Convergence of Markov Processes

January 13, 2016

Martin Hairer

Mathematics Department, University of Warwick
Email: `M.Hairer@Warwick.ac.uk`

Abstract

The aim of this minicourse is to provide a number of tools that allow one to determine at which speed (if at all) the law of a diffusion process, or indeed a rather general Markov process, approaches its stationary distribution. Of particular interest will be cases where this speed is subexponential. After an introduction to the general ergodic theory of Markov processes, the first part of the course is devoted to Lyapunov function techniques. The second part is then devoted to an elementary introduction to Malliavin calculus and to a proof of Hörmander's famous "sums of squares" regularity theorem.

1 General (ergodic) theory of Markov processes

In this note, we are interested in the long-time behaviour of Markov processes, both in discrete and continuous time. Recall that a discrete-time Markov process x on a state space $\mathcal{X}$ is described by a transition kernel $\mathcal{P}$, which we define as a measurable map from $\mathcal{X}$ into the space of probability measures on $\mathcal{X}$. In all that follows, $\mathcal{X}$ will always be assumed to be a Polish space, that is a complete, separable metric space. When viewed as a measurable space, we will always endow it with its natural Borel σ-algebra, that is the smallest σ-algebra containing all open sets. This ensures that $\mathcal{X}$ endowed with any probability measure is a Lebesgue space and that basic intuitive results of probability and measure theory (Fubini's theorem, regular conditional probabilities, etc) are readily available.

We will sometimes consider $\mathcal{P}$ as a linear operator on the space of signed measures on $\mathcal{X}$ and / or the space of bounded measurable functions by

$$(\mathcal{P}\mu)(A) = \int_{\mathcal{X}} \mathcal{P}(x,A)\,\mu(dx)\;, \qquad (\mathcal{P}\varphi)(x) = \int_{\mathcal{X}} \varphi(y)\,\mathcal{P}(x,dy)\;.$$

Hence we are using $\mathcal{P}$ both to denote the action on functions and its dual action on measures. Note that $\mathcal{P}$ extends trivially to measurable functions $\varphi\colon \mathcal{X} \to [0,+\infty]$. The main concept arising in the study in the long-time behaviour of a Markov chain is that of an invariant measure:

Definition 1.1 A positive measure μ on $\mathcal{X}$ is invariant for the Markov process x if $\mathcal{P}\mu = \mu$.

In the case of discrete state space, another key notion is that of transience, recurrence and positive recurrence of a Markov chain. The next subsection explores these notions and how they relate to the concept of an invariant measure.

1.1 Transience and recurrence

Take a Markov process on a countable state space, say $\mathbf{Z}$ and denote its transition probabilities by $P_{ij} = \mathcal{P}(j, \{i\})$. In this simple case, the qualitative long-time behaviour of the chain can be described in the following way. For a given state i, we denote by τ_i the first return time of the process to $\{i\}$, i.e.

$$\tau_i = \inf\{k \geq 1 \,:\, x_k = i\} \;.$$

We also denote by $\mathbf{P}_i$ the law of the process started at i. With these notations, we say that the state i is:

- Transient if $\mathbf{P}_i(\tau_i = \infty) > 0$.
- Recurrent if $\mathbf{P}_i(\tau_i < \infty) = 1$.
- Positive recurrent if it is recurrent and $\mathbf{E}_i\tau_i < \infty$.

We assume that P is irreducible in the sense that for every pair (i, j) there exists $n > 0$ such that $P^n_{ij} > 0$. Given a state i, we also denote by τ_i the first return time to i, that is $\tau_i = \inf\{n > 0 \,:\, x_n = i\}$. We have

Proposition 1.2 *If P is irreducible, then all of the states are of the same type.*

Proof. The proof is based on the following fact. If P is irreducible and $j \neq k$ are any two states, then $\mathbf{P}_j(\tau_k < \tau_j) > 0$. The argument goes by contradiction: assume that the probability vanishes. Then, by the strong Markov property, the process starting from j would never visit the state k. This is in contradiction with the irreducibility of P.

We first show that we cannot have one recurrent and one transient state. Assume by contradiction that i is transient and that $k \neq i$ is recurrent. Start the process in the state k so that, by the recurrence of k and the strong Markov property, it almost surely visits k infinitely often. The above remark on the other hand ensure that between each of these visits, the process has a strictly positive probability of visiting i. Since these 'trials' are independent, the process almost surely visits i infinitely often. From the strong Markov property applied at the first hitting time of i, we conclude that i must be recurrent as well.

It remains to show that if k is positive recurrent, then i must also be so. Denote by A the event that the process visits i between two successive visits to k, and denote by p the probability of A. If $p = 1$, then $\mathbf{E}_i\tau_i \le 2\mathbf{E}_k\tau_k$ and we are done, so we can assume that $p \in (0,1)$, so that

$$\mathbf{E}_k(\tau_k \mid A) \le \frac{\mathbf{E}_k\tau_k}{p}\;, \qquad \mathbf{E}_k(\tau_k \mid \bar{A}) \le \frac{\mathbf{E}_k\tau_k}{1-p}\;.$$

We now break the excursion from i into a trajectory from i to k, a number of excursions from k to k, and a final piece of trajectory from k back to i. Since the first and the last piece of trajectory are part of an excursion from k that does visit i, this yields the upper bound

$$\mathbf{E}_i\tau_i \le 2\frac{\mathbf{E}_k\tau_k}{p} + p\sum_{n\ge 0}(1-p)^n n\frac{\mathbf{E}_k\tau_k}{1-p} \le 3\frac{\mathbf{E}_k\tau_k}{p}\;,$$

thus concluding the proof. □

It turns out that if a Markov chain is positive recurrent, then it has a finite invariant measure. If it is only recurrent, then it still has a σ-finite invariant measure. If on the other hand it is transient, the notion of an invariant measure is useless since the process visits states only finitely often. To show this, it suffices to make the following construction. Fix a distinguished state $\triangle$ and denote by $\tilde{x}$ the process that starts at $\triangle$ and stops as soon as it reaches $\triangle$ again, but is conditioned to spend at least one time step outside of $\triangle$. The transition probabilities $\tilde{P}_{ij}$ for $\tilde{x}$ are then given by

$$\tilde{P}_{ij} = \begin{cases} P_{ij} & \text{if } j \neq \triangle, \\ 1 & \text{if } i = j = \triangle. \\ 0 & \text{otherwise.} \end{cases}$$

We now set $\pi_\triangle = 1$ and, for $j \neq \triangle$,

$$\pi_j = \sum_{k=1}^\infty \mathbf{P}(\tilde{x}_k = j) = \mathbf{E}\sum_{k=1}^\infty \mathbf{1}_{\{j\}}(\tilde{x}_k)\;.$$

Note first that since there is a non-zero probability that the process reaches $\triangle$ between any two successive return times to j, this quantity is finite for every j. Furthermore, it follows from the definition of $\tilde{x}$ that we have the identity

$$\pi_j = \begin{cases} \sum_{k=0}^\infty \sum_{i\neq\triangle} \tilde{P}_{ji}^k P_{i\triangle} & \text{if } j \neq \triangle, \\ 1 & \text{if } j = \triangle. \end{cases}$$

Therefore, for $\ell \neq \triangle$ we have

$$(P\pi)_\ell = P_{\ell\triangle} + \sum_{j\neq\triangle}\sum_{k\ge 0}\sum_{i\neq\triangle} P_{\ell j}\tilde{P}_{ji}^k P_{i\triangle} = P_{\ell\triangle} + \sum_{j}\sum_{k\ge 0}\sum_{i\neq\triangle} \tilde{P}_{\ell j}\tilde{P}_{ji}^k P_{i\triangle} = \pi_\ell\;.$$

On the other hand, we have

$$(P\pi)_\triangle = P_{\triangle\triangle} + \sum_{j\neq\triangle}\sum_{k=0}^\infty\sum_{i\neq\triangle} P_{\triangle j}\tilde{P}_{ji}^k P_{i\triangle} = 1\;,$$

since this is precisely the probability that the process x eventually returns to $\triangle$.

1.2 Foster-Lyapunov criteria

We have the following Foster-Lyapunov drift criteria on *countable* state space. As a shorthand notation, we define the 'discrete generator' of our Markov process by $\mathcal{L}f = \mathcal{P}f - f$.

Proposition 1.3 *Consider a Markov process with transition probabilities $\mathcal{P}$ on a countable state space. Then:*

- *It is transient if and only if there exists a function $V\colon \mathcal{X} \to \mathbf{R}_+$ and a non-empty set $A \subset \mathcal{X}$ such that $\mathcal{L}V(x) \le 0$ for all $x \notin A$ and there exists $x \notin A$ such that $V(x) < \inf_{y\in A} V(y)$.*
- *It is recurrent if and only if there exists a function $V\colon \mathcal{X} \to \mathbf{R}_+$ such that $\{x : V(x) \le N\}$ is finite for every $N > 0$ and such that $\mathcal{L}V(x) \le 0$ for all but finitely many values of x.*
- *It is positive recurrent if and only if there exists a function $V\colon \mathcal{X} \to \mathbf{R}_+$ such that $\mathcal{L}V(x) \le -1$ for all but finitely many values of x.*

Proof. In all three cases, we will show that the existence of a function V with the specified properties is sufficient to obtain the corresponding transience / recurrence properties. We then show how to construct such a function in an abstract way.

Let us first show the criterion for transience. Multiplying V by a positive constant, we can assume without loss of generality that $\inf_{y\in A} V(y) = 1$. Consider now $x \notin A$ such that $V(x) < 1$. Since $\mathcal{L}V(z) \le 0$ for $z \notin A$, we have

$$\begin{aligned} V(x) &\ge \int V(y)\mathcal{P}(x,dy) \ge \mathcal{P}(x,A) + \int_{A^c} V(y)\mathcal{P}(x,dy) \\ &\ge \mathcal{P}(x,A) + \int_{A^c} \mathcal{P}(y,A)\mathcal{P}(x,dy) + \int_{A^c}\int_{A^c} V(z)\mathcal{P}(y,dz)\mathcal{P}(x,dy) \ge \ldots \end{aligned} \tag{1.1}$$

Taking limits, we see that $V(x) \ge \mathbf{P}_x(\tau_A < \infty)$. Since $V(x) < 1$, this immediately implies that the process is transient. Conversely, if the process is transient, the function $V(x) = \mathbf{P}_x(\tau_A < \infty)$ satisfies the required conditions.

We now turn to the condition for recurrence. For $N \ge 0$, set $V_N(x) = V(x)/N$ and set $D_N = \{x : V_N(x) \ge 1\}$. It follows from the assumption that the sets D_N have finite complements. Denote furthermore $A = \{x : \mathcal{L}V(x) > 0\}$. A calculation virtually identical to (1.1) then shows that $V_N(x) \ge \mathbf{P}_x(\tau_{D_N} < \tau_A) \ge \mathbf{P}_x(\tau_A = \infty)$. In particular, one has $\mathbf{P}_x(\tau_A = \infty) = 0$ for every x, so that the process is recurrent. Conversely, assume that the process is transient and consider an arbitrary finite set A, as well as a sequence of decreasing sets D_N with finite complements and such that $\bigcap_{N>0} D_N = \phi$. We then set $W_N(x) = \mathbf{P}_x(\tau_{D_N} < \tau_A)$ with the convention that $W_N(x) = 0$ for $x \in A$ and $W_N(x) = 1$ for $x \in D_N$. It is straightforward do check that one does have $\mathcal{L}W_N \le 0$ for $x \notin A$. Furthermore, it follows from the recurrence of the process that $\lim_{N\to\infty} W_N(x) = 0$ for every x. We can therefore find a sequence $N_k \to \infty$ such that $V(x) = \sum_{k\ge 0} W_{N_k}(x) < \infty$

for every x. Since V grows at infinity by construction, it does indeed satisfy the required conditions.

We finally consider the criterion for positive recurrence. We define the set $A = \{x : \mathcal{L}V(x) > -1\}$ which is finite by assumption. For $x \notin A$ we now have

$$\begin{aligned} V(x) &\geq 1 + \int V(y)\mathcal{P}(x,dy) \geq 1 + \int_{A^c} V(y)\mathcal{P}(x,dy) \\ &\geq 1 + \mathcal{P}(x,A^c) + \int_{A^c}\int_{A^c} V(z)\mathcal{P}(y,dz)\mathcal{P}(x,dy) \geq \ldots \end{aligned}$$

Taking limits again, we obtain that $V(x) \geq \mathbf{E}_x\tau_A$, so that the first hitting time of A has finite expectation. Since it follows from our assumption that $\mathcal{L}V$ is bounded from above (in particular $\mathcal{L}V(x) < \infty$ on A), this shows that the return time to any fixed state has finite expectation. Conversely, if the process is positive recurrent, we set $V(x) = \mathbf{E}_x\tau_A$ for an arbitrary finite set A and we check as before that it does have the required properties. □

Actually the criterion for positive recurrence can be slightly strengthened:

Proposition 1.4 *if there exist functions $V, F \colon \mathcal{X} \to \mathbf{R}_+$ such that $\mathcal{L}V$ is bounded and $\mathcal{L}V(x) \leq -1 - F(x)$ for all but finitely many values of x, then the unique invariant measure π satisfies $\int F(x)\,\pi(dx) < \infty$.*

Proof. It suffices to show that π has the desired property. Denoting $V_N(x) = V(x) \wedge N$, we see that $G_N = \mathcal{L}V_N$ is bounded, negative outside a finite set, and satisfies $\lim_{N\to\infty} G_N(x) = G(x) = \mathcal{L}V(x)$. Furthermore, $\int G_N(x)\,\pi(dx) = 0$ by the invariance of π. The claim now follows from Fatou's lemma. □

Recurrence / transience of the random walk. The random walk on $\mathbf{Z}^d$ has generator

$$\mathcal{L}f(x) = \frac{1}{2d}\sum_{y\sim x}(f(y) - f(x)) .$$

Our aim is to use the above criteria to show that it is recurrent if and only if $d \leq 2$.

It is clear that it is irreducible and aperiodic. It cannot be positive recurrent since the density of any invariant probability measure would have to satisfy $\mathcal{L}\pi = 0$, so that there cannot be a point x with $p(x)$ maximal. The question of interest is whether the random walk is transient or recurrent. We use the fact that if f varies slowly, then $\mathcal{L}f(x) \approx \Delta f(x)$.

Note now that if we set $f(x) = \|x\|^{2\alpha}$, we have

$$\Delta f(x) = 2\alpha(d + 2\alpha - 2)\|x\|^{2\alpha-2} .$$

This simple calculation already shows that 2 is the critical dimension for the transience / recurrence transition since for $d < 2$ one can find $\alpha > 0$ so that f satisfies the conditions for the recurrence criterion, whereas for $d > 2$, one can find $\alpha < 0$

such that f satisfies the conditions for the transience criterion. What happens at the critical dimension? We consider

$$f(x) = (\log(x_1^2 + x_2^2))^\alpha \;,$$

so that

$$\Delta f(x) = \frac{4\alpha(\alpha-1)}{x_1^2+x_2^2}(\log(x_1^2 + x_2^2))^{\alpha-2} \;.$$

Choosing $\alpha \in (0,1)$ allows to construct a function such that the criterion for recurrence is satisfied.

1.3 Ergodic invariant measures

An important class of invariant measures are the ergodic invariant measures. Informally speaking, these are precisely those measures for which a law of large numbers holds. In order to describe them more precisely, we introduce the notion of a $\mathcal{P}$-invariant set:

Definition 1.5 Given a Markov kernel $\mathcal{P}$, a measurable set $A \subset \mathcal{P}$ is $\mathcal{P}$-invariant if $\mathcal{P}(x, A) = 1$ for every $x \in A$.

With this definition at hand, we have

Definition 1.6 An invariant probability measure μ for $\mathcal{P}$ is ergodic if every $\mathcal{P}$-invariant set has μ-measure either 0 or 1.

The importance of invariant measures can be seen in the following structural theorem, which is a consequence of Birkhoff's ergodic theorem:

Theorem 1.7 *Given a Markov kernel $\mathcal{P}$, denote by $\mathcal{I}$ the set of all invariant probability measures for $\mathcal{P}$ and by $\mathcal{E} \subset \mathcal{I}$ the set of all those that are ergodic. Then, $\mathcal{I}$ is convex and $\mathcal{E}$ is precisely the set of its extremal points. Furthermore, for every invariant measure $\mu \in \mathcal{I}$, there exists a probability measure $\mathcal{Q}_\mu$ on $\mathcal{E}$ such that*

$$\mu(A) = \int_{\mathcal{E}} \nu(A)\, \mathcal{Q}_\mu(d\nu) \;.$$

In other words, every invariant measure is a convex combination of ergodic invariant measures. Finally, any two distinct elements of $\mathcal{E}$ are mutually singular.

Remark 1.8 As a consequence, if a Markov process admits more than one invariant measure, it does admit at least two ergodic (and therefore mutually singular) ones. This leads to the intuition that, in order to guarantee the uniqueness of its invariant measure, it suffices to show that a Markov process explores its state space 'sufficiently thoroughly'.

1.4 Existence and uniqueness criteria

In this section, we present two main existence and uniqueness criteria for the invariant measure of a Markov process. First, existence essentially follows from a continuity and compactness result, which can be formulated in terms of the Feller property:

Definition 1.9 A Markov operator $\mathcal{P}$ is Feller if $\mathcal{P}\varphi$ is continuous for every continuous bounded function $\varphi\colon \mathcal{X} \to \mathbf{R}$. In other words, it is Feller if and only if the map $x \mapsto \mathcal{P}(x,\cdot)$ is continuous in the topology of weak convergence.

Theorem 1.10 (Krylov-Bogolioubov) *Let $\mathcal{P}$ be a Feller Markov operator over a Polish space $\mathcal{X}$. Assume that there exists $x_0 \in \mathcal{X}$ such that the sequence $\mathcal{P}^n(x_0,\cdot)$ is tight. Then, there exists at least one invariant probability measure for $\mathcal{P}$.*

Proof. Let μ_N be the sequence of probability measures defined by

$$\mu_N(A) = \frac{1}{N}\sum_{n=0}^{N-1} \mathcal{P}^n(x_0, A)\;. \tag{1.2}$$

Since our assumption immediately implies that $\{\mu_N\}_{N\ge 1}$ is tight, there exists at least one accumulation point μ_* and a sequence N_k with $N_k \to \infty$ such that $\mu_{N_k} \to \mu_*$ weakly. Take now an arbitrary test function $\varphi \in \mathcal{C}_b(\mathcal{X})$ and denote by $\|\cdot\|$ the supremum norm of φ. One has

$$\begin{aligned}\|(\mathcal{P}\mu_*)(\varphi) - \mu_*(\varphi)\| &= \|\mu_*(\mathcal{P}\varphi) - \mu_*(\varphi)\| = \lim_{k\to\infty} \|\mu_{N_k}(\mathcal{P}\varphi) - \mu_{N_k}(\varphi)\| \\ &= \lim_{k\to\infty} \|\mu_{N_k}(\mathcal{P}\varphi) - \mu_{N_k}(\varphi)\| = \lim_{k\to\infty} \frac{1}{N_k}\|\mathcal{P}^N\varphi - \varphi\| \\ &\le \lim_{k\to\infty} \frac{2\|\varphi\|}{N_k} = 0\;.\end{aligned}$$

Here, the second equality relies on the fact that $\mathcal{P}\varphi$ is continuous since $\mathcal{P}$ was assumed to be Feller. Since φ was arbitrary, this shows that $\mathcal{P}\mu_* = \mu_*$ as requested. □

Example 1.11 *Take $\mathcal{X} = [0,1]$ and consider the transition probabilities defined by*

$$\mathcal{P}(x,\cdot) = \begin{cases} \delta_{x/2} & \text{if } x > 0 \\ \delta_1 & \text{if } x = 0. \end{cases}$$

It is clear that this Markov operator cannot have any invariant probability measure. Indeed, assume that μ is invariant. Clearly, one must have $\mu(\{0\}) = 0$ since $\mathcal{P}(x,\{0\}) = 0$ for every x. Since, for $x \neq 0$, one has $\mathcal{P}(x,\{(1/2,1]\}) = 0$, one must also have $\mu((1/2,1]) = 0$. Proceeding by induction, we have that $\mu((1/2^n,1]) = 0$ for every n and therefore $\mu((0,1]) = 0$. Therefore, $\mu(\mathcal{X}) = 0$ which is a contradiction.

Endowing $\mathcal{X}$ with the usual topology, it is clear that the 'Feller' assumption of the Krylov-Bogolioubov criteria is not satisfied around 0*. The tightness criterion however is satisfied since $\mathcal{X}$ is a compact space. On the other hand, we could add the set $\{0\}$ to the topology of $\mathcal{X}$, therefore really interpreting it as $\mathcal{X} = \{0\} \sqcup (0, 1]$. Since $\{0\}$ already belongs to the Borel σ-algebra of $\mathcal{X}$, this change of topology does not affect the Borel sets. Furthermore, the space $\mathcal{X}$ is still a Polish space and it is easy to check that the Markov operator $\mathcal{P}$ now has the Feller property! However, the space $\mathcal{X}$ is no longer compact and a sequence $\{x_n\}$ accumulating at* 0 *is no longer a precompact set, so that it is now the tightness assumption that is no longer satisfied.*

2 Some simple uniqueness criteria

The following definition captures what we mean by the fact that a given point of the state space can be 'visited' by the dynamic:

Definition 2.1 Let $\mathcal{P}$ be a Markov operator over a Polish space $\mathcal{X}$ and let $x \in \mathcal{X}$. We say that x is *accessible* for $\mathcal{P}$ if, for every $y \in \mathcal{X}$ and every open neighborhood U of x, there exists $k > 0$ such that $\mathcal{P}^k(y, U) > 0$.

It is straightforward to show that if a given point is accessible, then it must belong to the topological support of every invariant measure of the semigroup:

Lemma 2.2 *Let $\mathcal{P}$ be a Markov operator over a Polish space $\mathcal{X}$ and let $x \in \mathcal{X}$ be accessible. Then, $x \in \operatorname{supp} \mu$ for every invariant probability measure μ.*

Proof. Let μ be invariant for the Markov operator $\mathcal{P}$ and define the resolvent kernel $\mathcal{R}$ by

$$\mathcal{R}(y, A) = \sum_{n>0} 2^{-n} \mathcal{P}^n(y, A) .$$

Clearly, the accessibility assumption implies that $\mathcal{R}(y, A) > 0$ for every $y \in \mathcal{X}$ and every neighborhood $U \subset \mathcal{X}$ of x. Then, the invariance of μ implies that

$$\mu(U) = \int_{\mathcal{X}} \mathcal{R}(y, U)\, \mu(dy) > 0 ,$$

as required. □

It is important to realise that this definition depends on the topology of $\mathcal{X}$ and not just on the Borel σ-algebra. Considering again Example 1.11, we see that the point 0 is reachable when $[0, 1]$ is endowed with its usual topology, whereas it is *not* reachable if we interpret the state space as $\{0\} \sqcup (0, 1]$. Therefore, as in the previous section, this definition can be useful only in conjunction with an appropriate regularity property of the Markov semigroup. The following example shows that the Feller property is too weak to serve our purpose.

Example 2.3 (Ising model) *The Ising model is one of the most popular toy models of statistical mechanics. It is one of the simplest models describing the evolution of a ferromagnet. The physical space is modelled by a lattice $\mathbf{Z}^d$ and the magnetisation at each lattice site is modelled by a 'spin', an element of $\{\pm 1\}$. The state space of the system is therefore given by $\mathcal{X} = \{\pm 1\}^{\mathbf{Z}^2}$, which we endow with the product topology. This topology can be metrized for example by the distance function*

$$d(x, y) = \sum_{k \in \mathbf{Z}^2} \frac{|x_k - y_k|}{2^{|k|}} ,$$

and the space $\mathcal{X}$ endowed with this distance function is easily seen to be separable.

The (Glauber) dynamic for the Ising model depends on a parameter β and can be described in the following way. At each lattice site, we consider independent clocks that ring at Poisson distributed times. Whenever the clock at a given site (say the site k) rings, we consider the quantity $\delta E_k(x) = \sum_{j \sim k} x_j x_k$, where the sum runs over all sites j that are nearest neighbors of k. We then flip the spin at site k with probability $\min\{1, \exp(-\beta\, \delta E_k(x))\}$.

Let us first show that every point is accessible for this dynamic. Fix an arbitrary configuration $x \in \mathcal{X}$ and a neighbourhood U containing x. By the definition of the product topology, U contains an 'elementary' neighbourhood $U_N(x)$ of the type $U_N(x) = \{y \in \mathcal{X} \,|\, y_k = x_k \;\forall\, |k| \le N\}$. Given now an arbitrary initial condition $y \in \mathcal{X}$, we can find a sequence of m spin flips at distinct locations $k_1, \ldots, k_m$, all of them located inside the ball $\{|k| \le N\}$, that allows to go from y into $U_N(x)$. Fix now $t > 0$. There is a very small but nevertheless strictly positive probability that within that time interval, the Poisson clocks located at $k_1, \ldots, k_m$ ring exactly once and exactly in that order, whereas all the other clocks located in the ball $\{|k| \le N + 2\}$ do not ring. Furthermore, there is a strictly positive probability that all the corresponding spin flips do actually happen. As a consequence, the Ising model is topologically irreducible in the sense that for any state $x \in \mathcal{X}$, any open set $U \subset \mathcal{X}$ and any $t > 0$, one has $\mathcal{P}_t(x, U) > 0$.

It is also relatively straightforward to show that the dynamic has the Feller property, but this is outside the scope of these notes. However, despite the fact that the dynamic is both Feller and topologically irreducible, one has the following:

Theorem 2.4 *For $d \ge 2$ there exists $\beta_c > 0$ such that the Ising model has at least two distinct invariant measures for $\beta > \beta_c$.*

The proof of this theorem is not simple and we will not give it here. It was a celebrated tour de force by Onsager to be able to compute the critical value $\beta_c = \ln(1 + \sqrt{2})/2$ explicitly in [Ons44] for the case $d = 2$. We refer to the monograph [Geo88] for a more detailed discussion of this and related models.

This example shows that if we wish to base a uniqueness argument on the accessibility of a point or on the topological irreduciblity of a system, we need to combine this with a stronger regularity property than the Feller property. One possible regularity property that yields the required properties is the *strong Feller* property:

Definition 2.5 A Markov operator $\mathcal{P}$ over a Polish space $\mathcal{X}$ has the *strong Feller* property if, for every function $\varphi \in \mathcal{B}_b(\mathcal{X})$, one has $\mathcal{P}\varphi \in \mathcal{C}_b(\mathcal{X})$.

With this definition, one has:

Proposition 2.6 *If a Markov operator $\mathcal{P}$ over a Polish space $\mathcal{X}$ has the strong Feller property, then the topological supports of any two mutually singular invariant measures are disjoint.*

Proof. Let μ and ν be two mutually singular invariant measures for $\mathcal{P}$, so that there exists a set $A \subset \mathcal{X}$ such that $\mu(A) = 1$ and $\nu(A) = 0$. The invariance of μ and ν then implies that $\mathcal{P}(x, A) = 1$ for μ-almost every x and $\mathcal{P}(x, A) = 0$ for ν-almost every x.

Set $\varphi = \mathcal{P}\mathbf{1}_A$, where $\mathbf{1}_A$ is the characteristic function of A. It follows from the previous remarks that $\varphi(x) = 1$ μ-almost everywhere and $\varphi(x) = 0$ ν-almost everywhere. Since φ is continuous by the strong Feller property, the claim now follows from the fact that if a continuous function is constant μ-almost everywhere, it must be constant on the topological support of μ. □

Corollary 2.7 *Let $\mathcal{P}$ be a strong Feller Markov operator over a Polish space $\mathcal{X}$. If there exists an accessible point $x \in \mathcal{X}$ for $\mathcal{P}$, then it can have at most one invariant measure.*

Proof. Combine Proposition 2.6 with Lemma 2.2 and the fact that if $\mathcal{I}$ contains more than one element, then by Theorem 1.7 there must be at least two distinct ergodic invariant measures for $\mathcal{P}$. □

2.1 Continuous time Markov processes

A continuous time Markov process is no longer described by a single Markov transition kernel $\mathcal{P}$, but by a family of transition kernels $\mathcal{P}_t$ satisfying the semigroup property $\mathcal{P}_{s+t} = \mathcal{P}_s\mathcal{P}_t$ and such that $\mathcal{P}_0$ is the identity: $\mathcal{P}_0(x, \cdot) = \delta_x$. Without further restrictions, a continuous-time Markov process could have very pathological properties. We will therefore always assume that $t \mapsto \mathcal{P}_t(x, A)$ is measurable for every x and every measurable set A and that, for every initial condition $x \in \mathcal{X}$, the process admits a version that is càdlàg (right-continuous with left limits) as a function of time.

In other words, we will assume that for every $x \in \mathcal{X}$, there exists a probability measure $\mathbf{P}_x$ on $\mathcal{D}(\mathbf{R}_+, \mathcal{X})$ such that its marginals on $\mathcal{X}^n$ at any finite number of times $t_1 < \ldots < t_n$ are given by the probability measure

$$\mathcal{P}_{t_1}(x, dx_1)\mathcal{P}_{t_2-t_1}(x_1, dx_2)\cdots\mathcal{P}_{t_n-t_n-1}(x_{n-1}, dx_n)\;.$$

In the case of continuous time, we say that a positive measure μ is invariant if $\mathcal{P}_t\mu = \mu$ for every $t \geq 0$. Note that in theory, it is always possible to restrict oneself to the case of discrete time in the study of the existence and uniqueness of an invariant measure:

Proposition 2.8 *Let $\mathcal{P}_t$ be a Markov semigroup over $\mathcal{X}$ and let $\mathcal{P} = \mathcal{P}_T$ for some fixed $T > 0$. Then, if μ is invariant for $\mathcal{P}$, the measure $\hat{\mu}$ defined by*

$$\hat{\mu}(A) = \frac{1}{T}\int_0^T \mathcal{P}_t\mu(A)\,dt$$

is invariant for the semigroup $\mathcal{P}_t$. □

Remark 2.9 The converse is not true at this level of generality. This can be seen for example by taking $\mathcal{P}_t(x,\cdot) = \delta_{x+t}$ with $\mathcal{X} = S^1$.

In the case of continuous-time Markov processes, it is however often convenient to formulate Lyapunov-Foster type conditions in terms of the generator $\mathcal{L}$ of the process. Formally, one has $\mathcal{L} = \partial_t\mathcal{P}_t|_{t=0}$, but it turns out that the natural domain of the generator with this definition may be too restrictive for our usage. We therefore take a rather pragmatic view of the definition of the generator $\mathcal{L}$ of a Markov process, in the sense that writing

$$\mathcal{L}F = G\;,$$

is considered to be merely a shorthand notation for the statement that the process $F(x_t,t) - \int_0^t G(x_s,s)\,ds$ is a martingale for every initial condition x_0. Similarly,

$$\mathcal{L}F \leq G\;,$$

is a shorthand notation for the statement that $F(x_t,t) - \int_0^t G(x_s,s)\,ds$ is a supermartingale for every x_0.

Remark 2.10 It is possible to have $\mathcal{L}F \leq G$ even in situations where there does not exist any function H such that $\mathcal{L}F = H$. Think of the case $F(x) = -|x|$ when the process x_t is a Brownian motion. There, one has $\mathcal{L}F \leq 0$, but F does not belong to the domain of the generator, even in the weakened sense described above.

3 Harris's theorem

The purpose of this section is to show that under a geometric drift condition and provided that $\mathcal{P}$ admits sufficiently large 'small sets', its transition probabilities converge towards a unique invariant measure at an exponential rate. This theorem dates back to [Har56] and extends the ideas of Doeblin to the unbounded state space setting. This is usually established by finding a Lyapunov function with "small" level sets [Has80, MT93]. If the Lyapunov function is strong enough, one then has a spectral gap in a weighted supremum norm [MT92, MT93].

Traditional proofs of this result rely on the decomposition of the Markov chain into excursions away from the small set and a careful analysis of the exponential tail of the length of these excursions [Num84, Cha89, MT92, MT93]. There have been other variations which have made use of Poisson equations or worked at getting explicit constants [KM05, DMR04, DMLM03, Bax05]. The proof given in the present notes is a slight simplification of the proof given in [HM10]. It is very direct, and relies instead on introducing a family of equivalent weighted norms indexed by a parameter β and to make an appropriate choice of this parameter that allows to combine in a very elementary way the two ingredients (existence of a Lyapunov function and irreducibility) that are crucial in obtaining a spectral gap. The original motivation was to provide a proof which could be extended to more general settings in which no "minorisation condition" holds. This has been applied successfully to the study of the two-dimensional stochastic Navier-Stokes equations in [HM08].

We will assume throughout this section that $\mathcal{P}$ satisfies the following geometric drift condition:

Assumption 3.1 *There exists a function $V : \mathcal{X} \to [0, \infty)$ and constants $K \geq 0$ and $\gamma \in (0, 1)$ such that*

$$(\mathcal{P}V)(x) \leq \gamma V(x) + K\;, \tag{3.1}$$

for all $x \in \mathcal{X}$.

Remark 3.2 One could allow V to also take the value $+\infty$. However, since we do not assume any particular structure on $\mathcal{X}$, this case can immediately be reduced to the present case by simply replacing $\mathcal{X}$ by $\{x : V(x) < \infty\}$.

Exercise 3.3 *Show that in the case of continuous time, a sufficient condition for Assumption 3.1 to hold is that there exists a measurable function $V : \mathcal{X} \to [0, \infty)$ and positive constants c, K such that $\mathcal{L}V \leq K - cV$.*

Assumption 3.1 ensures that the dynamic enters the "centre" of the state space regularly with tight control on the length of the excursions from the centre. We now assume that a sufficiently large level set of V is sufficiently "nice" in the sense that we have a uniform "minorisation" condition reminiscent of Doeblin's condition, but localised to the sublevel sets of V.

Assumption 3.4 *For every $R > 0$, there exists a constant $\alpha > 0$ so that*

$$\|\mathcal{P}(x, \cdot) - \mathcal{P}(y, \cdot)\|_{\mathrm{TV}} \leq 2(1 - \alpha)\,, \tag{3.2}$$

for all x, y such that $V(x) + V(y) \leq R$.

Remark 3.5 An alternative way of formulating (3.2) is to say that the bound

$$|\mathcal{P}\varphi(x) - \mathcal{P}\varphi(y)| \le 2(1-\alpha) ,$$

holds uniformly over all functions φ with absolute value bounded by 1.

In order to state the version of Harris' theorem under consideration, we introduce the following weighted supremum norm:

$$\|\varphi\| = \sup_x \frac{|\varphi(x)|}{1+V(x)} . \tag{3.3}$$

With this notation at hand, one has:

Theorem 3.6 *If Assumptions 3.1 and 3.4 hold, then $\mathcal{P}$ admits a unique invariant measure $\mu_\star$. Furthermore, there exist constants $C > 0$ and $\varrho \in (0,1)$ such that the bound*

$$\|\mathcal{P}^n\varphi - \mu_\star(\varphi)\| \le C\varrho^n \|\varphi - \mu_\star(\varphi)\|$$

holds for every measurable function $\varphi\colon \mathcal{X} \to \mathbf{R}$ such that $\|\varphi\| < \infty$.

While this result is well-known, the proofs found in the literature are often quite involved and rely on careful estimates of the return times to small sets, combined with a clever application of Kendall's lemma. See for example [MT93, Section 15].

The aim of this note is to provide a very short and elementary proof of Theorem 3.6 based on a simple trick. Instead of working directly with (3.3), we define a whole family of weighted supremum norms depending on a scale parameter $\beta > 0$ that are all equivalent to the original norm (3.3):

$$\|\varphi\|_\beta = \sup_x \frac{|\varphi(x)|}{1+\beta V(x)} .$$

The advantage of this scale of norms is that it allows us to prove that:

Theorem 3.7 *If Assumptions 3.1 and 3.4 hold, then there exist constants $\beta > 0$ and $\bar\varrho \in (0,1)$ such that the bound*

$$\|\mathcal{P}^n\varphi - \mu_\star(\varphi)\|_\beta \le \bar\varrho^n \|\varphi - \mu_\star(\varphi)\|_\beta ,$$

holds for every measurable function $\varphi\colon \mathcal{X} \to \mathbf{R}$ such that $\|\varphi\|_\beta < \infty$.

The interest of this result lies in the fact that it is possible to tune β in such a way that $\mathcal{P}$ is a strict contraction for the norm $\|\cdot\|_\beta$. In general, this does *not* imply that $\mathcal{P}$ is a contraction for $\|\cdot\|$, even though the equivalence of these two norms does of course imply that there exists $n > 0$ such that $\mathcal{P}^n$ is such a contraction.

3.1 Formulation as a Lipschitz norm

We now introduce an alternative definition of the norm $\|\cdot\|_\beta$. We begin by defining a metric d_β between points in $\mathcal{X}$ by

$$d_\beta(x,y) = \begin{cases} 0 & x = y\;, \\ 2 + \beta V(x) + \beta V(y) & x \neq y\;. \end{cases}$$

Though sightly odd looking, the reader can readily verify that since $V \geq 0$ by assumption, d_β indeed satisfies the axioms of a metric. This metric in turn induces a Lipschitz seminorm on measurable functions by

$$|\!|\!|\varphi|\!|\!|_\beta = \sup_{x\neq y} \frac{|\varphi(x) - \varphi(y)|}{d_\beta(x,y)}\;.$$

It turns out that this norm is almost identical to the one from the previous section. More precisely, one has:

Lemma 3.8 *One has the identity* $|\!|\!|\varphi|\!|\!|_\beta = \inf_{c\in\mathbf{R}} \|\varphi + c\|_\beta$.

Proof. It is obvious that $|\!|\!|\varphi|\!|\!|_\beta \leq \|\varphi\|_\beta$ and therefore $|\!|\!|\varphi|\!|\!|_\beta \leq \inf_{c\in\mathbf{R}} \|\varphi + c\|_\beta$, so it remains to show the reverse inequality.

Given any φ with $|\!|\!|\varphi|\!|\!|_\beta \leq 1$, we set $c = \inf_x(1 + \beta V(x) - \varphi(x))$. Observe that for any x and y, $\varphi(x) \leq |\varphi(y)| + |\varphi(x) - \varphi(y)| \leq |\varphi(y)| + 2 + \beta V(x) + \beta V(y)$. Hence $1 + \beta V(x) - \varphi(x) \geq -1 - \beta V(y) - |\varphi(y)|$. Since there exists at least one point with $V(y) < \infty$ we see that c is bounded from below and hence $|c| < \infty$.

Observe now that

$$\varphi(x) + c \leq \varphi(x) + 1 + \beta V(x) - \varphi(x) = 1 + \beta V(x)\;,$$

and

$$\begin{aligned} \varphi(x) + c &= \inf_y \varphi(x) + 1 + \beta V(y) - \varphi(y) \\ &\geq \inf_y 1 + \beta V(y) - |\!|\!|\varphi|\!|\!|_\beta \cdot d_\beta(x,y) \geq -(1 + \beta V(x))\;, \end{aligned}$$

so that $|\varphi(x) + c| \leq 1 + \beta V(x)$ as required. □

3.2 Proof of main theorem

Theorem 3.9 *If Assumptions 3.1 and 3.4 hold there exists an* $\bar\alpha \in (0,1)$ *and* $\beta > 0$ *such that*

$$|\!|\!|\mathcal{P}\varphi|\!|\!|_\beta \leq \bar\alpha |\!|\!|\varphi|\!|\!|_\beta\;.$$

Proof. Fix a test function φ with $|\!|\!|\varphi|\!|\!|_\beta \leq 1$. By Lemma 3.8, we can assume without loss of generality that one also has $\|\varphi\|_\beta \leq 1$. The claim then follows if we can exhibit $\bar\alpha < 1$ so that

$$|\mathcal{P}\varphi(x) - \mathcal{P}\varphi(y)| \leq \bar\alpha d_\beta(x,y)\;.$$

If $x = y$, the claim is true. Henceforth we assume $x \neq y$. We begin by assuming that x and y are such that

$$V(x) + V(y) \geq R \;. \tag{3.4}$$

Choosing any $\bar\gamma \in (\gamma, 1)$, we then have from (3.1) and (3.3) the bound

$$\begin{aligned} |\mathcal{P}\varphi(x) - \mathcal{P}\varphi(y)| &\leq 2 + \beta\mathcal{P}V(x) + \beta\mathcal{P}V(y) \\ &\leq 2 + \beta\gamma V(x) + \beta\gamma V(y) + 2\beta K \\ &\leq 2 + \beta\bar\gamma V(x) + \beta\bar\gamma V(y) + \beta(2K - (\bar\gamma - \gamma)R) \;. \end{aligned}$$

It follows that if we ensure that R is sufficiently large so that $(\bar\gamma - \gamma)R > 2K$, then there exists some $\alpha_1 < 1$ (depending on β) such that we do indeed have

$$|\mathcal{P}\varphi(x) - \mathcal{P}\varphi(y)| \leq \alpha_1 d_\beta(x, y) \;.$$

We emphasise that up to now β could be any positive number; only the precise value of α_1 depends on it (and gets "worse" for small values of β). The second part of the proof will determine a choice of $\beta > 0$.

Now consider the case of x and y such that $V(x) + V(y) \leq R$. To treat this case, we split the function φ as

$$\varphi(x) = \varphi_1(x) + \varphi_2(x) \;,$$

where

$$|\varphi_1(x)| \leq 1 \;, \qquad |\varphi_2(x)| \leq \beta V(x) \;, \qquad \forall x \in \mathcal{X} \;.$$

With this decomposition, we then obtain the bound

$$\begin{aligned} |\mathcal{P}\varphi(x) - \mathcal{P}\varphi(y)| &\leq |\mathcal{P}\varphi_1(x) - \mathcal{P}\varphi_1(y)| + |\mathcal{P}\varphi_2(x)| + |\mathcal{P}\varphi_2(y)| \\ &\leq 2(1 - \alpha) + \gamma\beta V(x) + \gamma\beta V(y) + 2\beta K \\ &\leq 2 - 2\alpha + \beta(\gamma R + 2K) \end{aligned}$$

Hence fixing for example $\beta = \alpha/(\gamma R + 2K)$, we obtain the bound

$$|\mathcal{P}\varphi(x) - \mathcal{P}\varphi(y)| \leq 2 - \alpha \leq (1 - \alpha/2) d_\beta(x, y) \;,$$

simply because $d_\beta(x, y) \geq 2$. Setting $\bar\alpha = \max\{1 - \alpha/2, \alpha_1\}$ concludes the proof. □

Remark 3.10 Actually, setting $\gamma_0 = \gamma + 2K/R < 1$, then for any $\alpha_0 \in (0, \alpha)$ one can choose $\beta = \alpha_0/K$ and $\bar\alpha = (1 - \alpha + \alpha_0) \vee (2 + R\beta\gamma_0)/(2 + R\beta)$.

4 Subgeometric rates of convergence

Let x be a strong Markov process on a metric space X with generator $\mathcal{L}$ and associated semigroup $\mathcal{P}_t$. We assume that x has a càdlàg modification and that $\mathcal{P}_t$ is Feller for every $t > 0$, so that in particular the hitting times of closed sets are stopping times. The aim of this note is to obtain a short self-contained result on the convergence rate of such a process to its invariant measure (when one exists).

The main result of this section is the following:

Theorem 4.1 *Assume that there exists a continuous function $V\colon X \to [1,\infty)$ with precompact sublevel sets such that*

$$\mathcal{L}V \le K - \varphi(V)\,, \tag{4.1}$$

for some constant K and for some strictly concave function $\varphi\colon \mathbf{R}_+ \to \mathbf{R}_+$ with $\varphi(0) = 0$ and increasing to infinity. Assume furthermore that sublevel sets of V are 'small' in the sense that for every $C > 0$ there exists $\alpha > 0$ and $T > 0$ such that

$$\|\mathcal{P}_T(x,\cdot) - \mathcal{P}_T(y,\cdot)\|_{\rm TV} \le 2(1-\alpha)\,,$$

for every (x,y) such that $V(x) + V(y) \le C$. Then the following hold:

1. *There exists a unique invariant measure μ for the Markov process x and μ is such that $\int \varphi(V(x))\,\mu(dx) \le K$.*
2. *Let H_φ be the function defined by*
$$H_\varphi(u) = \int_1^u \frac{ds}{\varphi(s)}\,.$$
Then, there exists a constant C such that for every $x, y \in X$, one has the bounds
$$\int_0^\infty (\varphi\circ H_\varphi^{-1})(t)\|\mathcal{P}_t(x,\cdot) - \mathcal{P}_t(y,\cdot)\|_{\rm TV}\,dt \le C(V(x)+V(y))$$
$$\|\mathcal{P}_t(x,\cdot) - \mathcal{P}_t(y,\cdot)\|_{\rm TV} \le C\frac{V(x)+V(y)}{H_\varphi^{-1}(t)}\,.$$
3. *There exists a constant C such that the bound*
$$\|\mathcal{P}_t(x,\cdot) - \mu\|_{\rm TV} \le \frac{CV(x)}{H_\varphi^{-1}(t)} + \frac{C}{(\varphi\circ H_\varphi^{-1})(t)}\,,$$
holds for every initial condition $x \in X$.

Remark 4.2 Since V is bounded from below by assumption, (4.1) follows immediately if one can check that the process

$$t \mapsto V(x_t) - Kt + \int_0^t (\varphi\circ V)(x_s)\,ds$$

is a *local* supermartingale.

The remainder of this section is devoted to the proof of Theorem 4.1. In a nutshell, the proof relies on the fact that if x_t and y_t are two processes that are both Markov with transition semigroup $\mathcal{P}_t$, then one has the coupling inequality

$$\|\mathcal{P}_t(x_0,\cdot) - \mathcal{P}_t(y_0,\cdot)\|_{\mathrm{TV}} \le 2\mathbf{P}(x_t \neq y_t)\;.$$

Of course, this bound will never converge to 0 if the two processes are independent, so the aim of the game is to introduce correlations in a suitable way. This will be done in Section 4.2, after some preliminary calculations that provide the main tools used in this section.

To the best of my knowledge, Theorem 4.1 was stated in the form given in these notes for the first time in the recent work [BCG08, DFG09]. However, it relies on a wealth of previous work, for example [DFMS04] for the same condition in the discrete-time case, early results by Nummelin, Tuominen and Tweedie [NT83, TT94], etc. See also [Ver99, RT99] for examples of early results on subgeometric convergence. The proof given in these notes is a simplification of the arguments given in [BCG08, DFG09] and is more self-contained as it avoids reducing oneself to the discrete-time case.

4.1 Preliminary calculations

4.1.1 Some renewal business

We start with the following result:

Lemma 4.3 *Let $H\colon \mathbf{R}_+ \to \mathbf{R}_+$ be an increasing strictly log-concave function such that $H(0) = 1$. Let $\mathcal{F}_n$ be an increasing sequence of σ-algebras over some probability space and let $\{X_n\}_{n\ge 0}$ be a sequence of positive $\mathcal{F}_n$-measurable random variables such that there exists $C_X > 0$ with $\mathbf{E}(H(X_n)\,|\,\mathcal{F}_{n-1}) < C_X$ almost surely for every n. Finally, let N be a geometrically distributed random variable such that $\{N = n\}$ is independent of $\mathcal{F}_n$ for every $n \ge 1$. Then, the random variable $X = \sum_{n=1}^N X_n$ satisfies $\mathbf{E}H(X) < \infty$.*

Proof. The strict log-concavity of H implies that for every $\varepsilon > 0$ there exists $K > 0$ such that

$$H(u+v) \le \frac{\varepsilon}{C_X} H(u)H(v)\;, \tag{4.2}$$

for every u, v such that $u \ge K$ and $v \ge K$.

Denoting $X^{(k)} = \sum_{n=1}^k X_n$, we then have

$$\begin{aligned}
H(X^{(k)}) &= \sum_{A\subset\{1,\ldots,k\}} H(X^{(k)}) \prod_{n\in A} \mathbf{1}_{\{X_n\le K\}} \prod_{m\notin A} \mathbf{1}_{\{X_m > K\}} \\
&\le \sum_{A\subset\{1,\ldots,k\}} H\Big(|A|K + \sum_{m\notin A} X_m\Big) \prod_{m\notin A} \mathbf{1}_{\{X_m > K\}} \\
&\le \sum_{A\subset\{1,\ldots,k\}} H(|A|K)\varepsilon^{|A^c|} \prod_{m\notin A} \frac{H(X_m)}{C_X}\;,
\end{aligned}$$

where we used (4.2) for the last inequality. It follows that

$$\mathbf{E}H(X^{(k)}) \le \sum_{A\subset\{1,\ldots,k\}} H(|A|K)\varepsilon^{k-|A|} \le H(kK) \sum_{A\subset\{1,\ldots,k\}} \varepsilon^{k-|A|}\;.$$

Note now that it follows from Stirling's formula for the binomial coefficients that there exists a constant C such that

$$\sum_{A\subset\{1,\ldots,k\}} \varepsilon^{k-|A|} \le C\sqrt{k}\sum_{n=0}^{k} \frac{k^k}{n^n(k-n)^{k-n}}\varepsilon^n\;.$$

Setting $n = kx$ for some $x \in [0,1]$, we deduce that

$$\mathbf{E}H(X^{(k)}) \le Ck^{3/2}H(kK)\Big(\sup_{x\in[0,1]} \frac{\varepsilon^x}{x^x(1-x)^{1-x}}\Big)^k \tag{4.3}$$

Since $\lim_{x\to 0} x^x = 1$ and the function $x \mapsto x^x$ is continuous for $x > 0$, it is possible for every $\delta > 0$ to find $\varepsilon > 0$ (and therefore $K > 0$) such that the supremum appearing in this expression is bounded by $1 + \delta$.

On the other hand, since N is geometrically distributed, there exists $\lambda > 1$ such that

$$\mathbf{P}(N = k) = (\lambda - 1)\lambda^{-k}\;, \qquad k \ge 1\;.$$

Combining this with (4.3), we deduce that

$$\mathbf{E}H(X) = (\lambda-1)\sum_{k\ge 1}\lambda^{-k}\mathbf{E}H(X^{(k)}) \le C\sum_{k\ge 1}\lambda^{-k}H(kK)(1+\delta)^k\;,$$

where we made use of the fact that $\{N = k\}$ is independent of $\mathcal{F}_k$ to get the first identity. We can first make K large enough so that $(1+\delta) < \lambda^{1/3}$. Then, we note that the strict subexponential growth of H implies that there exists a constant C such that $H(kK) \le C\lambda^{k/3}$ and the claim follows. □

4.1.2 Concave functions of semimartingales

We first make the following remark:

Proposition 4.4 *Let y be a real-valued càdlàg semimartingale and let $\Phi\colon \mathbf{R}_+ \times \mathbf{R} \to \mathbf{R}$ be a function that is $\mathcal{C}^1$ in its first argument, and $\mathcal{C}^2$ and concave in its second argument. Then, the process $\Phi(y_t) - \int_0^t \Phi'(y_{s-}, s)\,dy(s) - \int_0^t \partial_t\Phi(y_{s-}, s)\,dy(s)$ is decreasing.*

Proof. Since y is a semimartingale, we can decompose it as $y_t = A_t + M_t$ with M a martingale and A a process of finite variation. It follows from Itô's formula that

$$\Phi(y_t) = \Phi(y_0) + \int_0^t \Phi'(y_{s-}, s)\,dy(s) + \int_0^t \partial_t\Phi(y_{s-}, s)\,ds$$

$$+ \int_0^t \Phi''(y_s, s)\, d\langle M\rangle_t^c + \sum_{s\in[0,t]} \big(\Phi(y_{s+}, s) - \Phi(y_{s-}, s) - \Phi'(y_{s-}, s)\Delta y_s\big)\;,$$

where we denote by $\langle M\rangle_t^c$ the quadratic variation of the continuous part of M and we denote by Δy_s the size of the jump of y at time s. The claim then follows from the facts that $\langle M\rangle_t^c$ is an increasing process and that Φ'' is negative. □

As a consequence, we have that:

Corollary 4.5 *Let x_t be a continuous-time Markov process with generator $\mathcal{L}$ and let F and G be such that $\mathcal{L}F \le G$. Then, if Φ is as in Proposition 4.4 with the further property that $\Phi' \ge 0$, we have $\mathcal{L}\Phi(F) \le \partial_t\Phi + \Phi'(F)G$.*

Proof. Setting $y_t = F(x_t, t)$, it follows from our assumptions that one can write $dy_t = G(x_t, t)\, dt + dN_t + dM_t$, where M is a càdlàg martingale and N is a non-increasing process. It follows from Proposition 4.4 that

$$d\Phi(y_t) \le \Phi'(y_{t-}, t)(G(x_t, t)\, dt + dN_t + dM_t) + \partial_t\Phi(y_{t-}, t)\, dt\;,$$

so that the claim now follows from the fact that dN_t is a negative measure and Φ' is positive. □

In order to obtain bounds on the dynamic of x, starting from (4.1), it seems natural to consider the solution Φ to the ordinary differential equation

$$\partial_t\Phi = -\varphi\circ\Phi\;,\qquad \Phi(u, 0) = u\;.$$

This can be solved explicitly, yielding

$$\Phi(u, t) = H_\varphi^{-1}(H_\varphi(u) - t)\;,$$

where H_φ is as in the statement of Theorem 4.1. At first glance, one would 'naïvely' expect from (4.1) to have a bound of the type $\mathbf{E}V(x_t) \le \Phi(V(x_0), t)$. Unfortunately, it is easy to be convinced that the application of Jensen's inequality required to obtain such a bound does precisely go in the 'wrong direction'.

However, it turns out that one can obtain a very similar bound, namely (in the case $K = 0$) it is possible to obtain an inequality of the type $\mathbf{E}\Phi^{-1}(V(x_t), t) \le V(x_0)$! To see this, note first that one has the identity

$$\Phi^{-1}(x, t) = H_\varphi^{-1}(H_\varphi(u) + t)\;. \tag{4.4}$$

Combining this with the fact that $H_\varphi' = 1/\varphi$, it immediately follows that one has the identity

$$\partial_x\Phi^{-1}(x, t) = \frac{\partial_t\Phi^{-1}(x, t)}{\varphi(x)} = \frac{\varphi(\Phi^{-1}(x, t))}{\varphi(x)}\;.$$

Since φ is concave and increasing by assumption, this shows that Φ^{-1} is increasing and concave in its first argument for every fixed value of $t \geq 0$. It then follows from Corollary 4.5 that, provided that V satisfies (4.1), one has the bound

$$\begin{aligned}\mathcal{L}\Phi^{-1}(V(x),t) &\leq \partial_t\Phi^{-1}(V(x),t) + \partial_x\Phi^{-1}(V(x),t)\mathcal{L}V \\ &\leq \partial_t\Phi^{-1}(V(x),t) + \partial_x\Phi^{-1}(V(x),t)(K - \varphi(V(x))) \\ &= K\partial_x\Phi^{-1}(V(x),t)\;. \end{aligned} \tag{4.5}$$

If $K = 0$, the claimed bound then follows at once.

4.2 The coupling argument

We now turn to the proof of Theorem 4.1. The first point, namely the existence of an invariant measure follows immediately from the Krylov-Bogoliubov criterion, Theorem 1.10. Indeed, start with any initial condition x_0 such that $V(x_0) < \infty$ and define μ_T as in (1.2) (with the sum replaced by an integral). One then has

$$\frac{1}{T}(\mathbf{E}V(x_t) - V(x_0)) \leq K - \int_{\mathcal{X}} \varphi(V(x))\,\mu_T(dx)\;.$$

The tightness of the measures μ_T then follows immediately from the compactness of the sublevel sets of V. The required integrability also follows at one from Fatou's lemma.

Take two independent copies of the process x and define

$$W(x,y) = 2V(x) + 2V(y)\;.$$

We then have

$$\mathcal{L}_2W(x,y) \leq 4K - 2\varphi(V(x)) - 2\varphi(V(y)) \leq 4K - 2\varphi(V(x)+V(y))\;.$$

Since we assumed that φ is strictly concave, it follows that for every $C > 0$ there exists $V_C > 0$ such that $\varphi(2x) \leq 2\varphi(x) - C$, provided that $x > V_C$. It follows that there exists some V_C such that

$$\mathcal{L}_2W(x,y) \leq -\varphi(W(x,y))\;,$$

provided that $V(x) + V(y) > V_C$, and $\mathcal{L}_2W(x,y) \leq 4K$ otherwise. We denote from now on by $\mathcal{K}$ the set $\{(x,y) : V(x)+V(y) \leq V_C\}$ and by τ the first return time to this set.

Performing exactly the same calculation as in (4.5), but this time with the function $\Phi^{-1}(W(x,y),t)$, it then follows immediately that

$$\mathbf{E}_{(x,y)}H_\varphi^{-1}(\tau) \leq W(x,y)\;. \tag{4.6}$$

We can now construct a coupling between two copies of the process in the following way:

1. If the two copies are equal at some time, they remain so for future times.
2. If the two copies are different and outside of $\mathcal{K}$, then they evolve independently.
3. If the two copies are different and inside of $\mathcal{K}$, then they try to couple over the next unit time interval.

We denote by τ_C the coupling time for the two processes. By assumption, the maximal coupling has probability at least δ to succeed at every occurrence of step 3. As a consequence, we can construct the coupling in such a way that this probability is *exactly* equal to δ.

Note now that one has the coupling inequality

$$\|\mathcal{P}_t(x,\cdot) - \mathcal{P}_t(y,\cdot)\|_{\text{TV}} \le \mathbf{P}(\tau_C \ge t) , \tag{4.7}$$

which is the main tool to obtain bounds on the convergence speed. We denote by τ_n the n-th return time to $\mathcal{K}$ and by z_t the pair of processes (x_t, y_t). We also denote by $\mathcal{F}_n$ the σ-algebra generated by τ_n. Note that it then follows from (4.1) that

$$\mathbf{E}(W(z_{\tau_n+1}) \,|\, \mathcal{F}_n) \le 2V_C + 4K ,$$

so that there exists a constant C such that

$$\mathbf{E}(F(\tau_{n+1} - \tau_n) \,|\, \mathcal{F}_n) \le C ,$$

almost surely for every $n \ge 1$.

It then follows immediately from Lemma 4.3 and from (4.6) that

$$\mathbf{E}F(\tau_C) \le \mathbf{E}F(\tau_1)\,\mathbf{E}(F(\tau_C - \tau_1) \,|\, \mathcal{F}_1) \le CW(x,y) .$$

Combining this with (4.7), it follows that

$$\int_0^\infty F'(t)\, \|\mathcal{P}_t(x,\cdot) - \mathcal{P}_t(y,\cdot)\|_{\text{TV}}\, dt < C(V(x) + V(y)) ,$$

for some constant C. In particular, since the map $t \mapsto \|\mathcal{P}_t(x,\cdot) - \mathcal{P}_t(y,\cdot)\|_{\text{TV}}$ is non-increasing, we have the pointwise bound

$$\|\mathcal{P}_t(x,\cdot) - \mathcal{P}_t(y,\cdot)\|_{\text{TV}} \le \frac{C(V(x) + V(y))}{F(t)} . \tag{4.8}$$

4.3 Convergence to the invariant measure

In this section, we bound the speed of convergence towards the invariant measure, starting from an arbitrary initial condition x. Note that this is not necessarily the same speed as the one exhibited in the previous section. Indeed, if we knew that V

is integrable with respect to μ, then we would immediately obtain from (4.8) the bound

$$\begin{aligned}\|\mathcal{P}_t(x,\cdot)-\mu\|_{\mathrm{TV}} &\le \int \|\mathcal{P}_t(x,\cdot)-\mathcal{P}_t(y,\cdot)\|_{\mathrm{TV}}\,\mu(dy)\\ &\le \frac{C}{H_\varphi^{-1}(t)}\Big(V(x)+\int V(y)\,\mu(dy)\Big)\;.\end{aligned}$$

Unfortunately, all that we can deduce from (4.1) is that $\varphi(V)$ is integrable with respect to μ with

$$\int \varphi(V(x))\,\mu(dx) \le K\;.$$

This information can be used in the following way, by decomposing μ into the part where $V \le R$ and its complement:

$$\|\mathcal{P}_t(x,\cdot)-\mu\|_{\mathrm{TV}} \le \frac{C}{H_\varphi^{-1}(t)}\Big(V(x)+\int_{V\le R} V(y)\,\mu(dy)\Big)+\mu(V>R)\;.$$

This bound holds for every value of R, so we can try to optimise over it. Since we assumed that φ is concave, the function $x\mapsto\varphi(x)/x$ is decreasing so that on the set $\{V\le R\}$ one has $V(y)\le\varphi(V(y))R/\varphi(R)$. Furthermore, Chebychev's inequality yields

$$\mu(V>R)=\mu(\varphi(V)>\varphi(R))\le\frac{K}{\varphi(R)}\;.$$

Combining this with the previous bound, we obtain

$$\|\mathcal{P}_t(x,\cdot)-\mu\|_{\mathrm{TV}} \le \frac{C}{H_\varphi^{-1}(t)}\Big(V(x)+\frac{KR}{\varphi(R)}\Big)+\frac{K}{\varphi(R)}\;.$$

Setting $R=H_\varphi^{-1}(t)$, we finally obtain for some constant C the bound

$$\|\mathcal{P}_t(x,\cdot)-\mu\|_{\mathrm{TV}} \le \frac{CV(x)}{H_\varphi^{-1}(t)}+\frac{C}{(\varphi\circ H_\varphi^{-1})(t)}\;.$$

4.4 Convergence in stronger norms

Another standard set of results concerns the convergence towards the invariant measure in weighted total variation norms. For any two positive measures μ and ν on X that have densities $\mathcal{D}_\mu$ and $\mathcal{D}_\nu$ with respect to some common reference measure dx and any function $G\colon X\to[1,\infty)$, we set

$$\|\mu-\nu\|_G=\int|\mathcal{D}_\mu(x)-\mathcal{D}_\nu(x)|\,G(x)\,dx\;. \tag{4.9}$$

The usual total variation distance corresponds to the choice $G=1$. Note also that (4.9) is independent of the choice of reference measure.

With this notation, one has the following result:

Lemma 4.6 *Let $\Psi\colon \mathbf{R}_+ \to \mathbf{R}_+$ be an increasing function such that $u \mapsto \Psi(u)/u$ is increasing. Then, the bound*

$$\|\mu - \nu\|_G \le \|\mu - \nu\|_{\mathrm{TV}} \Psi^{-1}\left(\frac{\int_X (\Psi \circ G)(x)(\mu + \nu)(dx)}{\|\mu - \nu\|_{\mathrm{TV}}}\right) ,$$

holds for any pair of positive measures μ and ν such that the right hand side is finite.

Proof. From the definition of $\|\cdot\|_G$, we have

$$\|\mu - \nu\|_G \le R\|\mu - \nu\|_{\mathrm{TV}} + \int_{G(x)>R} G(x)(\mu + \nu)(dx) .$$

Since $\Psi(u)/u$ is increasing by assumption, the bound $u \le \Psi(u)R/\Psi(R)$ holds for $u \ge R$. We deduce that the bound

$$\|\mu - \nu\|_G \le R\|\mu - \nu\|_{\mathrm{TV}} + \int_X (\Psi \circ G)(x)(\mu + \nu)(dx)\frac{R}{\Psi(R)} ,$$

holds for every $R > 0$. Optimising over R concludes the proof. □

5 Lower bounds on convergence rates

The main idea for obtaining lower bounds on the convergence rate towards the invariant measure is the following. Suppose that on the one hand, we know that $\int F\, d\mu = \infty$ for some positive function F, so that we have some lower bound on the 'heaviness' of the tails of the invariant measure μ. Suppose on the other hand that we can find a function $G \gg F$ such that we have an *upper bound* on the growth of the expected value of G under the transition probabilities starting from some fixed starting point. Then, this can be turned into a lower bound on the convergence rate towards μ by arguing that if the convergence was too fast, then the non-integrability of F under μ would imply some lower bound on the expected value of G. This lower bound must be compatible with the existing upper bound.

Theorem 5.1 *Let x_t be a Markov process on a Polish space $\mathcal{X}$ with invariant measure $\mu_\star$ and let $G\colon \mathcal{X} \to [1,\infty)$ be such that:*

- *There exists a function $f\colon [1,\infty) \to [0,1]$ such that the function $\mathrm{Id}\cdot f\colon y \mapsto yf(y)$ is increasing to infinity and such that $\mu_\star(G \ge y) \ge f(y)$ for every $y \ge 1$.*
- *There exists a function $g\colon \mathcal{X} \times \mathbf{R}_+ \to [1,\infty)$ increasing in its second argument and such that $\mathbf{E}(G(x_t) \mid X_0 = x_0) \le g(x_0, t)$.*

Then, the bound

$$\|\mu_t - \mu_\star\|_{\mathrm{TV}} \ge \frac{1}{2} f((\mathrm{Id}\cdot f)^{-1}(2g(x_0,t))) , \tag{5.1}$$

holds for every $t > 0$, where μ_t is the law of x_t with initial condition $x_0 \in \mathcal{X}$.

Proof. It follows from the definition of the total variation distance and from Chebyshev's inequality that, for every $t \geq 0$ and every $y \geq 1$, one has the lower bound

$$\|\mu_t - \mu_\star\|_{\mathrm{TV}} \geq \mu_\star(G(x) \geq y) - \mu_t(G(x) \geq y) \geq f(y) - \frac{g(x_0, t)}{y} \;.$$

Choosing y to be the unique solution to the equation $yf(y) = 2g(x_0, t)$, the result follows. □

The problem is that unless $\mu_\star$ is known exactly, one does not in general have sufficiently good information on the tail behaviour of $\mu_\star$ to be able to apply Theorem 5.1 as it stands. However, it follows immediately from the proof that the bound (5.1) still holds for a subsequence of times t_n converging to ∞, provided that the bound $\mu_\star(G \geq y_n) \geq f(y_n)$ holds for a sequence y_n converging to infinity. This observation allows to obtain the following corollary that is more useful in situations where the law of the invariant measure is not known explicitly:

Corollary 5.2 *Let x_t be a Markov process on a Polish space $\mathcal{X}$ with invariant measure $\mu_\star$ and let $W\colon \mathcal{X} \to [1,\infty)$ be such that $\int W(x)\,\mu_\star(dx) = \infty$. Assume that there exist $F\colon [1,\infty) \to \mathbf{R}$ and $h\colon [1,\infty) \to \mathbf{R}$ such that:*

- *h is decreasing and $\int_1^\infty h(s)\,ds < \infty$.*
- *$F \cdot h$ is increasing and $\lim_{s\to\infty} F(s)h(s) = \infty$.*
- *There exists a function $g\colon \mathcal{X} \times \mathbf{R}_+ \to \mathbf{R}_+$ increasing in its second argument and such that $\mathbf{E}((F \circ W)(x_t)\,|\,x_0) \leq g(x_0, t)$.*

Then, for every $x_0 \in \mathcal{X}$, there exists a sequence of times t_n increasing to infinity such that the bound

$$\|\mu_{t_n} - \mu_\star\|_{\mathrm{TV}} \geq h((F \cdot h)^{-1}(g(x_0, t_n)))$$

holds, where μ_t is the law of x_t with initial condition $x_0 \in \mathcal{X}$.

Proof. Since $\int W(x)\,\mu_\star(dx) = \infty$, there exists a sequence w_n increasing to infinity such that $\mu_\star(W(x) \geq w_n) \geq 2h(w_n)$, for otherwise we would have the bound

$$\int W(x)\,\mu_\star(dx) = 1 + \int_1^\infty \mu_\star(W(x) \geq w)\,dw \leq 1 + 2\int_1^\infty h(w)\,dw < \infty \;,$$

thus leading to a contradiction. Applying Theorem 5.1 with $G = F \circ W$ and $f = 2h \circ F^{-1}$ concludes the proof. □

6 Malliavin calculus and Hörmander's theorem

One of the main ingredients in the convergence results of the previous sections was the existence of sufficiently large 'small sets' in which the transition probabilities of our Markov process can be uniformly bounded from below. One situation where this is relatively easy to show is that of a process with transition probabilities that have a continuous density with respect to some reference measure, i.e.

$$\mathcal{P}_t(x, dy) = p_t(x, y)\, \mu(dy) \;,$$

where p_t is continuous in both arguments. In that case, every point $x \in \mathcal{X}$ admits a neighbourhood U_x such that U_x is 'small'. If furthermore there exists $x_\star$ such that $x_\star \in \operatorname{supp} \mathcal{P}_t(x, \cdot)$ for every $x \in X$, then it is a straightforward exercise to check that every compact subset of $\mathcal{X}$ is small.

The aim of this section is to provide a reasonably self-contained proof of Hörmander's criterion which ensures that the transition probabilities of a diffusion with smooth coefficients have a smooth density with respect to Lebesgue measure. Our main object of study in this section is a stochastic differential equation of the form

$$dx = V_0(x)\, dt + \sum_{i=1}^m V_i(x) \circ dW_i \;, \tag{6.1}$$

where the V_i's are smooth vector fields on $\mathbf{R}^n$. We assume that these vector fields assume the coercivity assumptions necessary so that the solution to (6.1) is $\mathcal{C}^\infty$ with respect to its initial condition. An import tool for our analysis will be the linearisation of (6.1) with respect to its initial condition. This is given by the stochastic process $J_{0,t}$ defined as the solution to the SDE

$$dJ_{0,t} = DV_0(x)\, J_{0,t}\, dt + \sum_{i=1}^m DV_i(x)\, J_{0,t} \circ dW_i \;. \tag{6.2}$$

Higher order derivatives $J_{0,t}^{(k)}$ with respect to the initial condition can be defined similarly. Throughout this section, we will make the following standing assumption:

Assumption 6.1 *The vector fields V_i are $\mathcal{C}^\infty$ and all of their derivatives grow at most polynomially at infinity. Furthermore, they are such that*

$$\mathbf{E} \sup_{t \le T} |x_t|^p < \infty \;, \qquad \mathbf{E} \sup_{t \le T} |J_{0,t}^{(k)}|^p < \infty \;, \qquad \mathbf{E} \sup_{t \le T} |J_{0,t}^{-1}|^p < \infty \;,$$

for every initial condition $x_0 \in \mathbf{R}^n$, every terminal time $T > 0$, and every $p > 0$.

Note here that the inverse $J_{0,t}^{-1}$ of the Jacobian can be found by solving the SDE

$$dJ_{0,t}^{-1} = -J_{0,t}^{-1}\, DV_0(x)\, dt - \sum_{i=1}^m J_{0,t}^{-1}\, DV_i(x) \circ dW_i \;. \tag{6.3}$$

The aim of this section is to show that under a certain non-degeneracy assumption on the vector fields V_i, the law of the solution to (6.1) has a smooth density with respect to Wiener measure. To describe this non-degeneracy condition, recall that the Lie bracket $[U,V]$ between two vector fields U and V is the vector field defined by

$$[U,V](x) = DV(x)\,U(x) - DU(x)\,V(x)\;.$$

This notation is consistent with the usual notation for the commutator between two linear operators since, if we denote by A_U the first-order differential operator acting on smooth functions f by $A_U f(x) = \langle V(x), \nabla f(x)\rangle$, then we have the identity $A_{[U,V]} = [A_U, A_V]$.

With this notation at hand, we give the following definition:

Definition 6.2 Given an SDE (6.1), define a collection of vector fields V_k by

$$\mathrm{V}_0 = \{V_i \,:\, i > 0\}\;,\quad \mathrm{V}_{k+1} = \mathrm{V}_k \cup \{[U,V_j] \,:\, U \in \mathrm{V}_k \,\&\, j \ge 0\}\;.$$

We also define the vector spaces $\mathrm{V}_k(x) = \mathrm{span}\{V(x) \,:\, V \in \mathrm{V}_k\}$. We say that (6.1) satisfies the *parabolic Hörmander condition* if $\bigcup_{k\ge 1} \mathrm{V}_k(x) = \mathbf{R}^n$ for every $x \in \mathbf{R}^n$.

With these notations, Hörmander's theorem can be formulated as

Theorem 6.3 *Consider (6.1) and assume that Assumption 6.1 holds. If the corresponding vector fields satisfy the parabolic Hörmander condition, then its solutions admit a smooth density with respect to Lebesgue measure.*

A complete rigorous proof of Theorem 6.3 is far beyond the scope of these notes. However, we hope to be able to give a convincing argument showing why this result is true and what are the main steps involved in its probabilistic proof. The interested reader can find the technical details required to make the proof rigorous in [Mal78, KS84, KS85, KS87, Nor86, Nua95]. Hörmander's original, completely different, proof using fractional integrations can be found in [Hör67]. A yet completely different functional-analytic proof using the theory of pseudo-differential operators was developed by Kohn in [Koh78] and can also be found in [Hör85].

6.1 Simplified Malliavin calculus

The main tool in the proof is the Malliavin calculus with its integration by part formula in Wiener space, which was developed precisely in order to provide a probabilistic proof of Theorem 6.3. It essentially relies on the fact that the image of a Gaussian measure under a smooth submersion that is sufficiently integrable possesses a smooth density with respect to Lebesgue measure. This can be shown in the following way. First, one observes the following fact, which follows from standard Sobolev embedding results:

Lemma 6.4 *Let μ be a probability measure on $\mathbf{R}^n$ such that the bound*

$$\Big|\int_{\mathbf{R}^n} D^{(k)}G(x)\,\mu(dx)\Big| \le C_k \|G\|_\infty\;,$$

holds for every smooth bounded function G and every $k \ge 1$. Then μ has a smooth density with respect to Lebesgue measure. □

Consider now a sequence of N independent Gaussian random variables δw_k with variances δt_k for $k \in \{1, \ldots, N\}$, as well as a smooth map $X : \mathbf{R}^N \to \mathbf{R}^n$. We also denote by w the collection $\{\delta w_k\}_{k\geq 1}$ and we define the $n \times n$ matrix-valued map

$$\mathrm{M}_{ij}(w) = \sum_k \partial_k X_i(w) \partial_k X_j(w)\, \delta t_k \;,$$

where we use ∂_k as a shorthand for the partial derivative with respect to the variable δw_k. With this notation, X being a surjection is equivalent to $\mathrm{M}(w)$ being invertible for every w. We then have the following result:

Theorem 6.5 *Let X be as above, assume that $\mathrm{M}(w)$ is invertible for every w and that, for every $p > 1$ and every $m \geq 0$, we have*

$$\mathbf{E}|\partial_{k_1} \cdots \partial_{k_m} X(w)|^p < \infty \;, \qquad \mathbf{E}\|\mathrm{M}(w)^{-1}\|^p < \infty \;. \tag{6.4}$$

Then the law of $X(w)$ has a smooth density with respect to Lebesgue measure.

Before we turn to the proof of Theorem 6.5, note that if F_k and G are square integrable functions with square integrable derivatives, then we have the integration by parts formula

$$\begin{aligned}\mathbf{E} \sum_k \partial_k G(w) F_k(w)\, \delta t_k &= \mathbf{E} G(w) \sum_k F_k(w)\, \delta w_k - \mathbf{E} G(w) \sum_k \partial_k F_k(w)\, \delta t_k \\ &\stackrel{\text{def}}{=} \mathbf{E} G(w) \int F\, dw \;. \end{aligned} \tag{6.5}$$

Here we *defined* the Skorokhod integral $\int F\, dw$ by the expression on the first line. This Skorokhod integral satisfies the following isometry:

Proposition 6.6 *Let F_k be square integrable functions with square integrable derivatives, then*

$$\begin{aligned}\mathbf{E}\Big(\int F\, dw\Big)^2 &= \sum_k \mathbf{E} F_k^2(w)\, \delta t_k + \sum_{k,\ell} \mathbf{E} \partial_k F_\ell(w) \partial_\ell F_k(w)\, \delta t_k\, \delta t_\ell \\ &\leq \sum_k \mathbf{E} F_k^2(w)\, \delta t_k + \sum_{k,\ell} \mathbf{E} (\partial_k F_\ell(w))^2\, \delta t_k\, \delta t_\ell \;,\end{aligned}$$

holds.

Proof. It follows from the definition that one has the identity

$$\mathbf{E}\Big(\int F\, dw\Big)^2 = \sum_{k,\ell} \mathbf{E}(F_k F_\ell\, \delta w_k \delta w_\ell + \partial_k F_k \partial_\ell F_\ell\, \delta t_k \delta t_\ell - 2 F_k \partial_\ell F_\ell\, \delta w_k \delta t_\ell) \;.$$

Applying the identity $\mathbf{E} G\, \delta w_\ell = \mathbf{E} \partial_\ell G\, \delta t_\ell$ the first term in the above formula (with $G = F_k F_\ell\, \delta w_k$), we thus obtain

$$\ldots = \sum_{k,\ell} \mathbf{E}(F_k F_\ell \delta_{k,\ell} \delta t_\ell + \partial_k F_k\, \partial_\ell F_\ell\, \delta t_k \delta t_\ell + (F_\ell \partial_\ell F_k - F_k \partial_\ell F_\ell)\, \delta w_k \delta t_\ell) \;.$$

Applying the same identity to the last term then finally leads to

$$\ldots = \sum_{k,\ell} \mathbf{E}(F_k F_\ell \delta_{k,\ell} \delta t_\ell + \partial_k F_\ell \, \partial_\ell F_k \, \delta t_k \delta t_\ell) ,$$

which is the desired result. □

Proof of Theorem 6.5. We want to show that Lemma 6.4 can be applied. For $\eta \in \mathbf{R}^n$, we then have from the definition of M the identity

$$(D_j G)(X(w)) = \sum_{k,m} \partial_k (G(X(w))) \partial_k X_m(w) \, \delta t_k \, \mathrm{M}^{-1}_{mj}(w) . \tag{6.6}$$

Combining this identity with (6.5), it follows that

$$\mathbf{E} D_j G(X) = \mathbf{E} G(X(w)) \sum_m \int \partial . X_m(w) \, \mathrm{M}^{-1}_{mj}(w) \, dw .$$

Combining this with Proposition 6.6 and (6.4) immediately shows that the requested result holds for $k = 1$. Higher values of k can be treated similarly by repeatedly applying (6.6). □

6.2 Back to diffusions

The results in the previous section strongly suggest that one can define a "Malliavin derivative" operator D, acting on random variables and returning a stochastic process, that has all the usual properties of a derivative and such that

$$\mathrm{D}_t \int_0^T f(s) \, dW(s) = f(t) , \quad t \in [0,T] ,$$

for smooth (deterministic) functions f. If on the other hand f is a stochastic process, then the following chain rule holds:

$$\mathrm{D}_t \int_0^T f(s) \, dW(s) = f(t) + \mathrm{D}_t \int_0^T \mathrm{D}_t f(s) \, dW(s) , \quad t \in [0,T] ,$$

Using this identity, it then follows from differentiating (6.1) on both sides that one has for $r \le t$ the identity

$$\mathrm{D}^j_r X(t) = \int_r^t DV_0(X_s) \, \mathrm{D}^j_r X_s \, ds + \sum_{i=1}^m \int_r^t DV_i(X_s) \, \mathrm{D}^j_r X_s \circ dW_i(s) + V_j(X_r) .$$

We see that this equation is identical (save for the initial condition at time $t = r$ given by $V_j(X_r)$) to the equation giving the derivative of X with respect to its initial condition! Denoting by $J_{s,t}$ the Jacobian for the stochastic flow between times s and t, we therefore have for $s < t$ the identity

$$\mathrm{D}^j_s X_t = J_{s,t} V_j(X_s) . \tag{6.7}$$

(Since X is adapted, we have $D_s^j X_t = 0$ for $s \geq t$.) Note that the Jacobian has the composition property $J_{0,t} = J_{s,t} J_{0,s}$, so that $J_{s,t}$ can be rewritten in terms of the 'usual' Jacobian as $J_{s,t} = J_{0,t} J_{0,s}^{-1}$.

We now denote by $\mathrm{A}_{0,t}$ the operator $\mathrm{A}_{0,t} v = \int_0^t J_{s,t} V(X_s) v(s)\, ds$, where v is a square integrable, not necessarily adapted, $\mathbf{R}^m$-valued stochastic process and V is the $n \times m$ matrix-valued function obtained by concatenating the vector fields V_j for $j = 1, \ldots, m$. This allows us to define the Malliavin covariance matrix $\mathrm{M}_{0,t}$ of X_t in the same way as in the previous section by

$$\mathrm{M}_{0,t} = \mathrm{A}_{0,t} \mathrm{A}_{0,t}^* = \int_0^t J_{s,t} V(X_s) V^*(X_s) J_{s,t}^*\, ds \;.$$

Note that $\mathrm{M}_{0,t}$ is a random positive definite $n \times n$ matrix. We then have:

Proposition 6.7 *Consider a diffusion of the type (6.1) satisfying Assumption 6.1. If $\mathrm{M}_{0,t}$ is almost surely invertible and that $\mathbf{E}\|\mathrm{M}_{0,t}^{-1}\|^p < \infty$ for every $p > 0$, then the transition probabilities of (6.1) have a $\mathcal{C}^\infty$ density with respect to Lebesgue measure for every $t > 0$.*

Proof. This is essentially a version of Theorem 6.5. The technical details required to make it rigorous are well beyond the scope of these notes and can be found for example in [Nua95]. □

The remainder of this section is devoted to a proof of Hörmander's theorem, which gives a simple (but essentially sharp!) criterion for the invertibility of the Malliavin matrix of a diffusion process. Actually, it turns out that for technical reasons, it is advantageous to rewrite the Malliavin matrix as

$$\mathrm{M}_{0,t} = J_{0,t} \mathrm{C}_{0,t} J_{0,t}^* \;, \qquad \mathrm{C}_{0,t} = \int_0^t J_{0,s}^{-1} V(X_s) V^*(X_s) (J_{0,s}^{-1})^*\, ds \;,$$

where $\mathrm{C}_{0,t}$ is the *reduced Malliavin matrix* of our diffusion process. Then since we assumed that $J_{0,t}$ has inverse moments of all orders, the invertibility of $\mathrm{M}_{0,t}$ is equivalent to that of $\mathrm{C}_{0,t}$. Note first that since $\mathrm{C}_{0,t}$ is a positive definite symmetric matrix, the norm of its inverse is given by

$$\|\mathrm{C}_{0,t}^{-1}\| = \Big(\inf_{|\eta|=1} \langle \eta, \mathrm{C}_{0,t} \eta \rangle \Big)^{-1} \;.$$

A very useful observation is then the following:

Lemma 6.8 *Let M be a symmetric positive semidefinite $n \times n$ matrix-valued random variable such that $\mathbf{E}\|M\|^p < \infty$ for every $p \geq 1$ and such that, for every $p \geq 1$ there exists C_p such that*

$$\sup_{|\eta|=1} \mathbf{P}(\langle \eta, M\eta \rangle < \varepsilon) \leq C_p \varepsilon^p \;, \tag{6.8}$$

holds for every $\varepsilon \leq 1$. Then, $\mathbf{E}\|M^{-1}\|^p < \infty$ for every $p \geq 1$.

Proof. The non-trivial part of the result is that the supremum over η is taken outside of the probability in (6.8). For $\varepsilon > 0$, let $\{\eta_k\}_{k\leq N}$ be a sequence of vectors with $|\eta_k| = 1$ such that for every η with $|\eta| = 1$, there exists k such that $|\eta_k - \eta| \leq \varepsilon^2$. It is clear that one can find such a set with $N \leq C\varepsilon^{2-2n}$ for some $C > 0$ independent of ε. We then have the bound

$$\begin{aligned}\langle \eta, M\eta\rangle &= \langle \eta_k, M\eta_k\rangle + \langle \eta - \eta_k, M\eta\rangle + \langle \eta - \eta_k, M\eta_k\rangle \\ &\geq \langle \eta_k, M\eta_k\rangle - 2\|M\|\varepsilon^2\;,\end{aligned}$$

so that

$$\begin{aligned}\mathbf{P}\Big(\inf_{|\eta|=1}\langle \eta, M\eta\rangle \leq \varepsilon\Big) &\leq \mathbf{P}\Big(\inf_{k\leq N}\langle \eta_k, M\eta_k\rangle \leq 4\varepsilon\Big) + \mathbf{P}\Big(\|M\| \geq \frac{1}{\varepsilon}\Big) \\ &\leq C\varepsilon^{2-2n}\sup_{|\eta|=1}\mathbf{P}\Big(\langle \eta, M\eta\rangle \leq 4\varepsilon\Big) + \mathbf{P}\Big(\|M\| \geq \frac{1}{\varepsilon}\Big)\;.\end{aligned}$$

It now suffices to use (6.8) for p large enough to bound the first term and Chebychev's inequality combined with the moment bound on $\|M\|$ to bound the second term. □

As a consequence of this, Thoerem 6.3 is a corollary of:

Theorem 6.9 *Consider (6.1) and assume that Assumption 6.1 holds. If the corresponding vector fields satisfy the parabolic Hörmander condition then, for every initial condition $x \in \mathbf{R}^n$, we have the bound*

$$\sup_{|\eta|=1}\mathbf{P}(\langle \eta, \mathrm{C}_{0,1}\eta\rangle < \varepsilon) \leq C_p\varepsilon^p\;,$$

for suitable constants C_p and all $p \geq 1$.

Remark 6.10 The choice $t = 1$ as the final time is of course completely arbitrary. Here and in the sequel, we will always consider functions on the time interval $[0, 1]$.

Before we turn to the proof of this result, we introduce a useful notation. Given a family $A = \{A_\varepsilon\}_{\varepsilon\in(0,1]}$ of events depending on some parameter $\varepsilon > 0$, we say that A is 'almost true' if, for every $p > 0$ there exists a constant C_p such that $\mathbf{P}(A_\varepsilon) \geq 1 - C_p\varepsilon^p$ for all $\varepsilon \in (0, 1]$. Similarly for 'almost false'. Given two such families of events A and B, we say that 'A almost implies B' and we write $A \Rightarrow_\varepsilon B$ if $A \setminus B$ is almost false. It is straightforward to check that these notions behave as expected (almost implication is transitive, finite unions of almost false events are almost false, etc). Note also that these notions are unchanged under any reparametrisation of the form $\varepsilon \mapsto \varepsilon^\alpha$ for $\alpha > 0$. Given two families X and Y of real-valued random variables, we will similarly write $X \leq_\varepsilon Y$ as a shorthand for the fact that $\{X_\varepsilon \leq Y_\varepsilon\}$ is 'almost true'.

Before we proceed, we state the following useful result, where $\|\cdot\|_\infty$ denotes the L^∞ norm and $\|\cdot\|_\alpha$ denotes the best possible α-Hölder constant.

Lemma 6.11 *Let $f\colon [0,1] \to \mathbf{R}$ be continuously differentiable and let $\alpha \in (0,1]$. Then, the bound*

$$\|\partial_t f\|_\infty = \|f\|_1 \le 4\|f\|_\infty \max\Big\{1, \|f\|_\infty^{-\frac{1}{1+\alpha}} \|\partial_t f\|_\alpha^{\frac{1}{1+\alpha}}\Big\}$$

holds, where $\|f\|_\alpha$ denotes the best α-Hölder constant for f.

Proof. Denote by x_0 a point such that $|\partial_t f(x_0)| = \|\partial_t f\|_\infty$. It follows from the definition of the α-Hölder constant $\|\partial_t f\|_{\mathcal{C}^\alpha}$ that $|\partial_t f(x)| \ge \frac{1}{2}\|\partial_t f\|_\infty$ for every x such that $|x - x_0| \le (\|\partial_t f\|_\infty / 2\|\partial_t f\|_{\mathcal{C}^\alpha})^{1/\alpha}$. The claim then follows from the fact that if f is continuously differentiable and $|\partial_t f(x)| \ge A$ over an interval I, then there exists a point x_1 in the interval such that $|f(x_1)| \ge A|I|/2$. □

With these notations at hand, we have the following statement, which is essentially a quantitative version of the Doob-Meyer decomposition theorem. Originally, it appeared in [Nor86], although some form of it was already present in earlier works. The proof given here is a further simplification of the arguments in [Nor86].

Lemma 6.12 (Norris) *Let W be an m-dimensional Wiener process and let A and B be $\mathbf{R}$ and $\mathbf{R}^m$-valued adapted processes such that, for $\alpha = \frac{1}{3}$, one has $\mathbf{E}(\|A\|_\alpha + \|B\|_\alpha)^p < \infty$ for every p. Let Z be the process defined by*

$$Z_t = Z_0 + \int_0^t A_s\,ds + \int_0^t B_s\,dW(s)\;. \tag{6.9}$$

Then, there exists a universal constant $r \in (0,1)$ such that one has

$$\{\|Z\|_\infty < \varepsilon\} \quad \Rightarrow_\varepsilon \quad \{\|A\|_\infty < \varepsilon^r\}\ \&\ \{\|B\|_\infty < \varepsilon^r\}\;.$$

Proof. Recall the exponential martingale inequality [RY99, p. 153], stating that if M is any continuous martingale with quadratic variation process $\langle M\rangle(t)$, then

$$\mathbf{P}\Big(\sup_{t\le T} M(t) \ge x \quad \& \quad \langle M\rangle(T) \le y\Big) \le \exp(-x^2/2y)\;,$$

for every positive T, x, y. With our notations, this immediately implies that for any $q < \frac{1}{2}$, one has the almost implication

$$\{\|\langle M\rangle\|_\infty < \varepsilon\} \quad \Rightarrow_\varepsilon \quad \{\|M\|_\infty < \varepsilon^q\}\;. \tag{6.10}$$

With this bound in mind, we apply Itô's formula to Z^2, so that

$$Z_t^2 = Z_0^2 + 2\int_0^t Z_s\,A_s\,ds + 2\int_0^t Z_s\,B_s\,dW(s) + \int_0^t B_s^2\,ds\;. \tag{6.11}$$

Since $\|A\|_\infty \le_\varepsilon \varepsilon^{-1/2}$ (or any other negative exponent for that matter) by assumption and similarly for B, it follows from this and (6.10) that

$$\{\|Z\|_\infty < \varepsilon\} \quad \Rightarrow_\varepsilon \quad \Big\{\Big|\int_0^1 A_s\, Z_s\, ds\Big| \le \varepsilon^{\frac{1}{2}}\Big\} \ \&\ \Big\{\Big|\int_0^1 B_s\, Z_s\, dW(s)\Big| \le \varepsilon^{\frac{1}{2}}\Big\}\;.$$

Inserting these bounds back into (6.11) and applying Jensen's inequality then yields

$$\{\|Z\|_\infty < \varepsilon\} \quad \Rightarrow_\varepsilon \quad \Big\{\int_0^1 B_s^2\, ds \le \varepsilon^{1/2}\Big\} \quad \Rightarrow \quad \Big\{\int_0^1 |B_s|\, ds \le \varepsilon^{1/4}\Big\}\;.$$

We now use the fact that $\|B\|_\alpha \le_\varepsilon \varepsilon^{-q}$ for every $q > 0$ and we apply Lemma 6.11 with $\partial_t f(t) = |B_t|$ (we actually do it component by component), so that

$$\{\|Z\|_\infty < \varepsilon\} \quad \Rightarrow_\varepsilon \quad \{\|B\|_\infty \le \varepsilon^{1/17}\}\;,$$

say. In order to get the bound on A, note that we can again apply the exponential martingale inequality to obtain that this 'almost implies' the martingale part in (6.9) is 'almost bounded' in the supremum norm by $\varepsilon^{1/18}$, so that

$$\{\|Z\|_\infty < \varepsilon\} \quad \Rightarrow_\varepsilon \quad \Big\{\Big\|\int_0^\cdot A_s\, ds\Big\|_\infty \le \varepsilon^{1/18}\Big\}\;.$$

Finally applying again Lemma 6.11 with $\partial_t f(t) = A_t$, we obtain that

$$\{\|Z\|_\infty < \varepsilon\} \quad \Rightarrow_\varepsilon \quad \{\|A\|_\infty \le \varepsilon^{1/80}\}\;,$$

and the claim follows with $r = 1/80$. □

Remark 6.13 By making α arbitrarily close to $1/2$, keeping track of the different norms appearing in the above argument, and then bootstrapping the argument, it is possible to show that

$$\{\|Z\|_\infty < \varepsilon\} \quad \Rightarrow_\varepsilon \quad \{\|A\|_\infty \le \varepsilon^p\} \ \&\ \{\|B\|_\infty \le \varepsilon^q\}\;,$$

for p arbitrarily close to $1/5$ and q arbitrarily close to $3/10$. This is a slight improvement over the exponent $1/8$ that was originally obtained in [Nor86]. (Note however that the two results are not necessarily comparable since [Nor86] used L^2 norms instead of L^∞ norms.)

We now have all the necessary tools to prove Theorem 6.9:

Proof of Theorem 6.9. We fix some initial condition $x \in \mathbf{R}^n$ and some unit vector $\eta \in \mathbf{R}^n$. With the notation introduced earlier, our aim is then to show that

$$\{\langle \eta, \mathrm{C}_{0,1}\eta\rangle < \varepsilon\} \quad \Rightarrow_\varepsilon \quad \phi\;.$$

At this point, we introduce for an arbitrary vector field F on $\mathbf{R}^n$ the process Z_F defined by

$$Z_F(t) = \langle \eta, J_{0,t}^{-1} F(x_t)\rangle\;,$$

so that

$$\langle \eta, \mathrm{C}_{0,1}\eta\rangle = \sum_{k=1}^m \int_0^1 |Z_{V_k}(t)|^2\,dt \ge \sum_{k=1}^m \Bigl(\int_0^1 |Z_{V_k}(t)|\,dt\Bigr)^2\;. \tag{6.12}$$

The processes Z_F have the nice property that they solve the stochastic differential equation

$$dZ_F(t) = Z_{[F,V_0]}(t)\,dt + \sum_{i=1}^m Z_{[F,V_k]}(t)\circ dW_k(t)\;.$$

This can be rewritten in Itô form as

$$dZ_F(t) = \Bigl(Z_{[F,V_0]}(t) + \sum_{k=1}^m \frac{1}{2} Z_{[[F,V_k],V_k]}(t)\Bigr)\,dt + \sum_{i=1}^m Z_{[F,V_k]}(t)\,dW_k(t)\;. \tag{6.13}$$

Since we assumed that all derivatives of the V_j grow at most polynomially, we deduce from the Hölder regularity of Brownian motion that Z_F does indeed satisfy the assumptions on its Hölder norm required for the application of Norris' lemma.

We deduce from it that one has the implication

$$\{\|Z_F\|_\infty < \varepsilon\} \quad \Rightarrow_\varepsilon \quad \{\|Z_{[F,V_k]}\|_\infty < \varepsilon^r\}\;\&\;\{\|Z_G\|_\infty < \varepsilon^r\}\;,$$

for $k = 1,\ldots,m$ and for $G = [F,V_0] + \frac{1}{2}\sum_{k=1}^m [[F,V_k],V_k]$. Iterating this bound a second time, we obtain that

$$\{\|Z_F\|_\infty < \varepsilon\} \quad \Rightarrow_\varepsilon \quad \{\|Z_{[[F,V_k],V_\ell]}\|_\infty < \varepsilon^{r^2}\}\;,$$

so that we finally obtain the implication

$$\{\|Z_F\|_\infty < \varepsilon\} \quad \Rightarrow_\varepsilon \quad \{\|Z_{[F,V_k]}\|_\infty < \varepsilon^{r^2}\}\;, \tag{6.14}$$

for $k = 0,\ldots,m$.

At this stage, we are basically done. Indeed, combining (6.12) with Lemma 6.11 as above, we see that

$$\{\langle \eta, \mathrm{C}_{0,1}\eta\rangle < \varepsilon\} \quad \Rightarrow_\varepsilon \quad \{\|Z_{V_k}\|_\infty < \varepsilon^{1/5}\}\;.$$

Applying (6.14) iteratively, we see that for every $k > 0$ there exists some $q_k > 0$ such that

$$\{\langle \eta, \mathrm{C}_{0,1}\eta\rangle < \varepsilon\} \quad \Rightarrow_\varepsilon \quad \bigcap_{V\in \mathrm{V}_k} \{\|Z_V\|_\infty < \varepsilon^{q_k}\}\;.$$

Since $Z_V(0) = \langle \eta, V(x_0)\rangle$ and since there exists some $k > 0$ such that $\mathrm{V}_k(x_0) = \mathbf{R}^n$, the right hand side of this expression is empty for some sufficiently large value of k, which is precisely the desired result. □

7 Examples

In this section, we go through a number of examples where it is possible, using the results presented in these notes, to give very sharp upper and lower bounds on the rate of convergence to equilibrium.

7.1 Gradient dynamic

This is the simplest case of a reversible dynamic on $\mathbf{R}^n$, namely the solution to the SDE

$$dx = -\nabla V(x)\,dt + \sqrt{2}dW(t)\;, \tag{7.1}$$

where W is a d-dimensional Wiener process. In this case, the diffusion is elliptic, so it is straightforward that the assumptions of Hörmander's theorem are satisfied, so that it has smooth transition probabilities, provided that V is smooth. We assume that $V(x)$ is behaves like $|x|^{2k}$ at ∞, in the sense that the bounds

$$c|x|^{2k} \le V(x) \le C|x|^{2k}\;,\quad \langle x, \nabla V(x)\rangle \ge c|x|^{2k}\;,\quad |D^2V(x)| \le C|x|^{2k-2}\;,$$

hold for some positive constants c and C outside of some sufficiently large compact set. We also assume that all derivatives of V grow at most polynomially fast.

It then follows from Itô's formula that the generator of (7.1) is given by

$$\mathcal{L} = \Delta - \nabla V\,\nabla\;.$$

In particular, we have the identity

$$\mathcal{L}\exp(\alpha V) = \alpha(\Delta V + (\alpha-1)|\nabla V|^2)\exp(\alpha V)\;.$$

It follows that if $k \ge \frac{1}{2}$, then there exists $\alpha > 0$ and $C > 0$ such that

$$\mathcal{L}\exp(\alpha V) \le -C\exp(\alpha V)\;,$$

outside of some compact set. In particular, the assumptions of Harris's theorem hold and we do obtain exponential converge towards the invariant measure.

What about $k \in (0, \frac{1}{2})$? In this case, the conditions of Harris's theorem do not hold, but if we set $W(x) = \exp(\alpha V(x))$ for some $\alpha < 1$, then we have

$$\mathcal{L}W \le -C|x|^{4k-2}W \le -C\frac{W}{(\log W)^{\frac{1}{k}-2}}\;, \tag{7.2}$$

so that we can apply Theorem 4.1 with $\varphi(x) = Cx/(\log x)^\gamma$ for $\gamma = \frac{1}{k} - 2$. In this case, we have

$$H_\varphi(u) \propto \int_1^u (\log x)^\gamma\,\frac{dx}{x} = \int_0^{\log u} y^\gamma\,dy = \frac{(\log u)^{\gamma+1}}{\gamma+1}\;,$$

so that $H_\varphi^{-1}(u) \propto \exp(cu^{1/(\gamma+1)})$, for some $c > 0$. This immediately yields an upper bound on the convergence towards the invariant measure of the type

$$\|\mathcal{P}_t(x,\cdot) - \mu_\star\|_{\mathrm{TV}} \le C\exp(\alpha V(x) - ct^{\frac{k}{1-k}})\;. \tag{7.3}$$

Let us now turn to the lower bounds. Consider the function

$$G(x) = \exp(\kappa V(x)) ,$$

for a constant κ to be determined later. It is then a straightforward calculation to check, similarly to before, that there exists a constant C such that

$$\mathcal{L}G(x) \le C \frac{G(x)}{(\log G(x))^{\frac{1}{k}-2}} .$$

Note that the difference between this and (7.2) is that the right hand side is *positive* rather than negative. It follows from Jensen's inequality that the quantity $g(t) = \mathbf{E}G(x_t)$ satisfies the differential inequality

$$\frac{dg(t)}{dt} \le \frac{Cg(t)}{(\log g(t))^{\frac{1}{k}-2}} ,$$

so that there exist some constants c and C such that

$$\mathbf{E}G(x_t) \le CG(x_0)e^{ct^{k/(1-k)}} .$$

On the other hand, since the invariant measure for our system is known to be given by

$$\mu_\star(dx) \propto \exp(-V(x))\, dx ,$$

one can check that the bound

$$\mu_\star(G > R) = \mu_\star(\kappa V(x) > \log R) \ge CR^{-\gamma} ,$$

is valid for some $C > 0$ and some $\gamma > 0$ for sufficiently large values of R. The value of γ can be made arbitrarily small by increasing the value of κ. This enables us to apply Theorem 5.1 with $f(R) = CR^{-\gamma}$ and $g(x,t) = G(x)e^{ct^{k/(1-k)}}$, so that we obtain a lower bound of the form

$$\|\mathcal{P}_t(x,\cdot) - \mu_\star\|_{\mathrm{TV}} \ge C \exp\Big(\frac{\gamma\kappa}{1-\gamma} V(x) - ct^{\frac{k}{1-k}}\Big) ,$$

which essentially matches the upper bound previously obtained in (7.3).

7.2 Renewal processes

Let ν be a probability distribution on $\mathbf{R}_+ \setminus \{0\}$ with finite expectation. Then the renewal process with holding time distribution ν is a point process on $\mathbf{R}$ such the times between successive events are independent and distributed with law ν. This renewal process can be described by the Markov process x_t on $\mathbf{R}_+$ that is equal to the time left until the next event. The dynamic of x_t is given by $\dot{x}_t = -1$ as long as $x_t > 0$. When it reaches zero, then it jumps instantaneously to a positive value distributed according to ν. It is possible to convince oneself that the invariant measure μ of x is given by

$$\mu_\star(dx) = c\,x\,\nu(dx)\;, \tag{7.4}$$

for a suitable normalisation constant $c > 0$.

The generator of this process is then given by $\mathcal{L}f(x) = -\partial_x f(x)$, but a differentiable function f belongs to the domain of $\mathcal{L}$ only if

$$f(0) = \int_0^\infty f(x)\,\nu(dx)\;. \tag{7.5}$$

Suppose now that ν has a density p with respect to Lebesgue measure such that

$$\frac{c_-}{x^\zeta} \le p(x) \le \frac{c_+}{x^\zeta}\;,$$

for some $\zeta > 2$. How fast does such a process converge to its invariant measure? A natural choice of a Lyapunov function is to take V such that $V(x) = x^\alpha$ for $x > 1$ and to adjust it on $[0, 1]$ in such a way that the compatibility condition (7.5) holds. Note that this is possible if and only if one chooses $\alpha < \zeta - 1$.

In that case, one has

$$\mathcal{L}V(x) \sim -x^{\alpha-1} = -V^{1-\frac{1}{\alpha}}\;,$$

for sufficiently large values of x, so that the conditions of Theorem 4.1 hold with $\varphi(x) = x^\gamma$ and $\gamma = 1 - \frac{1}{\alpha}$. In this case, $H_\varphi(u) \sim u^{1/\alpha}$, so that we have an upper bound on the convergence towards the invariant measure given by

$$\|\mathcal{P}_t(x,\cdot) - \mu_\star\|_{\mathrm{TV}} \le \frac{C|x|^\alpha}{t^{\alpha-1}}\;. \tag{7.6}$$

Since α can be taken arbitrarily close to $\zeta - 1$, we have

$$\limsup_{t\to\infty} \frac{\log \|\mathcal{P}_t(x,\cdot) - \mu_\star\|_{\mathrm{TV}}}{\log t} \le 2 - \zeta\;. \tag{7.7}$$

Lower bounds can be obtained in a similar way by using the fact that $\mathcal{L}V$ is bounded from above, so that $\mathbf{E}x_t^\alpha \le x_0^\alpha + Ct$ for some constant C. Furthermore, since we know by (7.4) that the invariant measure $\mu_\star$ has a density behaving like $x^{1-\zeta}$ for large values of x, we have the bound

$$\mu_\star(V > R) = \mu_\star(x > R^{1/\alpha}) \sim R^{-\frac{\zeta-2}{\alpha}}\;.$$

This enables us again to apply Theorem 5.1 with $f(R) = CR^{-\frac{\zeta-2}{\alpha}}$ and $g(x,t) = |x|^\alpha + Ct$, so that we obtain a lower bound of the form

$$\|\mathcal{P}_t(x,\cdot) - \mu_\star\|_{\mathrm{TV}} \ge C(|x|^\alpha + t)^{\frac{\zeta-2}{\zeta-2-\alpha}}\;.$$

Taking again α close to $\zeta - 1$ and combining this with (7.7), we see that one does actually have the identity

$$\lim_{t\to\infty} \frac{\log \|\mathcal{P}_t(x,\cdot) - \mu_\star\|_{\mathrm{TV}}}{\log t} = 2 - \zeta\;.$$

7.3 Conclusive remarks

We have seen in the previous two examples that Theorems 4.1 and 5.1 are surprisingly sharp in the sense that we are able to match upper and lower bounds almost exactly. This is not a special feature of these two examples, but a widespread property of Markov processes that exhibit subgeometric convergence. In a nutshell, the reason is that Lyapunov function techniques essentially reduce a potentially very complicated Markov process to a one-dimensional situation, since all the information on the process is simply encoded by the value $V(x_t)$. It is not surprising then that these techniques tend to be sharp for one-dimensional systems, but one might wonder whether they still perform well for higher dimensional systems.

The reason why this is still the case is that the convergence of Markov process that exhibit subgeometric convergence to stationarity is dominated by some relatively rare but very long excursions into "bad regions" from which the process takes very long times to escape. In both of the previous examples, these were the regions where $|x|$ is large, but it may potentially be much more complicated regions.

Very often, there will be one "worst case scenario" degree of freedom that varies on a very slow timescale when the process gets trapped in one of these bad regions. The aim of the game is then to gain a good understanding of the mechanisms that allows one to construct a Lyapunov function that captures the dynamic of this slow degree of freedom. A good example of situations where this philosophy can be implemented in a more complicated setting is given in [HM09, Hai09].

References

[Bax05] P. H. BAXENDALE. Renewal theory and computable convergence rates for geometrically ergodic Markov chains. *Ann. Appl. Probab.* **15**, no. 1B, (2005), 700–738.

[BCG08] D. BAKRY, P. CATTIAUX, and A. GUILLIN. Rate of convergence for ergodic continuous Markov processes: Lyapunov versus Poincaré. *J. Funct. Anal.* **254**, no. 3, (2008), 727–759.

[Cha89] K. S. CHAN. A note on the geometric ergodicity of a Markov chain. *Adv. in Appl. Probab.* **21**, no. 3, (1989), 702–704.

[DFG09] R. DOUC, G. FORT, and A. GUILLIN. Subgeometric rates of convergence of f-ergodic strong Markov processes. *Stochastic Process. Appl.* **119**, no. 3, (2009), 897–923.

[DFMS04] R. DOUC, G. FORT, E. MOULINES, and P. SOULIER. Practical drift conditions for subgeometric rates of convergence. *Ann. Appl. Probab.* **14**, no. 3, (2004), 1353–1377.

[DMLM03] P. DEL MORAL, M. LEDOUX, and L. MICLO. On contraction properties of Markov kernels. *Probability Theory and Related Fields* **126**, no. 3, (2003), 395–420.

[DMR04] R. DOUC, E. MOULINES, and J. S. ROSENTHAL. Quantitative bounds on convergence of time-inhomogeneous Markov chains. *The Annals of Applied Probability* **14**, no. 4, (2004), 1643–1665.

[Geo88] H.-O. GEORGII. *Gibbs measures and phase transitions*, vol. 9 of *de Gruyter Studies in Mathematics*. Walter de Gruyter & Co., Berlin, 1988.

[Hai09] M. HAIRER. How hot can a heat bath get? *Comm. Math. Phys.* **292**, no. 1, (2009), 131–177.

[Har56] T. E. HARRIS. The existence of stationary measures for certain Markov processes. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, vol. II*, 113–124. University of California Press, Berkeley and Los Angeles, 1956.

[Has80] R. Z. HAS′MINSKII. *Stochastic stability of differential equations*, vol. 7 of *Monographs and Textbooks on Mechanics of Solids and Fluids: Mechanics and Analysis*. Sijthoff & Noordhoff, Alphen aan den Rijn, 1980. Translated from the Russian by D. Louvish.

[HM08] M. HAIRER and J. C. MATTINGLY. Spectral gaps in Wasserstein distances and the 2D stochastic Navier-Stokes equations. *Ann. Probab.* **36**, no. 6, (2008), 2050–2091.

[HM09] M. HAIRER and J. C. MATTINGLY. Slow energy dissipation in anharmonic oscillator chains. *Comm. Pure Appl. Math.* **62**, no. 8, (2009), 999–1032.

[HM10] M. HAIRER and J. C. MATTINGLY. Yet another look at Harris' ergodic theorem for Markov chains. In *Seminar on Stochastic Analysis, Random Fields and Applications VI*, Progr. Probab. Birkhäuser, Basel, 2010. To appear.

[Hör67] L. HÖRMANDER. Hypoelliptic second order differential equations. *Acta Math.* **119**, (1967), 147–171.

[Hör85] L. HÖRMANDER. *The Analysis of Linear Partial Differential Operators I–IV*. Springer, New York, 1985.

[KM05] I. KONTOYIANNIS and S. P. MEYN. Large deviations asymptotics and the spectral theory of multiplicatively regular markov processes. *Electronic Journal of Probability* **10**, (2005), no. 3, 61123 (electronic).

[Koh78] J. J. KOHN. Lectures on degenerate elliptic problems. In *Pseudodifferential operator with applications (Bressanone, 1977)*, 89–151. Liguori, Naples, 1978.

[KS84] S. KUSUOKA and D. STROOCK. Applications of the Malliavin calculus. I. In *Stochastic analysis (Katata/Kyoto, 1982)*, vol. 32 of *North-Holland Math. Library*, 271–306. North-Holland, Amsterdam, 1984.

[KS85] S. KUSUOKA and D. STROOCK. Applications of the Malliavin calculus. II. *J. Fac. Sci. Univ. Tokyo Sect. IA Math.* **32**, no. 1, (1985), 1–76.

[KS87] S. KUSUOKA and D. STROOCK. Applications of the Malliavin calculus. III. *J. Fac. Sci. Univ. Tokyo Sect. IA Math.* **34**, no. 2, (1987), 391–442.

[Mal78] P. MALLIAVIN. Stochastic calculus of variations and hypoelliptic operators. *Proc. Intern. Symp. SDE* 195–263.

[MT92] S. P. MEYN and R. L. TWEEDIE. Stability of Markovian processes. I. Criteria for discrete-time chains. *Adv. in Appl. Probab.* **24**, no. 3, (1992), 542–574.

[MT93] S. P. MEYN and R. L. TWEEDIE. *Markov chains and stochastic stability*. Communications and Control Engineering Series. Springer-Verlag London Ltd., London, 1993.

[Nor86] J. NORRIS. Simplified Malliavin calculus. In *Séminaire de Probabilités, XX, 1984/85*, vol. 1204 of *Lecture Notes in Math.*, 101–130. Springer, Berlin, 1986.

[NT83] E. NUMMELIN and P. TUOMINEN. The rate of convergence in Orey's theorem for Harris recurrent Markov chains with applications to renewal theory. *Stochastic Process. Appl.* **15**, no. 3, (1983), 295–311.

[Nua95] D. NUALART. *The Malliavin calculus and related topics*. Probability and its Applications (New York). Springer-Verlag, New York, 1995.

[Num84] E. NUMMELIN. *General irreducible Markov chains and nonnegative operators*, vol. 83 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 1984.

[Ons44] L. ONSAGER. Crystal statistics. I. A two-dimensional model with an order-disorder transition. *Phys. Rev. (2)* **65**, (1944), 117–149.

[RT99] G. O. ROBERTS and R. L. TWEEDIE. Bounds on regeneration times and convergence rates for Markov chains. *Stochastic Process. Appl.* **80**, no. 2, (1999), 211–229.

[RY99] D. REVUZ and M. YOR. *Continuous martingales and Brownian motion*, vol. 293 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, third ed., 1999.

[TT94] P. TUOMINEN and R. L. TWEEDIE. Subgeometric rates of convergence of f-ergodic Markov chains. *Adv. in Appl. Probab.* **26**, no. 3, (1994), 775–798.

[Ver99] A. Y. VERETENNIKOV. On polynomial mixing and the rate of convergence for stochastic differential and difference equations. *Teor. Veroyatnost. i Primenen.* **44**, no. 2, (1999), 312–327.