the ergodic theorem $\mu_1(E) = \mu_2(E) = 1$ and f^* is measurable with respect to $\mathcal{G}$. We have

$$\int_E f d\mu_i = \int_E f^* d\mu_i\,, \qquad i = 1, 2\,. \tag{40}$$

Since $\mu_1 = \mu_2$ on $\mathcal{G}$, f^* is $\mathcal{G}$-measurable, and $\mu_i(E) = 1$ for $i = 1, 2$, we see that

$$\int_X f d\mu_1 = \int_X f d\mu_2\,. \tag{41}$$

Since f is arbitrary this implies that $\mu_1 = \mu_2$ and this is a contradiction.

Conversely if μ is not ergodic, then there exists $A \in \mathcal{G}$ with $0 < \mu(A) < 1$. Let us define

$$\mu_1(B) \;=\; \frac{\mu(A\cap B)}{\mu(A)}\,, \quad \mu_2(B) \;=\; \frac{\mu(A^c\cap B)}{\mu(A^c)}\,. \tag{42}$$

Since $A \in \mathcal{G}$, it follows that μ_i are invariant and that $\mu = \mu(A)\mu_1 + \mu(A^c)\mu_2$. Thus μ is not an extreme point. □

A stronger property than ergodicity is the property of *mixing* . In order to formulate it we first note that we have

Lemma 3.6. *μ is ergodic if and only if*

$$\lim_{t\to\infty} \frac{1}{t} \int_0^t \mu(\phi_{-s}(A)\cap B) \;=\; \mu(A)\mu(B)\,, \tag{43}$$

for all A, $B \in \mathcal{F}$

Proof. If μ is ergodic, let $f = \chi_A$ be the characteristic function of A in the ergodic theorem, multiply by the characteristic function of B and use the bounded convergence theorem to show that Eq. (43) holds. Conversely let $E \in \mathcal{G}$ and set $A = B = E$ in Eq. (43). This shows that $\mu(E) = \mu(E)^2$ and therefore $\mu(E) = 0$ or 1. □

We say that an invariant measure μ is *mixing* if we have

$$\lim_{t\to\infty} \mu(\phi_{-t}(A)\cap B) \;=\; \mu(A)\mu(B) \tag{44}$$

for all A, $B \in \mathcal{F}$, i.e., we have convergence in Eq. (44) instead of convergence in the sense of Cesaro in Eq. (43).

Mixing can also be expressed in terms of the triviality of a suitable σ-algebra. We define the remote future σ-field, denoted $\mathcal{F}_\infty$, by

$$\mathcal{F}_\infty = \bigcup_{t \geq 0} \phi_{-t}(\mathcal{F}) . \tag{45}$$

Notice that a set $A \in \mathcal{F}_\infty$ if and only if for every t there exists a set $A_t \in \mathcal{F}$ such that $A = \phi_{-t} A_t$. Therefore the σ-field of invariant subsets $\mathcal{G}$ is a sub- σ-field of $\mathcal{F}^\infty$. We have

Lemma 3.7. *μ is mixing if and only if the σ-field $\mathcal{F}_\infty$ is trivial.*

Proof. Let us assume first that $\mathcal{F}_\infty$ is not trivial. There exists a set $A \in \mathcal{F}_\infty$ with $0 < \mu(A) < 1$ or $\mu(A)^2 \neq \mu(A)$ and for any t there exists a set A_t such that $A = \phi_{-t}(A_t)$. If μ were mixing we would have $\lim_{t\to\infty} \mu(\phi_{-t}(A) \cap A) = \mu(A)^2$. On the other hand

$$\mu(\phi_{-t}(A) \cap A) = \mu(\phi_{-t}(A) \cap \phi_{-t}(A_t)) = \mu(A \cap A_t) \tag{46}$$

and this converge to $\mu(A)$ as $t \to \infty$. This is a contradiction.

Let us assume that $\mathcal{F}_\infty$ is trivial. We have

$$\begin{aligned} \mu(\phi_{-t}(A) \cap B) - \mu(A)\mu(B) &= \mu(B \,|\, \phi_{-t}(A))\mu(\phi_{-t}(A)) - \mu(A)\mu(B) \\ &= (\mu(B \,|\, \phi_{-t}(A)) - \mu(B))\, \mu(A) \end{aligned} \tag{47}$$

The triviality of $\mathcal{F}_\infty$ implies that $\lim_{t\to\infty} \mu(B \,|\, \phi_{-t}(A)) = \mu(B)$. □

Given a stationary Markov process with a stationary distribution π one constructs a stationary Markov process with probability measure $\mathbf{P}_\pi$. We can extend this process in a natural way on $-\infty < t < \infty$. The marginal of $\mathbf{P}_\pi$ at any time t is π. Let Θ_s denote the shift transformation on Ω given by $\Theta_s(x_t(\omega)) = x_{t+s}(\omega)$. The stationarity of the Markov process means that Θ_s is a measure preserving transformation of $(\Omega, \mathcal{F}, \mathbf{P}_\pi)$.

In general given transition probabilities $P_t(x, dy)$ we can have several stationary distributions π and several corresponding stationary Markov processes. Let $\tilde{\mathcal{M}}$ denote the set of stationary distributions for $P_t(x, dy)$, i.e.,

$$\tilde{\mathcal{M}} = \{\pi \,:\, S_t \pi = \pi\} . \tag{48}$$

Clearly $\tilde{\mathcal{M}}$ is a convex set of probability measures. We have

Theorem 3.8. *A stationary distribution π for the Markov process with transition probabilities $P_t(x, dy)$ is an extremal point of $\tilde{\mathcal{M}}$ if and only if $\mathbf{P}_\pi$ is ergodic , i.e., an extremal point in the set of all invariant measures for the shift Θ_t.*

Proof. If P_π is ergodic then, by the linearity of the map $\pi \mapsto \mathbf{P}_\pi$, π must be an extreme point of $\mathcal{M}$.

To prove the converse let E be a nontrivial set in the σ-field of invariant subsets. Let $\mathcal{F}_\infty$ denote the far remote future σ-field and $\mathcal{F}^{-\infty}$ the far remote past σ-field which is defined similarly. Let also $\mathcal{F}_0^0$ be the σ-field generated by x_0 (this is the present). An invariant set is both in the remote future $\mathcal{F}_\infty$ as well as in the remote past $\mathcal{F}_\infty$. By Lemma 3.1 the past and the future are conditionally independent given the present. Therefore

$$\mathbf{P}_\pi[E \,|\, \mathcal{F}_0^0] \;=\; \mathbf{P}_\pi[E \cap E \,|\, \mathcal{F}_0^0] \;=\; \mathbf{P}_\pi[E \,|\, \mathcal{F}_0^0]\mathbf{P}_\pi[E \,|\, \mathcal{F}_0^0]\,. \tag{49}$$

and therefore it must be equal either to 0 or 1. This implies that for any invariant set E there exists a measurable set $A \subset \mathbf{R}^n$ such that $E = \{\omega \,:\, x_t(\omega) \in A \text{ for all } t \in \mathbf{R}\}$ up to a set of $\mathbf{P}_\pi$ measure 0. If the Markov process start in A or A^c it does not ever leaves it. This means that $0 < \pi(A) < 1$ and $P_t(x, A^c) = 0$ for π a.e. $x \in A$ and $P_t(x, A) = 0$ for π a.e. $x \in A^c$. This implies that π is not extremal.

Remark 3.9. Theorem 3.8 describes completely the structure of the σ-field of invariant subsets for a stationary Markov process with transition probabilities $P_t(x, dy)$ and stationary distribution π. Suppose that the state space can be partitioned non trivially, i.e., there exists a set A with $0 < \pi(A) < 1$ such that $P_t(x, A) = 1$ for π almost every $x \in A$ and for any $t > 0$ and $P_t(x, A^c) = 1$ for π almost every $x \in A^c$ and for any $t > 0$. Then the event

$$E \;=\; \{\omega\,;\, x_t(\omega) \in A \text{ for all } t \in \mathbf{R}\} \tag{50}$$

is a nontrivial set in the invariant σ-field. What we have proved is just the converse the statement.

We can therefore look at the extremal points of the sets of all stationary distribution, $S_t\pi \;=\; \pi$. Since they correspond to ergodic stationary processes, it is natural to call them ergodic stationary distributions. If π is ergodic then, by the ergodic theorem we have

$$\lim_{t\to\infty} \frac{1}{t}\int_0^t F(\theta_s(x.(\omega))\,ds \;=\; E_\pi\left[F(x.(\omega))\right]\,. \tag{51}$$

for $\mathbf{P}_\pi$ almost all ω. If $F(x.) = f(x_0)$ depends only on the state at time 0 and is bounded and measurable then we have

$$\lim_{t\to\infty} \frac{1}{t}\int_0^t f(x_s(\omega))\,ds \;=\; \int f(x) d\pi(x)\,. \tag{52}$$

for π almost all x and almost all ω. Integrating over ω gives that

$$\lim_{t\to\infty} \frac{1}{t}\int_0^t T_s f(x)\,ds \;=\; \int f(x) d\pi(x)\,. \tag{53}$$

for π almost all x.

The property of mixing is implied by the convergence of the probability measure $P_t(x, dy)$ to $\mu(dy)$. In which sense we have convergence depends on the problem under consideration, and various topologies can be used. We consider here the total variation norm (and variants of it later): let μ be a signed measure on $\mathbf{R}^n$, the *total variation norm* $\|\mu\|$ is defined as

$$\|\mu\| \;=\; \sup_{|f|\le 1} |\mu(f)| = \sup_A \mu(A) - \inf_A \mu(A)\,. \tag{54}$$

Clearly convergence in total variation norm implies weak convergence.

Let us assume that there exists a stationary distribution π for the Markov process with transition probabilities $P_t(x, dy)$ and that

$$\lim_{t\to\infty} \|P_t(x,\cdot) - \pi\| \;=\; 0\,, \tag{55}$$

for all x. The condition (55) implies mixing. By a simple density argument it is enough to show mixing for $E \in \mathcal{F}_s^{-\infty}$ and $F \in \mathcal{F}_\infty^t$. Since $\Theta_{-t}(\mathcal{F}_s^{-\infty}) = \mathcal{F}_{s-t}^{-\infty}$ we simply have to show that as $k = t - s$ goes to ∞, $\mu(E\cap F)$ converges to $\mu(E)\mu(F)$. We have

$$\begin{aligned}
\mu(E)\mu(F) \;&=\; \int_E \left(\int_{\mathbf{R}^n} \mathbf{P}_x(\Theta_{-t_1}F)d\pi(x)\right) d\mathbf{P}_\pi(\omega)\,,\\
\mu(E\cap F) \;&=\; \int_E \left(\int_{\mathbf{R}^n} \mathbf{P}_x(\Theta_{-t_1}F)P_k(x_{s_2}(\omega), dx)\right) d\mathbf{P}_\pi\,,
\end{aligned} \tag{56}$$

and therefore

$$\begin{aligned}
&\mu(E\cap F) - \mu(E)\mu(F)\\
&= \int_E \left(\int_{\mathbf{R}^n} \mathbf{P}_x(\Theta_{-t_1}F)\left(P_k(x_{s_2}(\omega), dx) - \pi(dx)\right)\right) d\mathbf{P}_\pi\,,
\end{aligned} \tag{57}$$

from which we conclude mixing.

4 Brownian Motion

An important example of a Markov process is the Brownian motion. We will take as a initial distribution the delta mass at x, i.e., the process starts at x. The transition probability function of the process has the density $p_t(x, y)$ given by

$$p_t(x,y) \;=\; \frac{1}{(2\pi t)^{n/2}} \exp\left(-\frac{(x-y)^2}{2t}\right)\,. \tag{58}$$

Then for $0 \le t_1 < t_2 < \cdots < t_k$ and for Borel sets F_i we define the finite dimensional distributions by

$$\begin{aligned}
&\nu_{t_1,\ldots,t_x}(F_1 \times \cdots \times F_x)\\
&= \int p_{t_1}(x, x_1)p_{t_2-t_1}(x_1,x_2)\cdots p_{t_x-t_{x-1}}(x_{x-1},x_x)dx_1\;\cdots dx_x\,,
\end{aligned} \tag{59}$$

with the convention

$$p_0(x, x_1) \;=\; \delta_x(x_1)\,. \tag{60}$$

By Kolmogorov Consistency Theorem this defines a stochastic process which we denote by B_t with probability distribution $\mathbf{P}_x$ and expectation $\mathbf{E}_x$. This process is the *Brownian motion starting at* x.

We list now some properties of the Brownian motion. Most proofs are left as exercises (use your knowledge of Gaussian random variables).

(a) The Brownian motion is a *Gaussian process*, i.e., for any $k \geq 1$, the random variable $Z \equiv (B_{t_1}, \cdots, B_{t_k})$ is a $\mathbf{R}^{nk}$-valued normal random variable. This is clear since the density of the finite dimensional distribution (59) is a product of Gaussian (the initial distribution is a degenerate Gaussian). To compute the mean and variance consider the characteristic function which is given for $\alpha \in \mathbf{R}^{nk}$ by

$$\mathbf{E}_x\left[\exp(i\alpha^T Z)\right] \;=\; \exp\left(-\frac{1}{2}\alpha^T C\alpha + i\alpha^T M\right)\,, \tag{61}$$

where

$$M \;=\; \mathbf{E}_x[Z] \;=\; (x, \cdots, x)\,, \tag{62}$$

is the mean of Z and the covariance matrix $C_{ij} = \mathbf{E}_x[Z_i Z_j]$ is given by

$$C \;=\; \begin{pmatrix} t_1\mathbf{I}_n & t_1\mathbf{I}_n & \cdots & t_1\mathbf{I_n} \\ t_1\mathbf{I}_n & t_2\mathbf{I}_n & \cdots & t_2\mathbf{I_n} \\ \vdots & \vdots & \cdots & \vdots \\ t_1\mathbf{I}_n & t_2\mathbf{I}_n & \cdots & t_k\mathbf{I}_n \end{pmatrix}\,, \tag{63}$$

where $\mathbf{I}_n$ is n by n identity matrix. We thus find

$$\mathbf{E}_x[B_t] \;=\; x\,, \tag{64}$$

$$\mathbf{E}_x[(B_t - x)(B_s - x)] \;=\; n\min(t, s)\,, \tag{65}$$

$$\mathbf{E}_x[(B_t - B_s)^2] \;=\; n|t - s|\,, \tag{66}$$

(b) If $B_t = (B_t^{(1)}, \cdots, B_t^{(n)})$ is a m-dimensional Brownian motion, $B_t^{(j)}$ are independent one-dimensional Brownian motions.

(c) The Brownian motion B_t has *independent increments* , i.e., for $0 \leq t_1 < t_2 < \cdots < t_k$ the random variables $B_{t_1}, B_{t_2} - B_{t_1}, \cdots B_{t_k} - B_{t_{k-1}}$ are independent. This easy to verify since for Gaussian random variables it is enough to show that the correlation $\mathbf{E}_x[(B_{t_i} - B_{t_{i-1}})(B_{t_j} - B_{t_{j-1}})]$ vanishes.

(d) The Brownian motion has *stationary increments* , i.e., $B_{t+h} - B_t$ has a distribution which is independent of t. Since it is Gaussian it suffices to check $\mathbf{E}_x[B_{t+h} - B_t] = 0$ and $\mathbf{E}_x[(B_{t+h} - B_t)^2]$ is independent of t.

(d) A stochastic process $\tilde{x}_t$ is called *a modification* of x_t if $\mathbf{P}\left\{x_t = \tilde{x}_t\right\}$ holds for all t. Usually one does not distinguish between a stochastic process and its modification.

However the properties of the paths can depend on the choice of the modification, and for us it is appropriate to choose a modification with particular properties, i.e., the paths are continuous functions of t. A criterion which allows us to do this is given by (another) famous theorem from Kolmogorov

Theorem 4.1. (Kolmogorov Continuity Theorem) *Suppose that there exists positive constants α, β, and C such that*

$$\mathbf{E}[|x_t - x_s|^\alpha] \le C|t-s|^{1+\beta}. \tag{67}$$

Then there exists a modification of x_t such that $t \mapsto x_t$ is continuous a.s.

In the case of Brownian motion it is not hard to verify (use the characteristic function) that we have

$$\mathbf{E}[|B_t - B_s|^4] = 3|t-s|^2, \tag{68}$$

so that the Brownian motion has a continuous version, i.e. we may (and will) assume that $x_t(\omega) \in \mathcal{C}([0,\infty);\mathbf{R}^n)$ and will consider the measure $\mathbf{P}_x$ as a measure on the function space $\mathcal{C}([0,\infty);\mathbf{R}^n)$ (this is a complete topological space when equipped with uniform convergence on compact sets). This version of Brownian motion is called the *canonical* Brownian motion.

5 Stochastic Differential Equations

We start with a few purely formal remarks. From the properties of Brownian motion it follows, formally, that its time derivative $\xi_t = \dot{B}_t$ satisfies $\mathbf{E}[\xi_t] = 0$, $\mathbf{E}[(\xi_t)^2] = \infty$, and $\mathbf{E}[\xi_t \xi_s] = 0$ if $t \neq s$, so that we have formally, $\mathbf{E}[\xi_t \xi_s] = \delta(t-s)$. So, intuitively, $\xi(t)$ models an time-uncorrelated random noise. It is a fact however that the paths of B_t are a.s. nowhere differentiable so that ξ_t cannot be defined as a random process on $(\mathbf{R}^n)^T$ (it can be defined if we allow the paths to be distributions instead of functions, but we will not discuss this here). But let us consider anyway an equation of the form

$$\dot{x}_t = b(x_t) + \sigma(x_t)\dot{B}_t, \tag{69}$$

where, $x \in \mathbf{R}^n$, $b(x)$ is a vector field, $\sigma(x)$ a $n \times m$ matrix, and B_t a m-dimensional Brownian motion. We rewrite it as integral equation we have

$$x_t(\omega) = x_0(\omega) + \int_0^t b(x_s(\omega)ds + \int_0^t \sigma(x_s(\omega))\dot{B}_s ds. \tag{70}$$

Since $\dot{B}_u$ is uncorrelated $x_t(\omega)$ will depend on the present, $x_0(\omega)$, but not on the past and the solution of such equation should be a Markov process. The goal of this chapter is to make sense of such differential equation and derive its properties. We rewrite (69) with the help of differentials as

$$dx_t = b(x_t)dt + \sigma(x_t)dB_t, \tag{71}$$

by which one really means a solution to the integral equation

$$x_t - x_0 = \int_0^t b(x_s)ds + \int_0^t \sigma(x_s)dB_s \,. \tag{72}$$

The first step to make sense of this integral equation is to define *Ito integrals* or *stochastic integrals*, i.e., integrals of the form

$$\int_0^t f(s,\omega)dB_s(\omega)\,, \tag{73}$$

for a suitable class of functions. Since, as mentioned before B_t is nowhere differentiable, it is not of bounded variation and thus Eq. (73) cannot be defined as an ordinary Riemann-Stieljes integral.

We will consider the class of functions $f(t,\omega)$ which satisfy the following three conditions

1. The map $(s,\omega)\mapsto f(s,\omega)$ is measurable for $0\le s\le t$.
2. For $0\le s\le t$, the function $f(s,\omega)$ depends only upon the history of B_s up to time s, i.e., $f(s,\omega)$ is measurable with respect to the σ-algebra $\mathcal{N}_s^0$ generated by sets of the form $\{B_{t_1}(\omega)\in F_1,\cdots,B_{t_k}(\omega)\in F_k\}$ with $0\le t_1<\cdots<t_k\le s$.
3. $\mathbf{E}\left[\int_0^t f(s,\omega)^2 ds\right]<\infty$.

The set of functions $f(s,\omega)$ which satisfy these three conditions is denoted by $\mathcal{V}[0,t]$.

It is natural, in a theory of integration, to start with *elementary functions* of the form

$$f(t,\omega) = \sum_j f(t_j^*,\omega)\mathbf{1}_{[t_j,t_{j+1})}(t)\,, \tag{74}$$

where $t_j^*\in[t_j,t_{j+1}]$. In order to satisfy Condition 2. one chooses the right-end point $t_j^*=t_j$ and we then write

$$f(t,\omega) = \sum_j e_j(\omega)\mathbf{1}_{[t_j,t_{j+1})}(t)\,, \tag{75}$$

and $e_j(\omega)$ is $\mathcal{N}_{t_j}$ measurable. We define the stochastic integral to be

$$\int_0^t f(s,\omega)dB_s(\omega) = \sum_j e_j(\omega)(B_{t_{j+1}}-B_{t_j})\,. \tag{76}$$

This is the *Ito integral*. To extend this integral from elementary functions to general functions, one uses Condition 3. together with the so called Ito isometry

Lemma 5.1. (Ito isometry) *If $\phi(s,\omega)$ is bounded and elementary*

$$\mathbf{E}\left[\left(\int_0^t \phi(s,\omega)dB_s(\omega)\right)^2\right] = \mathbf{E}\left[\int_0^t f(s,\omega)^2 ds\right]\,. \tag{77}$$

Proof. Set $\varDelta B_j = B_{t_{j+1}} - B_{t_j}$. Then we have

$$\mathbf{E}\left[e_i e_j \varDelta B_i \varDelta B_j\right] = \begin{cases} 0 & i \neq j \\ \mathbf{E}[e_j^2](t_{j+1} - t_j) & i = j \end{cases}, \tag{78}$$

using that $e_j e_i \varDelta B_i$ is independent of $\varDelta B_j$ for $j > i$ and that e_j is independent of B_j by Condition 2. We have then

$$\begin{aligned} \mathbf{E}\left[\left(\int_0^t \phi(s,\omega) dB_s(\omega)\right)^2\right] &= \sum_{i,j} \mathbf{E}\left[e_i e_j \varDelta B_i \varDelta B_j\right] \\ &= \sum_j \mathbf{E}\left[e_j^2\right](t_{j+1} - t_j) \\ &= \mathbf{E}\left[\int_0^t f(s,\omega)^2 dt\right]. \end{aligned} \tag{79}$$

□

Using the Ito isometry one extends the Ito integral to functions which satisfy conditions (a)-(c). One first shows that one can approximate such a function by elementary bounded functions, i.e., there exists a sequence $\{\phi_n\}$ of elementary bounded such that

$$\mathbf{E}\left[\int_0^t \left(f(s,\omega) - \phi_n(s,\omega)\right)^2 ds\right] \to 0\,. \tag{80}$$

This is a standard argument, approximate first f by a bounded, and then by a bounded continuous function. The details are left to the reader. Then one defines the stochastic integral by

$$\int_0^t f(s,\omega) dB_s(\omega) = \lim_{n\to\infty} \int_0^t \phi_n(s,\omega) dB_s(\omega)\,, \tag{81}$$

where the limit is the $L^2(P)$-sense. The Ito isometry shows that the integral does not depend on the sequence of approximating elementary functions. It easy to verify that the Ito integral satisfy the usual properties of integrals and that

$$\mathbf{E}\left[\int_0^t f dB_s\right] = 0\,. \tag{82}$$

Next we discuss Ito formula which is a generalization of the chain rule. Let $v(t,\omega) \in \mathcal{V}[0,t]$ for all $t > 0$ and let $u(t,\omega)$ be a measurable function with respect to $\mathcal{N}_t^0$ for all $t > 0$ and such that $\int_0^t |u(s,\omega)|\, ds$ is a.s. finite. Then the *Ito process* x_t is the stochastic integral with differential

$$dx_t(\omega) = u(t,\omega)dt + v(t,\omega)dB_t(\omega)\,. \tag{83}$$

Theorem 5.2. (Ito Formula) *Let x_t be an one-dimensional Ito process of the form (83). Let $g(x) \in \mathcal{C}^2(\mathbf{R})$ be bounded with bounded first and second derivatives. Then $y_t = g(x_t)$ is again an Ito process with differential*

$$dy_t(\omega) = \left(\frac{dg}{dx}(x_t)u(t,\omega)dt + \frac{1}{2}\frac{d^2g}{dx^2}(x_t)v^2(t,\omega)\right)dt + \frac{dg}{dx}(x_t)v(t,\omega)dB_t(\omega)\,.$$

Proof. We can assume that u and v are elementary functions. We use the notations $\Delta t_j = t_{j+1} - t_j$, $\Delta x_j = x_{j+1} - x_j$, and $\Delta g(x_j) = g(x_{j+1}) - g(x_j)$. Since g is $\mathcal{C}^2$ we use a Taylor expansion

$$\begin{aligned} g(x_t) &= g(x_0) + \sum_j \Delta g(x_j) \\ &= g(x_0) + \sum_j \frac{dg}{dx}(x_{t_j})\Delta x_j + \frac{1}{2}\sum_j \frac{d^2g}{d^2x}(x_{t_j})(\Delta x_j)^2 + R_j\,, \end{aligned} \tag{84}$$

where $R_j = o((\Delta x_j)^2)$. For the second term on the r.h.s. of Eq. (84) we have

$$\begin{aligned} \lim_{\Delta t_j \to 0} \sum_j \frac{dg}{dx}(x_{t_j})\Delta x_j &= \int \frac{dg}{dx}(x_s)dx_s \\ &= \int \frac{dg}{dx}(x_s)u(s,\omega)ds + \int \frac{dg}{dx}(x_s)v(s,\omega)dB_s\,. \end{aligned} \tag{85}$$

We can rewrite the third term on the r.h.s. of Eq. (84) as

$$\sum_j \frac{d^2g}{d^2x}(\Delta x_j)^2 = \sum_j \frac{d^2g}{d^2x}\left(u_j^2(\Delta t_j)^2 + 2u_jv_j\Delta t_j\Delta B_j + v_j^2(\Delta B)_j^2\right)\,. \tag{86}$$

The first two terms on the r.h.s. of Eq. (86) go to zero as $\Delta t_j \to 0$. For the first it is obvious while for the second one uses

$$\mathbf{E}\left[\left(\frac{d^2g}{d^2x}(x_{t_j})u_jv_j\Delta t_j\Delta B_j\right)^2\right] = \mathbf{E}\left[\left(\frac{d^2g}{d^2x}(x_{t_j})u_jv_j\right)^2\right](\Delta t_j)^3 \to 0\,, \tag{87}$$

as $\Delta t_j \to 0$. We claim that the third term on the r.h.s. of Eq. (86) converges to

$$\int_0^t \frac{d^2g}{d^2x}(x_s)v^2ds\,, \tag{88}$$

in $L^2(P)$ as $\Delta t_j \to 0$. To prove this let us set $a(t) = \frac{d^2g}{d^2x}(x_t)v^2(t,\omega)$ and $a_i = a(t_i)$. We have

$$\mathbf{E}\left[\left(\sum_j a_j((\Delta B_j)^2 - \Delta t_j)\right)^2\right] = \sum_{i,j}\mathbf{E}\left[a_ia_j((\Delta B_i)^2 - \Delta t_i)((\Delta B_j)^2 - \Delta t_j)\right] \tag{89}$$

If $i < j$, $a_ia_j((\Delta B_i)^2 - \Delta t_i)$ is independent of $((\Delta B_j)^2 - \Delta t_j)$. So we are left with

$$\sum_j \mathbf{E}\left[a_j^2((\Delta B_j)^2 - \Delta t_j)^2\right]$$
$$= \sum_j \mathbf{E}[a_j^2]\,\mathbf{E}[(\Delta B_j)^4 - 2(\Delta B_j)^2\Delta t_j + (\Delta t_j)^2]$$
$$= \sum_j \mathbf{E}[a_j^2](3(\Delta t_j)^2 - 2(\Delta t_j)^2 + (\Delta t_j)^2] \;=\; 2\sum_j \mathbf{E}[a_j^2]\Delta t_j^2\,, \qquad (90)$$

and this goes to zero as Δt_j goes to zero. □

Remark 5.3. Using an approximation argument, one can prove that it is enough to assume that $g \in \mathcal{C}^2$ without boundedness assumptions.

In dimension $n > 1$ one proceeds similarly. Let B_t be a m-dimensional Brownian motion, $u(t,\omega) \in \mathbf{R}^n$, and $v(t,\omega)$ an $n \times m$ matrix and let us consider the Ito differential

$$dx_t(\omega) \;=\; u(t,\omega)dt + v(t,\omega)dB_t(\omega) \qquad (91)$$

then $y_t = g(x_t)$ is a one dimensional Ito process with differential

$$dy_t(\omega) \;=\; \sum_j \left(\frac{\partial g}{\partial x_j}(x_t)u_j(t,\omega) + \frac{1}{2}\sum_{i,j}\frac{\partial^2 g}{\partial x_i \partial x_j}(x_t)(vv^T)_{ij}(t,\omega)\right)dt$$
$$+ \sum_{ij}\frac{\partial g}{\partial x_j}(x_t)v_{ij}(t,\omega)dB_t^{(i)}\,. \qquad (92)$$

We can apply this to a stochastic differential equation

$$dx_t(\omega) \;=\; b(x_t(\omega))dt + \sigma(x_t(\omega))dB_t(\omega)\,, \qquad (93)$$

with

$$u(t,\omega) = b(x_t(\omega))\,, \quad v(t,\omega) = \sigma(x_t(\omega))\,, \qquad (94)$$

provided we can show that existence and uniqueness of the integral equation

$$x_t(\omega) \;=\; x_0 + \int_0^t b(x_s(\omega))ds + \int_0^t \sigma(x_t(\omega))dB_s(\omega)\,. \qquad (95)$$

As for ordinary ODE's, if b and σ are locally Lipschitz one obtains uniqueness and existence of local solutions. If if one requires, in addition, that b and σ are linearly bounded

$$|b(x)| + |\sigma(x)| \;\le\; C(1+|x|)\,, \qquad (96)$$

one obtains global in time solutions. This is proved using Picard iteration, and one obtains a solution x_t with continuous paths, each component of which belongs to $\mathcal{V}[0,T]$, in particular x_t is measurable with respect to $\mathcal{N}_t^0$.

Let us now introduce the probability distribution $\mathbf{Q}_x$ of the solution $x_t = x_t^x$ of (93) with initial condition $x_0 = x$. Let $\mathcal{F}$ be the σ-algebra generated by the random variables $x_t(\omega)$. We define $\mathbf{Q}_x$ by

$$\mathbf{Q}_x\left[x_{t_1} \in F_1, \cdots, x_{t_n} \in F_n\right] = \mathbf{P}\left[\omega\,;\, x_{t_1} \in F_1, \cdots, x_{t_n} \in F_n\right] \tag{97}$$

where $\mathbf{P}$ is the probability law of the Brownian motion (where the Brownian motion starts is irrelevant since only increments matter for x_t). Recall that $\mathcal{N}_t^0$ is the σ-algebra generated by $\{B_s, 0 \le s \le t\}$. Similarly we let $\mathcal{F}_t^0$ the σ-algebra generated by $\{x_s, 0 \le s \le t\}$. The existence and uniqueness theorem for SDE's proves in fact that x_t is measurable with respect to $\mathcal{N}_t$ so that we have $\mathcal{F}_t \subset \mathcal{N}_t$.

We show that the solution of a stochastic differential equation is a Markov process.

Proposition 5.4. (Markov property) *Let f be a bounded measurable function from $\mathbf{R}^n$ to $\mathbf{R}$. Then, for $t, h \ge 0$*

$$\mathbf{E}_x\left[f(x_{t+h})\,|\,\mathcal{N}_t\right] = \mathbf{E}_{x_t(\omega)}\left[f(x_h)\right]\,. \tag{98}$$

Here $\mathbf{E}_x$ denote the expectation w.r.t to $\mathbf{Q}_x$, that is $\mathbf{E}_y\left[f(x_h)\right]$ means $\mathbf{E}\left[f(x_h^y)\right]$ where $\mathbf{E}$ denotes the expectation w.r.t to the Brownian motion measure $\mathbf{P}$.

Proof. Let us write $x_t^{s,x}$ the solution a stochastic differential equation with initial condition $x_s = x$. Because of the uniqueness of solutions we have

$$x_{t+h}^{0,x} = x_{t+h}^{t,x_t}\,. \tag{99}$$

Since x_{t+h}^{t,x_t} depends only x_t, it is measurable with respect to $\mathcal{F}_t^0$. The increments of the Brownian paths over the time interval $[t, t+h]$ are independent of $\mathcal{F}_t^0$, and the b and σ do not depend on t. Therefore

$$\begin{aligned} \mathbf{P}\left[x_{t+h} \in F\,|\,\mathcal{F}_t^0\right] &= \mathbf{P}\left[x_{t+h}^{t,x_t} \in F\,|\,\mathcal{F}_t^0\right] \\ &= \mathbf{P}\left[x_{t+h}^{t,y} \in F\right]\,|_{y=x_t(\omega)} \\ &= \mathbf{P}\left[x_h^{0,y} \in F\right]\,|_{y=x_t(\omega)}\,. \end{aligned} \tag{100}$$

and this proves the claim. □

Since $\mathcal{N}_t^0 \subset \mathcal{F}_t^0$ we have

Corollary 5.5. *Let f be a bounded measurable function from $\mathbf{R}^n$ to $\mathbf{R}$. Then, for $t, h \ge 0$*

$$\mathbf{E}_x\left[f(x_{t+h})\,|\,\mathcal{F}_t^0\right] = \mathbf{E}_{x_t(\omega)}\left[f(x_h)\right]|\,, \tag{101}$$

i.e. x_t is a Markov process.

Proof. Since $\mathcal{F}_t^0 \subset \mathcal{N}_t^0$ we have

$$\begin{aligned}\mathbf{E}_x\left[f(x_{t+h})\,|\,\mathcal{F}_t^0\right] &= \mathbf{E}_x\left[\mathbf{E}_x\left[f(x_{t+h})\,|\,\mathcal{N}_t^0\right]\,|\,\mathcal{F}_t^0\right] \\ &= \mathbf{E}_x\left[\mathbf{E}_{x_t}\left[f(x_h)\right]\,|\,\mathcal{F}_t^0\right] \\ &= \mathbf{E}_{x_t}\left[f(x_h)\right] . \end{aligned} \tag{102}$$

□

Let $f \in \mathcal{C}_0^2$ (i.e. twice differentiable with compact support) and let L be the second order differential operator given by

$$Lf \;=\; \sum_j b_j(x)\frac{\partial f}{\partial x_j}(x_t) + \frac{1}{2}\sum_{i,j} a_{ij}(x)\frac{\partial^2 f}{\partial x_i \partial x_j}(x)\,, \tag{103}$$

with $a_{ij}(x) = (\sigma(x)\sigma(x)^T)_{ij}$. Applying Ito formula to the solution of an SDE with $x_0 = x$, i.e. with $u(t,\omega) = b(x_t(\omega))$ and $v(t,\omega) = \sigma(x_t,\omega)$, we find

$$\begin{aligned}\mathbf{E}_x[f(x_t)] - f(x) &= \mathbf{E}_x\left[\int_0^t Lf(x_s)ds + \sum_{ij}\frac{\partial f}{\partial x_j}(x_s)\sigma_{ji}(x_s)dB_s^{(i)}\right] \\ &= \int_0^t \mathbf{E}_x\left[Lf(x_s)ds\right] . \end{aligned} \tag{104}$$

Therefore

$$Lf(x) \;=\; \lim_{t\to 0}\frac{\mathbf{E}_x\left[f(x_t)\right] - f(x)}{t}\,, \tag{105}$$

i.e., L is the generator of the diffusion x_t. By the semigroup property we also have

$$\frac{d}{dt}T_t f(x) \;=\; LT_f(x)\,, \tag{106}$$

so that L is the generator of the semigroup T_t and its domain contains $\mathcal{C}_0^2$.

Example 5.6. Let $p, q \in \mathbf{R}^n$ and let $V(q) : \mathbf{R}^n \to \mathbf{R}$ be a C^2 function and let B_t be a n-dimensional Brownian motion. The SDE

$$\begin{aligned} dq &= p\,dt\,, \\ dp &= \left(-\nabla V(q) - \lambda^2 p\right)\,dt + \lambda\sqrt{2T}dB_t\,, \end{aligned} \tag{107}$$

has unique local solutions, and has global solutions if $\|\nabla V(q)\| \le C(1+\|q\|)$. The generator is given by the partial differential operator

$$L \;=\; \lambda(T\nabla_p\cdot\nabla_p - p\cdot\nabla_p) + p\cdot\nabla_q - (\nabla_q V(q))\cdot\nabla_p\,. \tag{108}$$

We now introduce a strengthening of the Markov property, the *strong Markov property*. It says that the Markov property still holds provided we replace the time t

by a random time $\tau(\omega)$ in a class called stopping times. Given an increasing family of σ-algebra $\mathcal{M}_t$, a function $\tau : \Omega \to [0, \infty]$ is called a *stopping time* w.r.t to $\mathcal{M}_t$ if

$$\{\omega \,:\, \tau(\omega) \le t\} \in \mathcal{M}_t, \text{ for all } t \ge 0 \,. \tag{109}$$

This means that one should be able to decide whether or not $\tau \le t$ has occurred based on the knowledge of $\mathcal{M}_t$.

A typical example is the *first exit time* of a set U for the solution of an SDE: Let U be an open set and

$$\sigma_U \;=\; \inf\{t > 0\,;\, x_t \notin U\} \tag{110}$$

Then σ^U is a stopping time w.r.t to either $\mathcal{N}_t^0$ or $\mathcal{F}_t^0$.

The Markov property and Ito's formula can be generalized to stopping times. We state here the results without proof.

Proposition 5.7. (Strong Markov property) *Let f be a bounded measurable function from $\mathbf{R}^n$ to $\mathbf{R}$ and let τ be a stopping time with respect to $\mathcal{F}_t^0$, $\tau < \infty$ a.s. Then*

$$\mathbf{E}_x\left[f(x_{\tau+h})\,|\,\mathcal{F}_\tau^0\right] \;=\; \mathbf{E}_{x_\tau}\left[f(x_h)\right]\,, \tag{111}$$

for all $h \ge 0$.

The Ito's formula with stopping time is called *Dynkin's formula*.

Theorem 5.8. (Dynkin's formula) *Let f be C^2 with compact support. Let τ be a stopping time with $\mathbf{E}_x[\tau] < \infty$. Then we have*

$$\mathbf{E}_x\left[f(x_\tau)\right] \;=\; f(x) + \mathbf{E}_x\left[\int_0^\tau LF(x_s)\,ds\right]\,. \tag{112}$$

As a first application of stopping time we show a method to extend local solutions to global solutions for problems where the coefficients of the equation are locally Lipschitz, but not linearly bounded. We call a function $W(x)$ a *Liapunov function* if $W(x) \ge 1$ and

$$\lim_{|x|\to\infty} W(x) = \infty \tag{113}$$

i.e., W has compact level sets.

Theorem 5.9. *Let us consider a SDE*

$$dx_t = b(x_t)dt + \sigma(x_t)dB_t\,, \quad x_0 = x\,, \tag{114}$$

with locally Lipschitz coefficients. Let us assume that there exists a Liapunov function W which satisfies

$$LW \;\le\; cW\,, \tag{115}$$

for some constant c. Then the solution of Eq. (114) is defined for all time and satisfies

$$\mathbf{E}\left[W(x_t)\right] \;\le\; W(x)e^{ct}\,. \tag{116}$$

Proof. Since b and σ are locally Lipschitz we have a local solution $x_t(\omega)$ which is defined at least for small time. We define

$$\tau_n(\omega) \,=\, \inf\{t>0, W(x_t)\ge n\}\,, \tag{117}$$

i.e. τ_n is the first time exits the compact set $\{W\le n\}$. It is easy to see that τ_n is a stopping time. We define

$$\tau_n(t) \,=\, \inf\{\tau_n, t\}\,. \tag{118}$$

We now consider a new process

$$\tilde{x}_t = x_{\tau_n(t)}\,, \tag{119}$$

We have $\tilde{x}_t = x_{\tau_n}$ for all $t>\tau_n$, i.e., $\tilde{x}_t$ is stopped when it reaches the boundary of $\{W\le n\}$. Since τ_n is a stopping time, by Proposition 5.7 and Theorem 5.8, $\tilde{x}_t$ is a Markov process which is defined for all $t>0$. Its Ito differential is given by

$$d\tilde{x}_t \,=\, \mathbf{1}_{\{\tau_n>t\}}b(\tilde{x}_t)dt + \mathbf{1}_{\{\tau_n>t\}}\sigma(\tilde{x}_t)dB_t\,. \tag{120}$$

From Eq. (115) we have

$$(\frac{\partial}{\partial t} + L)We^{-ct} \,\le\, 0\,, \tag{121}$$

and thus

$$\mathbf{E}\left[W(x_{\tau_n(t)})e^{-c\tau_n(t)}\right] - W(x) \,=\, \mathbf{E}\left[\int_0^{\tau_n(t)}(\frac{\partial}{\partial s}+L)W(x_s)e^{-cs}ds\right] \,\le\, 0\,. \tag{122}$$

Since $\tau_n(t)\le t$, we obtain

$$\mathbf{E}\left[W(x_{\tau_n(t)})\right] \,\le\, W(x)e^{ct}\,. \tag{123}$$

On the other hand we have

$$\mathbf{E}\left[W(x_{\tau_n(t)})\right] \,\ge\, \mathbf{E}\left[W(x_{\tau_n(t)})\mathbf{1}_{\tau_n<t}\right] \,=\, n\mathbf{P}_x\{\tau_n<t\} \tag{124}$$

so that we obtain

$$\mathbf{P}_x\{\tau_n<t\} \,\le\, \frac{e^{ct}W(x)}{n} \to 0\,, \tag{125}$$

as $n\to\infty$. This implies that the paths of the process almost surely do not reach infinity in a finite time, if $\tau = \lim_{n\to\infty}\tau_n$ then

$$\mathbf{P}_x\{\tau=\infty\} \,=\, 1\,. \tag{126}$$

Taking the limit $n\to\infty$ in Eq. (123) and using Fatou's lemma gives Eq. (116). □

Example 5.10. Consider the SDE of Example (5.6). If $V(q)$ is of class $\mathcal{C}^2$ and $\lim_{\|q\|\to\infty} V(q) = \infty$, then the Hamiltonian $H(p,q) = p^2/2 + V(q)$ satisfy

$$LH(p,q) \;=\; \lambda(n - p^2) \;\leq\; \lambda n\,. \tag{127}$$

Since H is bounded below we can take $H+c$ as a Liapunov function, and by Theorem 115 the solutions exists for all time.

Finally we mention two important results of Ito calculus (without proof). The first result is a simple consequence of Ito's formula and give a probabilistic description of the semigroup $L - q$ where q is the multiplication operator by a function $q(x)$ and L is the generator of a Markov process x_t. The proof is not very hard and is an application of Ito's formula.

Theorem 5.11. (Feynman-Kac formula) *Let x_t is a solution of a SDE with generator L. If f is $\mathcal{C}^2$ with bounded derivatives and if g is continuous and bounded. Put*

$$v(t,x) \;=\; \mathbf{E}_x\left[e^{-\int_0^t q(x_s)ds} f(x_t)\right]\,. \tag{128}$$

Then, for $t > 0$,

$$\frac{\partial v}{\partial t} \;=\; Lv - qv\,, \quad v(0,x) \;=\; f(x)\,. \tag{129}$$

The second result describe the change of the probability distribution when the drift in an SDE is modified. The proof is more involved.

Theorem 5.12. (Girsanov formula) *Let x_t be the solution of the SDE*

$$dx_t \;=\; b(x_t)dt + \sigma(x_t)dB_t\,, \quad x_0 = x\,, \tag{130}$$

and let y_t be the solution of the SDE

$$dy_t \;=\; a(y_t)dt + \sigma(x_t)dB_t\,, \quad y_0 = x\,. \tag{131}$$

Suppose that there exist a function u such that

$$\sigma(x)u(x) \;=\; b(x) - a(x)\,, \tag{132}$$

and u satisfy Novikov Condition

$$\mathbf{E}\left[\exp\left(\frac{1}{2}\int_0^t u^2(y_t(\omega))\,ds\right)\right] \;<\; \infty\,. \tag{133}$$

Then on the interval $[0,t]$ the probability distribution $\mathbf{Q}_x^{[0,t]}$ of y_t is absolutely continuous with respect to the probability distribution $\mathbf{P}_x^{[0,t]}$ of x_t with a Radon-Nikodym derivative given by

$$d\mathbf{Q}_x^{[0,t]}(\omega) \;=\; e^{-\int_0^t u(y_s)dB_s - \frac{1}{2}\int_0^t u^2(y_s)ds} d\mathbf{P}_x^{[0,t]}(\omega)\,. \tag{134}$$

6 Control Theory and Irreducibility

To study the ergodic properties of Markov process one needs to establish which sets can be reached from x in time t, i.e. to determine when $P_t(x, A) > 0$.

For solutions of stochastic differential equations there are useful tools which from control theory. For the SDE

$$dx_t = b(x_t)dt + \sigma(x_t)dB_t \,, \quad x_0 = x \tag{135}$$

let us replace the Brownian motion B_t by a piecewise polygonal approximation

$$B_t^{(N)} \;=\; B_{k/N} + N(t - \frac{k}{N})(B_{(k+1)/N} - B_{k/N}) \,, \quad \frac{k}{N} \leq t \leq \frac{k+1}{N} \,. \tag{136}$$

Then its time derivative $\dot{B}_t^{(N)}$ is piecewise constant. One can show that the solutions of

$$dx_t^{(N)} = b(x_t^{(N)})dt + \sigma(x_t^{(N)})dB_t^{(N)} \,, \quad x_0 = x \tag{137}$$

converge almost surely to x_t uniformly on any compact interval $[t_1, t_2]$ to the solution of

$$dx_t = b(x_t) + \sigma\sigma'(x_t)dt + \sigma(x_t)dB_t \,, \tag{138}$$

The supplementary term in (138) is absent if $\sigma(x) = \sigma$ is independent of x and is related to Stratonovich integrals. Eq. (137) has the form

$$\dot{x} \;=\; b(x_t) + \sigma u_t \,, \quad x_0 = x \,, \tag{139}$$

where $t \mapsto u(t) = (u_1(t), \cdots, u_m(t))$ is a piecewise constant function. This is an ordinary (non-autonomous) differential equation. The function u is called a *control* and Eq. (139) a control system. The support theorem of Stroock and Varadhan shows that several properties of the SDE Eq. (135) (or (138)) can be studied and expressed in terms of the control system Eq. (139). The control system has the advantage of being a system of ordinary diffential equations.

Let us denote by $\mathcal{S}_x^{[0,t]}$ the support of the diffusion x_t, i.e., $\mathcal{S}_x$ is the smallest closed (in the uniform topology) subset of $\{f \in \mathcal{C}([0,t], \mathbf{R}^n) \,,\, f(0) = x\}$ such that

$$\mathbf{P}\left\{x_s(\omega) \in \mathcal{S}_x^{[0,t]}\right\} \;=\; 1 \,. \tag{140}$$

Note that Girsanov formula, Theorem 5.12 implies that the supports of (135) and (138) are identical.

A typical question of control theory is to determine for example the set of all possible points which can be reached in time t by choosing an appropriate control in a given class. For our purpose we will denote by $\mathcal{U}$ the set of all locally constant functions u. We will say a point y is accessible from x in time t if there exists a control $u \in \mathcal{U}$ such that the solution $x_t^{(u)}$ of the equation Eq. (139) satisfies $x^{(u)}(0) = x$ and $x^{(u)}(t) = y$. We denote by $A_t(x)$ the set of accessible points from x in time t. Further we define $C_x^{[0,t]}(\mathcal{U})$ to be the subset of all solutions of Eq. (139) as u varies in $\mathcal{U}$. This is a subset of $\{f \in \mathcal{C}([0,t], \mathbf{R}^n) \,,\, f(0) = x\}$.

Theorem 6.1. (Stroock-Varadhan Support Theorem)

$$\mathcal{S}_x^{[0,t]} \;=\; \overline{C^{[0,t]}(\mathcal{U})} \tag{141}$$

where the bar indicates the closure in the uniform topology.

As an immediate consequence if we denote $\operatorname{supp}\mu$ the support of a measure μ on $\mathbf{R}^n$ we obtain

Corollary 6.2.

$$\operatorname{supp} P_t(x,\,\cdot\,) \;=\; \overline{A_t(x)}\,. \tag{142}$$

For example if can show that all (or a dense subset) of the points in a set F are accessible in time t, then we have

$$P_t(x,F) > 0\,, \tag{143}$$

that is the probability to reach F from x in the time t is positive.

Example 6.3. Let us consider the SDE

$$dx_t \;=\; b(x_t)\,dt + \sigma dB_t\,, \tag{144}$$

where b is such that there is a unique solution for all times. Assume further that $\sigma \,:\, \mathbf{R}^n \to \mathbf{R}^n$ is invertible. For any $t > 0$ and any $x \in \mathbf{R}^n$, the support of the diffusion $\mathcal{S}_x^{[0,t]} = \{f \in \mathcal{C}([0,t],\mathbf{R}^n)\,,\, f(0) = x\}$ and, for all open set F, we have $P_t(x,F) > 0$. To see this, let ϕ_t be a $\mathcal{C}^1$ path in $\mathbf{R}^n$ such that $\phi_0 = x$ and define define the (smooth) control $u_t \;=\; \sigma^{-1}\left(\dot{\phi}_t - b(\phi_t)\right)$. Clearly ϕ_t is a solution for the control system $\dot{x}_t = b(x_t) + \sigma u_t$. A simple approximation argument shows that any continuous paths can be approximated by a smooth one and then any smooth path can be approximated by replacing the smooth control by a piecewise constant one.

Example 6.4. Consider the SDE

$$\begin{aligned} dq \;&=\; p\,dt\,, \\ dp \;&=\; \left(-\nabla V(q) - \lambda^2 p\right)\,dt + \lambda\sqrt{2T}dB_t\,, \end{aligned} \tag{145}$$

under the same assumptions as in Example (6.4). Given $t > 0$ and two pair of points (q_0,p_0) and (q_t,p_t), let $\phi(s)$ be any $\mathcal{C}^2$ path in $\mathbf{R}^n$ which satisfy $\phi(0) = q_0$, $\phi(t) = q_t$, $\phi'(0) = p_0$ and $\phi'(t) = p_t$. Consider the control u given by

$$u_t \;=\; \frac{1}{\lambda\sqrt{2T}}\left(\ddot{\phi}_t + \nabla V(\phi_t) + \lambda^2\dot{\phi}_t\right)\,. \tag{146}$$

By definition $(\phi_t,\dot{\phi}_t)$ is a solution of the control system with control u_t, so that u_t drives the system from (q_0,p_0) to (q_t,p_t). This implies that $A_t(x,F) = \mathbf{R}^n$, for all $t > 0$ and all $x \in \mathbf{R}^n$. From the support theorem we conclude that $P_t(x,F) > 0$ for all $t > 0$, all $x \in \mathbf{R}^n$, and all open set F.

7 Hypoellipticity and Strong-Feller Property

Let x_t denote the solution of the SDE

$$dx_t \;=\; b(x_t)dt + \sigma(x_t)dB_t\,, \tag{147}$$

and we assume here that $b(x)$ and $\sigma(x)$ are $\mathcal{C}^\infty$ and such that the equation has global solutions.

The generator of the semigroup T_t is given, on sufficiently smooth function, by the second-order differential operator

$$L \;=\; \sum_i b_i(x)\frac{\partial}{\partial x_i} + \frac{1}{2}\sum_{ij} a_{ij}(x)\frac{\partial^2}{\partial x_i \partial x_j}\,, \tag{148}$$

where

$$a_{ij}(x) \;=\; (\sigma(x)\sigma^T(x))_{ij}\,. \tag{149}$$

The matrix $A(x) = (a_{ij}(x))$ is non-negative definite for all x, $A(x) \ge 0$. The adjoint (in $L^2(dx)$) operator L^* is given by

$$L^* \;=\; \sum_i \frac{\partial}{\partial x_i} b_i(x) + \frac{1}{2}\sum_{ij}\frac{\partial^2}{\partial x_i \partial x_j} a_{ij}(x)\,. \tag{150}$$

It is called the Fokker-Planck operator. We have

$$T_t f(x) \;=\; \mathbf{E}\left[f(x_t)\right] \;=\; \int_{\mathbf{R}^n} P_t(x,dy) f(y)\,, \tag{151}$$

and we write

$$P_t(x,dy) \;=\; p_t(x,y)\,dy\,. \tag{152}$$

Although, in general, the probability measure $P_t(x,dy)$ does not necessarily have a density with respect to the Lebesgue measure, we can always interpret Eq. (152) in the sense of distributions. Since L is the generator of the semigroup L we have, in the sense of distributions,

$$\frac{\partial}{\partial t} p_t(x,\cdot) \;=\; L p_t(x,\cdot)\,. \tag{153}$$

The dual S_t of T_t acts on probability measure and if we write, in the sense of distributions, $d\pi(x) = \rho(x)\,dx$ we have

$$d(S_t\pi)(x) \;=\; T_t^*\rho(x)\,dx\,, \tag{154}$$

so that

$$\frac{\partial}{\partial t} p_t(\cdot,y) \;=\; L^* p_t(\cdot,y)\,. \tag{155}$$

In particular if π is an invariant measure $S_t\pi = \pi$ and we obtain the equation

$$L^*\rho(x) \,=\, 0\,. \tag{156}$$

If $A(x)$ is positive definite, $A(x) \,\geq\, c(x)\mathbf{1}$, $c(x) > 0$, we say that L is *elliptic*. There is an well-known elliptic regularity result: Let $H_s^{\rm loc}$ denote the local Sobolev space of index s. If A is elliptic then we have

$$Lf = g \ \text{ and } g \in H_s^{\rm loc} \quad\Longrightarrow\quad f \in H_{s+2}^{\rm loc}\,. \tag{157}$$

If L is elliptic then L^* is also elliptic. It follows, in particular that all eigenvectors of L and L^* are $\mathcal{C}^\infty$.

Let $X_i \,=\, \sum_j X_i^j(x)\frac{\partial}{\partial x_j}$, $i = 0, \cdots, M$ be $\mathcal{C}^\infty$ vectorfields. We denote by X_i^* its formal adjoint (on L^2). Let $f(x)$ be a $\mathcal{C}^\infty$ function. Let us consider operators K of the form

$$K \,=\, \sum_{j=1}^{M} X_j^*(x)X_j(x) + X_0(x) + f(x)\,. \tag{158}$$

Note that L, L^*, $\frac{\partial}{\partial t} - L$, and $\frac{\partial}{\partial t} - L^*$ have this form.

In many interesting physical applications, the generator fails to be elliptic. There is a theorem due to Hörmander which gives a very useful criterion to obtain the regularity of $p_t(x,y)$. We say that the family of vector fields $\{X_j\}$ satisfy *Hörmander condition* if the Lie algebra generated by the family

$$\{X_i\}_{i=0}^M \;,\; \{[X_i, X_j]\}_{i,j=0}^M \;,\; \{[[X_i, X_j], X_k]\}_{i,j,k=0}^M \;,\; \cdots\,, \tag{159}$$

has maximal rank at every point x.

Theorem 7.1. (Hörmander theorem) *If the family of vector fields $\{X_j\}$ satisfy Hörmander condition then there exists $\epsilon > 0$ such that*

$$Kf = g \ \text{ and } g \in H_s^{\rm loc} \quad\Longrightarrow\quad f \in H_{s+\epsilon}^{\rm loc}\,. \tag{160}$$

We call an operator which satisfies (160) an hypoelliptic operator. An analytic proof of Theorem 7.1 is given in [1], there are also probabilistic proofs which use Malliavin calculus, see [8] for a simple exposition.

As a consequence we have

Corollary 7.2. *Let $L = \sum_j Y_j(x)^* Y_j(x) + Y_0(x)$ be the generator of the diffusion x_t and let us assume that assume that (note that Y_0 is omitted!)*

$$\{Y_i\}_{i=1}^M \;,\; \{[Y_i, Y_j]\}_{i,j=0}^M \;,\; \{[[Y_i, Y_j], Y_k]\}_{i,j,k=0}^M \;,\; \cdots\,, \tag{161}$$

has rank n at every point x. Then L, L^*, $\frac{\partial}{\partial t} - L$, and $\frac{\partial}{\partial t} - L^*$ are hypoelliptic. The transition probabilities $P_t(x,y)$ have densities $p_t(x,y)$ which are C^∞ functions of (t,x,y) and the semigroup T_t is strong-Feller. The invariant measures, if they exist, have a C^∞ density $\rho(x)$.

Example 7.3. Consider the SDE

$$\begin{aligned} dq &= p\,dt\,, \\ dp &= \left(-\nabla V(q) - \lambda^2 p\right)\,dt + \lambda\sqrt{2T}dB_t\,, \end{aligned} \tag{162}$$

with generator

$$L = \lambda(T\nabla_p \cdot \nabla_p - p\cdot\nabla_p) + p\cdot\nabla_q - (\nabla_q V(q))\cdot\nabla_p\,. \tag{163}$$

In order to put in the form (158) we set

$$\begin{aligned} X_j(p,q) &= \lambda\sqrt{T}\frac{\partial}{\partial p_j}\,, \quad j = 1,2,\cdots,n\,, \\ X_0(p,q) &= -\lambda p\cdot\nabla_p + p\cdot\nabla_q - (\nabla_q V(q))\cdot\nabla_p\,, \end{aligned} \tag{164}$$

so that $L = -\sum_{j=1}^n X_j^* X_j + X_0$. The operator L is not elliptic since the matrix a_{ij} has only rank n. But L satisfies condition (161) since we have

$$[X_j\,,\,X_0] = -\lambda^2\sqrt{T}\frac{\partial}{\partial p_j} + \lambda\sqrt{T}\frac{\partial}{\partial q_j}\,, \tag{165}$$

and so the set $\{X_j, [X_j\,,\,X_0]\}_{j=1,n}$ has rank $2n$ at every point (p,q). This implies that L and L^* are hypoelliptic. The operator $\frac{\partial}{\partial t} - L$, and $\frac{\partial}{\partial t} - L^*$ are also hypoelliptic by considering the same set of vector fields together with X_0.

Therefore the transition probabilities $P_t(x,dy)$ have smooth densities $p_t(x,y)$. For that particular example it is easy to check that

$$\rho(x) = Z^{-1}e^{-\frac{1}{T}\left(\frac{p^2}{2}+V(q)\right)} \qquad Z = \int_{\mathbf{R}^{2n}} e^{-\frac{1}{T}\left(\frac{p^2}{2}+V(q)\right)}\,dpdq\,. \tag{166}$$

is the smooth density of an invariant measure, since it satisfies $L^*\rho = 0$. In general the explicit form of an invariant measure is not known and Theorem 7.1 implies that an invariant measure must have a smooth density, provided it exists.

8 Liapunov Functions and Ergodic Properties

In this section we will make the following standing assumptions

- **(H1)** The Markov process is irreducible aperiodic, i.e., there exists $t_0 > 0$ such that
$$P_{t_0}(x,A) > 0\,, \tag{167}$$
for all $x \in \mathbf{R}^n$ and all open sets A.
- **(H2)** The transition probability function $P_t(x,dy)$ has a density $p_t(x,y)$ which is a smooth function of (x,y). In particular T_t is strong-Feller, it maps bounded measurable functions into bounded continuous functions.

We are not optimal here and both condition $\mathbf{H1}$ can certainly be weakened. Note also that $\mathbf{H1}$ together with Chapman-Kolmogorov equations imply that (167) holds in fact for all $t > t_0$. We have discussed in Sections 6 and 7 some useful tools to establish $\mathbf{H1}$ and $\mathbf{H2}$.

Proposition 8.1. *If conditions* **H1** *and* **H2** *are satisfied then the Markov process* x_t *has at most one stationary distribution. The stationary distribution, if it exists, has a smooth everywhere positive density.*

Proof. By **H2** the dual semigroup S_t acting on measures maps measures into measures with a smooth density with respect to Lebesgue measure: $S_t\pi(dx) = \rho_t(x)dx$ for some smooth function $\rho_t(x)$. If we assume there is a stationary distribution

$$S_t\pi(dx) \;=\; \pi(dx)\,, \tag{168}$$

then clearly $\pi(dx) = \rho(x)dx$ must have a smooth density. Suppose that the invariant measure is not unique, then we might assume that π is not ergodic and thus, by Theorem 3.8 and Remark 3.9, there exists a nontrivial set A such that if the Markov process starts in A or A^c it never leaves it. Since π has a smooth density, we can assume that A is an open set and this contradicts ($\mathbf{H1}$).

If π is the stationary distribution then by invariance

$$\pi(A) \;=\; \int \pi(dx)P_t(x,A)\,, \tag{169}$$

and so $\pi(A) > 0$ for all open sets A. Therefore the density of π, denoted by ρ, is almost everywhere positive. Let us assume that for some y, $\rho(y) = 0$ then

$$\rho(y) \;=\; \int \rho(x)p_t(x,y)\,dx\,. \tag{170}$$

This implies that $p_t(x,y) = 0$ for almost all x and thus since it is smooth the function $p_t(\cdot,y)$ is identically 0 for all $t > 0$. On the other hand $p_t(x,y) \to \delta(x-y)$ as $t \to 0$ and this is a contradiction. So we have shown that $\rho(x) > 0$ for all x. □

Condition **H1** and **H2** do not imply in general the existence of a stationary distribution. For example Brownian motion obviously satisfies both conditions, but has no finite invariant distribution.

Remark 8.2. If the Markov process x_t has a compact phase space X instead of $\mathbf{R}^n$, then x_t always has a stationary distribution. To see this choose an arbitrary x_0 and set

$$\pi_t(dy) \;=\; \frac{1}{t}\int_0^t P_s(x_0,dy)\,ds\,. \tag{171}$$

The sequence of measures π_t has accumulation points in the weak topology. (Use Riesz representation Theorem to identify Borel measure as linear functional on $\mathcal{C}(X)$ and use the fact that the set of positive normalized linear functional on $\mathcal{C}(x)$ is weak-$*$ compact if X is compact). Furthermore the accumulation points are invariants. The details are left to the reader.

If the phase is not compact as in many interesting physical problems we need to find conditions which ensure the existence of stationary distributions. The basic idea is to show the existence of a compact set on which the Markov process spends "most of his time" except for rare excursions outside. On can express these properties systematically in terms of hitting times and Liapunov functions.

Recall that a Liapunov function is, by definition a positive function $W \geq 1$ with compact level sets ($\lim_{|x|\to\infty} W(x) = +\infty$).

Our first results gives a condition which ensure the existence of an invariant measure. If K is a subset of $\mathbf{R}^n$ we denote by $\tau_K = \tau_K(\omega) = \inf\{t, x_t(\omega) \in K\}$ the first time the process x_t enters the set K, τ_K is a stopping time.

Theorem 8.3. *Let x_t be a Markov process with generator L which satisfies condition* **H1** *and* **H2**. *Let us assume that there exists positive constant b and c, a compact set K, and a Liapunov function W such that*

$$LW \leq -c + b\mathbf{1}_K x \,. \tag{172}$$

Then we have

$$\mathbf{E}_x[\tau_K] < \infty \,. \tag{173}$$

for all $x \in \mathbf{R}^n$ and the exists a unique stationary distribution π.

Remark 8.4. One can show the converse statement that the finiteness of the expected hitting time do imply the existence of a Liapunov function which satisfies Eq. (172).

Remark 8.5. It turns out that sometimes it is more convenient to show the existence of a Liapunov function expressed in terms of the semigroup T_{t_0} for a fixed time t_0 rather than in terms of the generator L. If we assume that there exists constants positive b, c, a compact set K, and a Liapunov function W which satisfies

$$T_{t_0}W - W \leq -c + b\mathbf{1}_K x \,. \tag{174}$$

then the conclusion of the theorem still hold true.

Proof. We first prove the assertion on the hitting time. If $x \in K$ then clearly $\mathbf{E}_x[\tau_K] = 0$. So let us assume that $x \notin K$. Let us choose n so large that $W(x) < n$. Now let us set

$$\tau_{K,n} = \inf\{t\,,\, x_t \in K \cup \{W(x) \geq n\}\}\,, \quad \tau_{K,n}(t) = \inf\{\tau_{K,n}, t\}\,. \tag{175}$$

Obviously $\tau_{K,n}$ and $\tau_{K,n}(t)$ are stopping time and using Ito formula with stopping time we have

$$\begin{aligned}
\mathbf{E}_x\left[W(x_{\tau_{K,n}(t)})\right] - W(x) &= \mathbf{E}_x\left[\int_0^{\tau_{K,n}(t)} LW(x_s)\,ds\right] \\
&\leq \mathbf{E}_x\left[\int_0^{\tau_{K,n}(t)} -c + b\mathbf{1}_K(x)\,ds\right] \\
&\leq \mathbf{E}_x\left[\int_0^{\tau_{K,n}(t)} -c\right] \\
&\leq -c\mathbf{E}_x\left[\tau_{K,n}(t)\right]\,.
\end{aligned} \tag{176}$$

Since $W \geq 1$ we obtain

$$\mathbf{E}_x\left[\tau_{K,n}(t)\right] \leq \frac{1}{c}\left(W(x) - \mathbf{E}_x\left[W(x_{\tau_{K,n}(t)})\right]\right) \leq \frac{1}{c}W(x)\,. \tag{177}$$

Proceeding as in Theorem 5.9, using Fatou's lemma, we first take the limit $n \to \infty$ and, since $\lim_{n\to\infty} \tau_{K,n}(t) = \tau_K(t)$ we obtain

$$\mathbf{E}_x\left[\tau_K(t)\right] \leq \frac{W(x)}{c}\,. \tag{178}$$

Then we take the limit $t \to \infty$ and, since $\lim_{t\to\infty} \tau_K(t) = \tau_K$, we obtain

$$\mathbf{E}_x\left[\tau_K\right] \leq \frac{W(x)}{c}\,. \tag{179}$$

We show next the existence of an invariant measure. The construction goes via an embedded (discrete-time) Markov chain. Let us choose a compact set $\tilde{K}$ with K contained in the interior of $\tilde{K}$. We assume they have smooth boundaries which we denote by Γ and $\tilde{\Gamma}$ respectively. We divide now an arbitrary path x_t into cycles in the following way. Let $\tau_0 = 0$, let τ_1' be the first time after τ_0 at which x_t reaches $\tilde{\Gamma}$, τ_1 is the first time after τ_1' at which x_t reaches Γ and so on. It is not difficult to see that, under our assumptions, τ_j and τ_j' are almost surely finite. We define now a discrete-time Markov chain by $X_0 = x \in \Gamma$ and $X_i = x_{\tau_i}$. We denote by $\tilde{P}(x, dy)$ the one-step transition probability of X_n. We note that the Markov chain X_n has a compact phase space and so it possess a stationary distribution $\mu(dx)$ on Γ, by the same argument as the one sketched in Remark (8.2).

We construct now the invariant measure for x_t in the following way. Let $A \subset \mathbf{R}^n$ be a measurable set. We denote σ_A the time spent in A by x_t during the first cycle between 0 and τ_1. We define an unnormalized measure π by

$$\pi(A) = \int_\Gamma \mu(dx)\mathbf{E}_x[\sigma_A]\,. \tag{180}$$

Then for any bounded continuous function we have

$$\int_{\mathbf{R}^n} f(x)\pi(dx) = \int_\Gamma \mu(dx)\mathbf{E}_x\left[\int_0^{\tau_1} f(x_s)ds\right]\,. \tag{181}$$

In order to show that π is stationary we need to show that, for any bounded continuous f,

$$\int_{\mathbf{R}^n} T_t f(x)\pi(dx) = \int_{\mathbf{R}^n} f(x)\pi(dx)\,, \tag{182}$$

i.e., using our definition of π

$$\int_\Gamma \mu(dx)\mathbf{E}_x\left[\int_0^{\tau_1} \mathbf{E}_{x_s}[f(x_t)]ds\right] = \int_\Gamma \mu(dx)\mathbf{E}_x\left[\int_0^{\tau_1} f(x(s))ds\right]\,. \tag{183}$$

For any measurable continuous function we have

$$\begin{aligned}\mathbf{E}_x\left[\int_0^{\tau_1} f(x_{t+s})\,ds\right] &= \mathbf{E}_x\left[\int_0^{\infty} \mathbf{E}_x[\mathbf{1}_{s<\tau_1} f(x_{t+s})]\,ds\right] \\ &= \mathbf{E}_x\left[\int_0^{\infty} \mathbf{1}_{s<\tau_1}\mathbf{E}_{x_s}[f(x(t))]\,ds\right] \\ &= \mathbf{E}_x\left[\int_0^{\tau_1} \mathbf{E}_{x_s}[f(x_t)]\,ds\right]. \end{aligned} \tag{184}$$

Thus we have, using Eqs. (182), (183), and (184),

$$\begin{aligned}&\int_{\mathbf{R}^n} T_t f(x)\pi(dx) = \int_\Gamma \mu(dx)\mathbf{E}_x\left[\int_0^{\tau_1} \mathbf{E}_{x_s} f(x_t)\,ds\right] \\ &= \int_\Gamma \mu(dx)\mathbf{E}_x\left[\int_0^{\tau_1} f(x_{t+s})ds\right] = \int_\Gamma \mu(dx)\mathbf{E}_x\left[\int_t^{\tau_1+t} f(x_u)\,du\right] \\ &= \int_\Gamma \mu(dx)\mathbf{E}_x\left[\int_0^{\tau_1} f(x_u)\,du + \int_t^{t+\tau_1} f(x_u)\,du - \int_0^t f(x_u)\,du\right]. \end{aligned} \tag{185}$$

Since μ is a stationary distribution for X_n, for any bounded measurable function on Γ

$$\int_\Gamma \mu(dx)\mathbf{E}_x[g(X_1)] = \int_\Gamma \mu(dx)g(X)\,, \tag{186}$$

and so

$$\begin{aligned}\int_\Gamma \mu(dx)\mathbf{E}_x\left[\int_t^{t+\tau_1} f(x_u)\,du\right] &= \int_\Gamma \mu(dx)\mathbf{E}_x\left[\mathbf{E}_{X_1}\left[\int_0^t f(x_u)\,du\right]\right] \\ &= \int_\Gamma \mu(dx)\mathbf{E}_x\left[\int_0^t f(x_u)\,du\right]. \end{aligned} \tag{187}$$

Combining Eqs. (185) and (187) we obtain that

$$\int_{\mathbf{R}^n} T_t f(x)\pi(dx) = \int_{\mathbf{R}^n} f(x)\pi(dx)\,. \tag{188}$$

This shows that the measure π satisfies $S_t\pi = \pi$. Finally we note that

$$\pi(\mathbf{R}^n) = \int_\Gamma \mu(dx)\mathbf{E}_x[\tau_K] < \infty\,. \tag{189}$$

and so π can be normalized to a probability measure. □

Remark 8.6. One can also show (see e.g. [15] or [7]) that that under the same conditions one has convergence to the stationary state: for any x we have

$$\lim_{t\to\infty} \|P_t(x,\cdot) - \pi(\cdot)\| = 0\,, \tag{190}$$

where $\|\mu\| = \sup_{\|f\|\le 1}\left|\int f(x)\mu(dx)\right|$ is the total variation norm of a signed measure μ. Thus the measure π is also mixing. We don't prove this here, but we will prove an exponential convergence result using a stronger condition than (172).

It is often useful to to quantify the rate of convergence at which an initial distribution converges to the stationary distribution and we prove two such results in that direction which provide exponential convergence. We mention here that polynomial rate of convergence can be also expressed in terms of Liapunov functions, and this makes these functions a particularly versatile tool.

We introduce some notations and definitions. If W is a Liapunov function and μ a signed measure we introduce a weighted total variation norm given by

$$\|\mu\|_W \;=\; \sup_{|f|\le W} |\int f(x)\mu(dx)|\,, \tag{191}$$

and a norm on functions $\|\,\cdot\,\|_W$ given by

$$\|f\|_W \;=\; \sup_{x\in\mathbf{R}^n} \frac{|f(x)|}{W(x)}\,, \tag{192}$$

and a corresponding Banach space $\mathcal{H}_W$ given by

$$\mathcal{H}_W \;=\; \{f\,,\, \|f\|_W < \infty\}\,. \tag{193}$$

Theorem 8.7. (Quasicompactness) *Suppose that the conditions* **H1** *and* **H2** *hold. Let K be a compact set and let W be a Liapunov function W. Assume that either of the following conditions hold*

1. *There exists constants $a > 0$ and $b < \infty$ such that*

$$LW(x) \;\le\; -aW(x) + b\mathbf{1}_K(x)\,. \tag{194}$$

2. *We have $LW \le cW$ for some $c > 0$ and there exists constants $\kappa < 1$ and $b < \infty$, and a time $t_0 > 0$ such that*

$$T_{t_0}W(x) \;\le\; \kappa W(x) + b\mathbf{1}_K(x)\,. \tag{195}$$

Then for δ small enough

$$\mathbf{E}_x\left[e^{\delta\tau_K}\right] \;<\; \infty\,, \tag{196}$$

for all $x \in \mathbf{R}^n$. The Markov process has a stationary distribution x_t and there exists a constants $C > 0$ and $\gamma > 0$ such that

$$\|P_t(x,dy) - \pi(dy)\|_W \;\le\; CW(x)e^{-\gamma t}\,, \tag{197}$$

or equivalently

$$\|T_t - \pi\|_W \;\le\; Ce^{-\gamma t}\,. \tag{198}$$

Remark 8.8. In Theorem 8.7 one can replace $\mathcal{H}_W$ by any of the space

$$\mathcal{H}_{W,p} \;=\; \left\{f\,,\, \frac{|f|}{W} \in L^p(dx)\right\}\,, \tag{199}$$

with $1 < p < \infty$.

Proof. We will prove here only the part of the argument which is different from the proof of Theorem 8.9. We recall that a bounded operator T acting on a Banach space $\mathcal{B}$ is quasi compact if its spectral radius is M and its essential spectral radius is $\theta < M$. By definition, it means that outside the disk of radius θ, the spectrum of T consists of isolated eigenvalues of finite multiplicity. The following formula for the essential spectral radius θ is proved in [9]

$$\theta \;=\; \lim_{n\to\infty} \left(\inf\{\|T^n - C\| \,|\, C \text{ compact}\}\right)^{1/n} . \tag{200}$$

Let us assume that Condition 2. holds. If $\|f\| \in \mathcal{H}_W$ then, by definition, we have $|f(x)| \;\le\; \|f\|_W W(x)$ and, since T_t is positive, we have

$$|T_t f(x)| \;\le\; \|f\|_W T_t W(x) \,. \tag{201}$$

We consider a fixed $t > t_0$. We note that the Liapunov condition implies the bound

$$T_t W(x) \;\le\; \kappa W(x) + b \,. \tag{202}$$

Iterating this bound and using that $T_t 1 = 1$ we have, for all $n \ge 1$, the bounds

$$T_{nt} W(x) \;\le\; \kappa^n W(x) + \frac{b}{1-\kappa} \,. \tag{203}$$

Let K be any compact set, we have the bound

$$\begin{aligned} |\mathbf{1}_{K^c}(x) T_{nt} f(x)| \;&\le\; W(x) \sup_{y\in K^c} \frac{|T_{nt} f(y)|}{W(y)} \\ &\le\; W(x)\|f\|_W \sup_{y\in K^c} \frac{T_{nt} W(y)}{W(y)} \\ &\le\; W(x)\|f\|_W \left(\kappa^n + \frac{b}{1-\kappa} \sup_{y\in K^c} \frac{1}{W(x)}\right) . \end{aligned} \tag{204}$$

Since $\lim_{\|x\|\to\infty} W(x) = \infty$, given $\epsilon > 0$ and $n > 1$ we can choose a compact set K_n such that

$$\|\mathbf{1}_{K_n^c} T_{nt}\|_W \;\le\; (\kappa + \epsilon)^n \,. \tag{205}$$

On the other hand since T_t has a smooth kernel the set, for any compact K, the set

$$\{\mathbf{1}_K(x) T_t f(x) \,|\, \|f\|_W = 1\}$$

is compact by Arzelà-Ascoli. Therefore we have

$$\inf\{\|T_{nt} - C\| \,|\, C \text{ compact}\} \;\le\; \|T_{nt} - \mathbf{1}_{K_n} T_{nt}\| \le (\kappa+\epsilon)^n \,. \tag{206}$$

and therefore the essential spectral radius of T_t is less than κ. In order to obtain the exponential convergence from this one must prove that there is no other eigenvalue than 1 on the unit disk (or our outside the unit disk) and prove that 1 is a simple eigenvalue. We will prove this in Theorem 8.9 and the same argument apply here. Also the assertion on hitting times is proved as in Theorem 8.9. □

Theorem 8.9. (Compactness) *Suppose that the conditions* **H1** *and* **H2** *hold. Let* $\{K_n\}$ *be a sequence of compact sets and let* W *be a Liapunov function. Assume that either of the following conditions hold*

1. *There exists constants* $a_n > 0$ *with* $lim_{n \to \infty} a_n = \infty$ *and constants* $b_n < \infty$ *such that*
$$LW(x) \le -a_n W(x) + b_n \mathbf{1}_{K_n}(x) . \tag{207}$$
2. *There exists a constant c such that* $LW \le cW$ *and there exists constants* $\kappa_n < 1$ *with* $\lim_{n \to \infty} \kappa_n = 0$*, constants* $b_n < \infty$*, and a time* $t_0 > 0$ *such that*
$$T_{t_0} W(x) \le \kappa_n W(x) + b_n \mathbf{1}_{K_n}(x) . \tag{208}$$

Then for any (arbitrarily large) δ *there exists a compact set* $C = C(\delta)$ *such that*

$$\mathbf{E}_x \left[e^{\delta \tau_C} \right] < \infty , \tag{209}$$

for all $x \in \mathbf{R}^n$*. The Markov process* x_t *has a unique stationary distribution* π*. The semigroup* T_t *acting on* $\mathcal{H}_W$ *is a* **compact** *semigroup for* $t > t_0$ *and there exists a constants* $C > 0$ *and* $\gamma > 0$ *such that*

$$\| P_t(x, dy) - \pi(dy) \| \le C W(x) e^{-\gamma t} . \tag{210}$$

or equivalently

$$\| T_t - \pi \|_W \le C e^{-\gamma t} . \tag{211}$$

Proof. We will assume that conditions 2. holds. Let us prove the assertion on hitting times. Let X_n be the Markov chain defined by $X_0 =$ and $X_n = x_{nt_0}$ and for a set K let N_K be the least integer such that $X_{N_K} \in K$. We have $N_K \le \tau_K$ so it is sufficient to prove that $\mathbf{E}_x[e^{\delta N_K}] < \infty$.

Let K_n be the compact set given in Eq. (208). We can assume, by increasing K_n is necessary that K_n is a level set of W, i.e. $K_n = \{W(x) \le W_n\}$

Using the Liapunov condition and Chebyshev inequality we obtain the following tail estimate

$$
\begin{aligned}
&\mathbf{P}_x\left\{N_{K_n} > j\right\} \\
&= \mathbf{P}_x\left\{W(X_j) > W_n\,,\, X_i \in K_n^c\,,\, 0 \le i \le j\right\} \\
&= \mathbf{P}_x\left\{\prod_{i=i}^{j} \frac{W(X_i)}{W(x_{i-1})} > \frac{W_n}{W(x)}\,,\, X_i \in K_n^C\right\} \\
&\le \frac{W(x)}{W_n}\mathbf{E}_x\left[\prod_{i=i}^{j} \frac{W(X_i)}{W(x_{i-1})}\,,\, X_i \in K_n^C\right] \\
&\le \frac{W(x)}{W_n}\mathbf{E}_x\left[\prod_{i=i}^{j-1} \frac{W(X_i)}{W(x_{i-1})}\mathbf{E}_{X_{j-1}}\left[\frac{W_{X_j}}{W_{X_{j-1}}}\right]\,,\, X_i \in K_n^C\right] \\
&\le \frac{W(x)}{W_n}\sup_{y\in K_n^C}\mathbf{E}_y\left[\frac{W(x_1)}{W(y)}\right]\mathbf{E}_x\left[\prod_{i=i}^{j-1} \frac{W(X_i)}{W(x_{i-1})}\,,\, X_i \in K_n^C\right] \\
&\le \cdots \le \frac{W(x)}{W_n}\left(\sup_{y\in K_n^C}\mathbf{E}_y\left[\frac{W(x_1)}{W(y)}\right]\right)^j \\
&\le \frac{W(x)}{W_n}(\kappa_n)^j\,. \qquad (212)
\end{aligned}
$$

We thus have geometric decay of $P_{>j} \equiv P\{N_{k_n} > j\}$ in j. Summing by parts we obtain

$$
\begin{aligned}
&\mathbf{E}_x\left[e^{\delta N_{k_n}}\right] = \sum_{j=1}^{\infty} e^{\delta j}\mathbf{P}_x\{N_{K_n} = j\} \\
&= \lim_{M\to\infty}\left(\sum_{j=1}^{M} P_{>j}(e^{\delta(j+1)} - e^{\delta j}) + e^{\delta}P_{>0} - e^{\delta(M+1)}P_{>M}\right) \\
&\le e^{\delta} + \frac{W(x)}{W_n}(e^{\delta}-1)\sum_{j=1}^{\infty}\kappa_n^j e^{j\delta} \\
&\le e^{\delta} + \frac{W(x)}{W_n}(e^{\delta}-1)\frac{\kappa_n e^{\delta}}{1-\kappa_n e^{\delta}}\,, \qquad (213)
\end{aligned}
$$

provided $\delta < \ln(\kappa_n^{-1})$. Since we can choose κ_n arbitrarily small, this proves the claim about the hitting time.

If $\|f\| \in \mathcal{H}_W$ then by definition we have $|f(x)| \le \|f\|_W W(x)$ and thus $|T_t f(x)| \le \|f\|_W T_t W(x)$.

The compactness of T_t is a consequence of the following estimate. Using the Liapunov condition we have for $t > t_0$

$$\begin{aligned} \left|\mathbf{1}_{K_n^c}(x)T_tf(x)\right| &\leq W(x)\sup_{y\in K_n^c}\frac{|T_tf(y)|}{W(y)} \\ &\leq W(x)\|f\|_W\sup_{y\in K_n^c}\frac{T_tW(y)}{W(y)} \\ &\leq \kappa_n W(x)\|f\|_W\,, \end{aligned} \tag{214}$$

or

$$\|\mathbf{1}_{K_n^C}T_t\|_W \leq \kappa_n\,. \tag{215}$$

We have thus

$$\lim_{n\to\infty}\|\mathbf{1}_{K_n^C}T_t\|_W = 0\,, \tag{216}$$

i.e., $\mathbf{1}_{K_n^C}T_t$ converges to 0 in norm. On the other hand since T_t has a smooth kernel and K_n is compact, the operator

$$\mathbf{1}_{K_n}T_t\mathbf{1}_{K_n} \tag{217}$$

is a compact operator by Arzela-Ascoli Theorem. We obtain

$$\begin{aligned} T_t &= \left(\mathbf{1}_{K_n}+\mathbf{1}_{K_n^C}\right)T_{t-\epsilon}\left(\mathbf{1}_{K_n}+\mathbf{1}_{K_n}\right)T_\epsilon \\ &= \lim_{n\to\infty}\mathbf{1}_{K_n}T_{t-\epsilon}\mathbf{1}_{K_n}T_\epsilon\,. \end{aligned} \tag{218}$$

So T_t is the limit in norm of compact operators, hence it is compact. Its spectrum consists of 0 and eigenvalues with finite multiplicity.

We show that there no eigenvalues of modulus bigger than one. Assume the contrary. Then there is f and λ with $|\lambda|>1$ such that $T_tf = \lambda f$. Since T_t is positive we have

$$|\lambda||f| = |T_tf| \leq T_t|f|\,, \tag{219}$$

and therefore

$$T_t|f|-|f| \geq (|\lambda|-1)|f|\,. \tag{220}$$

Integrating with the strictly positive stationary distribution we have

$$\int T_t|f|\pi(dx)-\int|f|\pi(dx) \geq (|\lambda|-1)\int|f|\pi(dx) > 0\,. \tag{221}$$

This is a contradiction since, by stationarity $\int T_t|f|\pi(dx) = \int|f|\pi(dx)$, and so the r.h.s. of Eq. (221) is 0.

Next we show that 1 is a simple eigenvalue with eigenfunction given by the constant function. Clearly 1 is algebraically simple, if there is another eigenfunction with eigenvalue one, the stationary distribution is not unique. We show that 1 is also geometrically simple. Assume the contrary, then by Jordan decomposition there is a function g such that $T_tg = 1+g$, so $T_tg-g = 1$. Integrating with respect to the stationary distribution gives a contradiction.

Finally we show that there no other eigenvalues on the unit disk. Assume there exists f, $f \neq 1$ such that $T_tf = \lambda f$ with $|\lambda| = 1$. By the semigroup property T_t

is compact for all $t > t_0$ and $T_t f = e^{i\alpha t} f$, with $\alpha \in \mathbf{R}$, $\alpha \neq 0$. If we choose $t = 2n\pi/\alpha$ we obtain $T_{2n\pi/\alpha} f = f$ and so $f = 1$ which is a contradiction.

Using now the spectral decomposition of compact operators we find that there exists $\gamma > 0$ such that

$$\|T_t - \pi\| \le Ce^{-\gamma t} . \tag{222}$$

This concludes the proof of Theorem 8.9 □

To conclude we show that the correlations in the stationary distribution decay exponentially (exponential mixing).

Corollary 8.10. *Under the assumptions of Theorem 8.7 or 8.9, the stationary distribution π is exponentially mixing: for all f, g such that $f^2, g^2 \in \mathcal{H}_W$ we have*

$$| \int f(x) T_t g(x) \pi(dx) - \int f(x)\pi(dx) \int g(x)\pi(dx) | \, C \|f^2\|_W^{1/2} \|g^2\|_W^{1/2} e^{-\gamma t} . \tag{223}$$

Proof. If $f^2 \in \mathcal{H}_W$, then we have

$$|f(x)| \le \|f^2\|_W^{1/2} W^{1/2} . \tag{224}$$

Further if we have

$$T_{t_0} W(x) \le \kappa_n W(x) + b_n \mathbf{1}_{K_n}(x) , \tag{225}$$

then, using Jensen inequality, the inequality $\sqrt{1+y} \le 1 + y/2$ and $W \ge 1$ we have

$$\begin{aligned} T_{t_0} \sqrt{W}(x) &\le \sqrt{T_{t_0} W(x)} \\ &\le \sqrt{\kappa_n W(x) + b_n \mathbf{1}_{K_n}(x)} \\ &\le \sqrt{\kappa_n} \sqrt{W(x)} + \frac{b_n \mathbf{1}_{K_n}(x)}{2\kappa_n} . \end{aligned} \tag{226}$$

So we have

$$T_{t_0} \sqrt{W}(x) \le \kappa_n' \sqrt{W}(x) + b_n' \mathbf{1}_{K_n}(x) , \tag{227}$$

with $\kappa_n' = \sqrt{\kappa_n}$ and $b_n' = b_n/2\kappa_n$. Applying Theorem 8.9 or 8.7 with the Liapunov function $\sqrt{W}$ there exist constants $C > 0$ and $\gamma > 0$ such that

$$\left| T_t g(x) - \int g(x)\pi(dx) \right| \le C \sqrt{W(x)} \|g^2\|_W^{1/2} e^{-\gamma t} . \tag{228}$$

Therefore combining Eqs. (224) and (228) we obtain

$$\begin{aligned} &\left| \int f(x) T_t g(x) \pi(dx) - \int f(x)\pi(dx) \int g(x)\pi(dx) \right| \\ &\le \int |f(x)| \left| T_t g(x) - \int g(x)\pi(dx) \right| \pi(dx) \\ &\le C \int W(x)\pi(dx) \|f^2\|^{1/2} \|g^2\|^{1/2} e^{-\gamma t} . \end{aligned} \tag{229}$$

Finally since π is the solution of $S_t\pi = \pi$ we have

$$\int W(x)\pi(dx) \leq \|\pi\|_W < \infty \tag{230}$$

and this concludes the proof of Corollary 8.10. □

References

1. Hörmander, L.: *The Analysis of linear partial differential operators*. Vol **III**, Berlin: Springer, 1985
2. Has'minskii, R.Z.: *Stochastic stability of differential equations*. Alphen aan den Rijn—Germantown: Sijthoff and Noordhoff, 1980
3. Karatzas, I. and Shreve S.E.: *Brownian Motion and Stochastic Calculus*. Graduate Texts in Mathematics, Berlin: Springer (1997)
4. Kliemann, W.: Recurrence and invariant measures for degenerate diffusions. Ann. of Prob. **15**, 690–702 (1987)
5. Kunita, H.: *Stochastic flows and stochastic differential equations*. Cambridge Studies in Advanced Mathematics, 24. Cambridge: Cambridge University Press (1990)
6. Kunita, H.: Supports of diffusion process and controllability problems. In: *Proc. Inter. Symp. SDE Kyoto 1976*. New-York: Wiley, 1078, pp. 163-185
7. Meyn, S.P. and Tweedie, R.L.: *Markov Chains and Stochastic Stability.* Communication and Control Engineering Series, London: Springer-Verlag London, 1993
8. Norriss, J.: Simplified Malliavin Calculus. In *Séminaire de probabilités XX*, Lectures Note in Math. **1204**, 0 Berlin: Springer, 1986, pp. 101–130
9. Nussbaum R.D.: The radius of the essential spectral radius. Duke Math. J. **37**, 473–478 (1970)
10. Oksendal, B.: *Stochastic differential equations. An introduction with applications.* Fifth edition. Universitext. Berlin: Springer-Verlag, 1998
11. Rey-Bellet L.: Open classical systems. In this volume.
12. Stroock, D.W. and Varadhan, S.R.S.: On the support of diffusion processes with applications to the strong maximum principle. In *Proc. 6-th Berkeley Symp. Math. Stat. Prob.*, Vol **III**, Berkeley: Univ. California Press, 1972, pp. 361–368
13. Varadhan, S.R.S: *Probability Theory*. Courant Institute of Mathematical Sciences - AMS, 2001
14. Varadhan, S.R.S: Lecture notes on Stochastic Processes (Fall 00). http://www.math.nyu.edu/faculty/varadhan/processes.html