Ergodicity of diffusion processes

Lazić, Petra

Doctoral thesis / Disertacija

2022

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zagreb, Faculty of Science / Sveučilište u Zagrebu, Prirodoslovno-matematički fakultet**

Permanent link / Trajna poveznica: https://urn.nsk.hr/urn:nbn:hr:217:478121

Rights / Prava: In copyright / Zaštićeno autorskim pravom.

Download date / Datum preuzimanja: **2025-01-03**

Repository / Repozitorij:

Repository of the Faculty of Science - University of Zagreb

University of Zagreb

FACULTY OF SCIENCE
DEPARTMENT OF MATHEMATICS

Petra Lazić

Ergodicity of Diffusion Processes

DOCTORAL DISSERTATION

Zagreb, 2022.

University of Zagreb

FACULTY OF SCIENCE

DEPARTMENT OF MATHEMATICS

Petra Lazić

Ergodicity of Diffusion Processes

DOCTORAL DISSERTATION

Supervisor:

izv. prof. dr. sc. Nikola Sandrić

Zagreb, 2022.

Sveučilište u Zagrebu

PRIRODOSLOVNO–MATEMATIČKI FAKULTET

MATEMATIČKI ODSJEK

Petra Lazić

Ergodičnost procesa difuzija

DOKTORSKI RAD

Mentor:

izv. prof. dr. sc. Nikola Sandrić

Zagreb, 2022.

Acknowledgements

I am utterly grateful to my supervisor Dr. Nikola Sandrić for guiding me through my PhD studies and helping me to do my first steps as a young researcher and mathematician. I immensely appreciate the hours he spent discussing and planning my research with me, checking my results, correcting my mistakes and answering all the questions in the quickest time possible. I always knew I could ask for help if I had any problems.

I am specially thanking Ivana for being the best roommate in the world, listening to my problems, cheering me up in the hard times and always saying the right word of support.

I am also extremely grateful to Marija for being the best possible colleague and always finding time to listen to me and help me. Thank you for constant support, positive energy and the best cakes.

A special thanks goes to my best friend Ana for always being there for me when I needed it, no matter what, for all the walks and talks.

Last but not least, I deeply thank my family for all the help and support during all these years, for putting up with me and standing by my side throughout good and bad times.

SUMMARY

The topic of this work is ergodicity (stochastic stability) of variuos types of stochastic processes. The urge for analysis of random processes exists in every area of science and real life - medicine, biology, chemistry, physics, finance, etc., as many phenomena naturally exhibit some sort of random behaviour in their movement. Mathematical models used to describe those random movements are called stochastic differential equations (SDEs). Since the solutions of SDEs often have a very complicated structure or are impossible to obtain explicitly, it is usually hard to analyse them directly. Therefore, the emphasis has been placed on analysing their long-term stability. This includes detecting their equilibria (stacionary distributions) as well as the rate at which convergence occurs. The convergence is observed with respect to some appropriate distance function. In my work the emphasis has been placed on the quantitative aspect of this problem, namely, on finding explicit bounds on the rate of the convergence with respect to two distance functions: to total variation distance and the class od Wasserstein distances (which provides convergence in some weaker sense). As most of the existing results in this area correspond to the geometric ergodicity (that is, the case when the rate of the convergence is exponential), and known conditions ensuring sub-goemetric ergodicity are far from being optimal because there are many known examples of sub-geometrically ergodic systems that do not satisfy those conditions, the focus of my work is to find sharp and general conditions in terms of coefficients of the process that will ensure sub-geometric ergodicity of a wide range of processes.

The results of my research can be divided in two parts. Firstly, I will consider classical diffusion processes - Markov processes with continuous trajectories (here, the class of processes with singular diffusion coefficients will be of special interest as they were not investigated so far).

The second part of the research deals with somewhat more complicated random processes - diffusion processes with Markovian switching, which are processes that, beside the continuous, diffusive one, have a second, discrete component which changes the behaviour of the process at random times. This theory is relatively new so here we still have many interesting open questions and uninvestigated phenomena that are not characteristic for classical diffusion processs.

Also, in both cases, I will extend the results on a class of processes with jumps.

Keywords: stochastic differential equations, diffusion processes, diffusion processes with Markovian switching, ergodicity, sub-geometric ergodicity, φ-irreducibility, aperiodicity, total variation distance, Wasserstein distance.

SAŽETAK

Svrha mog rada je istražiti problem ergodičnosti (stohastičke stabilnosti) različitih tipova slučajnih procesa. Slučajni procesi iznimno su važni jer se koriste za modeliranje pojava u gotovo svim područjima znanosti i svakodnevnog života – primjenjuju se u medicini, biologiji, kemiji, fizici, financijama itd. Naime, u svim tim područjima često se dolazi do zaključka da se pojave ne mogu opisati determinističkim modelima jer su neki aspekti njihovog ponašanja slučajni. Matematički modeli koji se koriste za opisivanje ovakvih pojava su stohastičke diferencijalne jednadžbe (SDJ). Međutim, budući da rješenja SDJ-ova imaju jako kompliciranu strukturu i iznimno ih je teško analizirati direktnim putem, naglasak se stavlja na analizu njihove dugoročne stabilnosti. To uključuje određivanje njihovih ekvilibrija (stacionarnih distribucija), ali i brzine kojom konvergiraju prema tim ekvilibrijima. Konvergencija se promatra s obzirom na neku određenu funkciju udaljenosti. U mojem istraživanju naglasak je stavljen na kvantitativni aspekt ovog problema, odnosno na eksplicitne ocjene brzine konvergencije s obzirom na dvije funkcije udaljenosti: udaljenost totalne varijacije i klasu Wassersteinovih udaljenosti (koja predstavlja konvergenciju u nešto slabijem smislu). Kako se većina dosadašnjih rezultata odnosi na geometrijsku ergodičnost (tj. slučaj kada je brzina konvergencije eksponencijalna), a poznati uvjeti za subgeometrijsku ergodičnost nisu blizu optimalnih jer su poznati mnogi subgeometrijski ergodični sustavi koji te uvjete ne zadovoljavaju, fokus mog rada je pronaći oštre i opće uvjete u terminima koeficijenata samog procesa koji će osigurati subgeometrijsku ergodičnost široke klase procesa.

Rezultate mog istraživanja mogu podijeliti u dvije cjeline. U prvoj ću proučavati klasične difuzije - Markovljeve procese neprekidnih trajektorija (gdje će od posebnog značaja biti klasa procesa sa singularnim difuzijskim koeficijentima koji do sada nisu razmatrani).

Drugi dio rada proučava nešto složenije procese - difuzije sa slučajnim prebacivanjem, koji osim neprekidne, difuzijske komponente sadrže i drugu, diskretnu komponentu koja u slučajnim trenucima mijenja ponašanje procesa. Ova teorija je relativno nova pa tu ima još puno zanimljivih otvorenih pitanja i neistraženih pojava koje nisu karakteristične za klasične procese difuzija.

Također, u oba slučaja, rezultate ću primijeniti na klasu procesa sa skokovima.

Ključne riječi: stohastičke diferencijalne jednadžbe, difuzije, difuzije sa slučajnim prebacivanjem, ergodičnost, subgeometrijska ergodičnost, φ-ireducibilnost, aperiodičnost, udaljenost totalne varijacije, Wassersteinova udaljenost.

PROŠIRENI SAŽETAK

Svrha mog rada je istražiti problem ergodičnosti (stohastičke stabilnosti) različitih tipova slučajnih procesa. Slučajni procesi iznimno su važni jer se koriste za modeliranje pojava u gotovo svim područjima znanosti i svakodnevnog života – primjenjuju se u medicini, biologiji, kemiji, fizici, financijama itd. Naime, u svim tim područjima često se dolazi do zaključka da se pojave ne mogu opisati determinističkim modelima jer su neki aspekti njihovog ponašanja slučajni. Matematički modeli koji se koriste za opisivanje ovakvih pojava su stohastičke diferencijalne jednadžbe (SDJ). Međutim, budući da rješenja SDJ-ova imaju jako kompliciranu strukturu i iznimno ih je teško analizirati direktnim putem, naglasak se stavlja na analizu njihove dugoročne stabilnosti. To uključuje određivanje njihovih ekvilibrija (stacionarnih distribucija), ali i brzine kojom konvergiraju prema tim ekvilibrijima. Konvergencija se promatra s obzirom na neku određenu funkciju udaljenosti. U mojem istraživanju naglasak je stavljen na kvantitativni aspekt ovog problema, odnosno na eksplicitne ocjene brzine konvergencije marginalnih distribucija procesa prema invarijantnoj mjeri s obzirom na dvije funkcije udaljenosti: udaljenost totalne varijacije i klasu Wassersteinovih udaljenosti (koja predstavlja konvergenciju u nešto slabijem smislu). Kako se većina dosadašnjih rezultata odnosi na geometrijsku ergodičnost (tj. slučaj kada je brzina konvergencije eksponencijalna), a poznati uvjeti za subgeometrijsku ergodičnost nisu blizu optimalnih jer su poznati mnogi subgeometrijski ergodični sustavi koji te uvjete ne zadovoljavaju, fokus mog rada je pronaći oštre i opće uvjete u terminima koeficijenata samog procesa koji će osigurati subgeometrijsku ergodičnost široke klase procesa.

Rezultate mog istraživanja mogu podijeliti u dvije cjeline. U prvoj ću proučavati klasične difuzije - Markovljeve procese neprekidnih trajektorija (gdje će od posebnog značaja biti klasa procesa sa singularnim difuzijskim koeficijentima koji do sada nisu razmatrani).

Drugi dio rada proučava nešto složenije procese - difuzije sa slučajnim prebacivanjem, koji osim neprekidne, difuzijske komponente sadrže i drugu, diskretnu komponentu koja u slučajnim trenucima mijenja ponašanje procesa. Ova teorija je relativno nova pa tu ima još puno zanimljivih otvorenih pitanja i neistraženih pojava koje nisu karakteristične za klasične procese difuzija. Tako su za ovaj tip procesa do sada promatrane samo eksponencijalne brzine konvergencije.

Također, u oba slučaja, rezultate ću primijeniti na dvije klase procesa sa skokovima. Prvu klasu procesa ćemo dobiti tako da skokove dodamo putem infinitezimalnog generatora (dodajući u generator Lévyjevu jezgru skokova ν). Drugu klasu procesa sa skokovima ću dobiti subordiniranjem u smislu Bochnera: napravit ću promjenu u vremenskoj komponenti našeg procesa - umjesto da se uzme njegovu vrijednost u trenutku t, uzima se vrijednost u trenutku S_t određena nekim rastućim nenegativnim procesom $\{S_t\}_{t\geq 0}$ kojeg nazivamo subordinatorom.

Za dobivanje ergodičnosti koristit ću dvije osnovne metode. U slučaju udaljenosti totalne varijacije, primjenit ću tzv. Lyapunovljevu metodu energije. Cilj ove metode je pronaći odgovarajuću testnu funkciju za koju će tzv. Lyapunovljev uvjet drifta biti zadovoljen. Dodatno, da bi se dobila konvergenciju s obzirom na udaljenost totalne varijacije, proces mora imati određena strukturalna svojstva koja će osigurati određenu razinu regularnosti. Preciznije, mora biti φ-ireducibilan i aperdiodičan. U slučaju kada to nije zadovoljeno, konvergenciju moramo promatrati u nekom slabijem smislu. Tada se tipično u literaturi koristi klasa Wassersteinovih udaljenosti kao funkcija udaljenosti između mjera. U tom slučaju, za dobivanje ergodičnosti primijenit ću tzv. metodu sparivanja refleksijom. Tu će za dobivanje brzine konvergencije biti ključna pomoćna lema koja je zapravo nadogradnja poznate Grönwallove leme. Obje metode su već poznate u literaturi, no ne u obliku koji bi bio prikladan za ovaj tip procesa pa ću ih morati proširiti i prilagoditi.

Kako su strukturalna svojstva procesa nužna da bismo dobili konvergenciju, u radu ću također razmotriti i neke dovoljne uvjete za dobivanje tih svojstava za sve tipove procesa, a koji će biti jednostavniji za provjeriti od same definicije tih svojstava.

Osim predstavljanja samih rezultata i metoda, dat ću i primjere koji će ilustrirati primjenu samih teorema na nekim konkretnim procesima.

Contents

Introduction to the Topic

Stochastic modelling has come to play a very important role in science and industry. This is due to the fact that randomness is incorporated in the behaviour of many phenomena appearing in nature and everyday life. Numerous applications include modelling of financial markets, automatic control of stochastic systems, turbulent fluent and gas molecule movements, modelling of population dynamics, disease transmission etc.

Mathematical models used to describe such movements are called stochastic differential equations (SDEs). However, real solutions of SDEs are often not explicitly known, are hard to obtain or given in a closed form suitable to applications. They also often have a very complicated structure so it is not easy to analyse them directly. For all these reasons, the emphasis has been placed on analysing the stability of the considered stochastic model. Accordingly, the first step is to seek for conditions that ensure a stable behaviour of the model. That includes detecting both the equilibria as well as the rates at which convergence occurs. Finding the rates is very important as it allows us to predict movements of our process through time and detect expected period of stabilization. Such questions have been tackled before for various types of processes. Depending on the class of the process, certain aspects of the problem have been fully analysed, but usually most of them are still completely or partially unsettled. In some cases the solution is given only within very large boundaries. Therefore, the need for deeper studying of stochastic stability and searching for more efficient solutions constantly exists.

This work studies such questions for two classes of processes: classical diffusion processes and diffusion processes with switching. In classical literature a term diffusion process corresponds to a stochastic process that has continuous trajectories, that is, the process whose movement through time can be displayed by a continuous curve. Naturally, the behaviour of many phenomena is much more complex than that and cannot be captured by such models so they have gradually been generalized. One of the first steps in that direction is to include jumps: this represents a situation in which the sample path of our process exhibits jumps of random amplitude at random time points, and moves continuously in between. The behaviour of classical diffusion (that is, the type of the observed curve) is not changing with time: after the jump, the process continues to move in the same way as before. Further generalization allows the process to "change its mine" and start to move in a different way at some point. The idea behind this upgrade of the problem comes from observing hybrid models.

A random process is a movement that can not be completely predicted. Every time we start the process from the beginning, it will exhibit a different curve. However, some regularities in its behaviour do exist: certain values will be more or less likely to happen, depending on the time point. For a fixed time point t, the probabilities of being at the state x from the state space is determined by the so-called distribution function. The distribution changes as time changes: it is not the same all the time. So, many things in the behaviour of the process are random, but under certain conditions, we are hoping to be able to predict its movement in the future to some extent, or at least to be able to control it. The more control we have over the process, the nicer the process is to us: and nice processes will be called stable.

Stability for random processes does not mean that we can determine its value in the future since the movement is always random, but rather that we can control its distribution with time. Precisely, for a stable process the distribution at time t is starting to look more and more like some fixed distribution, usually denoted by π, as time passes by, that is, as t grows. This problem can be schematically portrayed like this:

$$r(t)\,\|\,\mathbb{P}_t - \pi\,\|\;\underset{t\to+\infty}{\overset{\mathbf{?}}{\longrightarrow}}\;0,$$

where

- $\mathbb{P}_t$ ... distribution of the process at time t
- π ... equilibrium distribution
- $r(t)$ ... rate of the convergence

- $\|\cdot\|$ ... an appropriate norm (function that measures how far away we are at time t from the equilibrium)

As the scheme suggests, the problem raises several questions (and we will try to answer all of them in this work):

⇝ under which conditions does the unique equilibrium state exist?

⇝ if we use function $\|\cdot\|$ to measure the distance between two states (for us, states are distributions), how far away our process is from that equilibrium at time t?

⇝ does this distance becomes smaller with time and, possibly, goes to 0 as time goes by? If so, we will say that our process converges to its equilibrium.

⇝ at which rate does the convergence occur?

⇝ do we get the same results for some more general processes, namely, for processes that jump at random time points?

⇝ can we extend this analysis to some even more general stochastic processes like hybrid models (where we have a discrete component that changes the behaviour of our process at random time points)?

To answer these questions, we will adopt two main techniques: the so called *Lyapunov energy method* and the *coupling method*.

Lyapunov energy method relies on the idea of Aleksandr M. Lyapunov from 1892 which states that the energy of the stable system does not dissipate. How does this work? Idea, that comes from physics, is to find a function that describes the total energy of the system (*Lyapunov function*) and analyse its behaviour:

- if V is such a function and x a current state of the system (i.e. $V(x)$ is the energy of the system when it is in state x, so $\dot{V}(x)$ can be interpreted as velocity of movement or dissipation of that energy), we want

$$\dot{V}(x) = \langle \nabla V(x), \dot{x} \rangle \leq 0$$

  - $\dot{V}(x) \geq 0 \Longrightarrow$ the system is unstable as its energy has a tendency to leave the center of the state space (the energy is dissipated in the space)

- $\dot{V}(x) \leq 0 \Longrightarrow$ the system is stable as its total energy is moving towards the center of the state space (and therefore, it is being ”contained” in the sytem)

On the other hand, the idea of the coupling method is to construct the optimal coupling strategy - in the sense of the optimal transport theory. This theory is concerned with the problem of minimizing the cost of the transport of goods between producers (factories) and consumers. If each factory can produce a limited number of products, the question is which products should be delivered to which consumer so that all consumers get the quantity they need and the total cost of the transport is as small as possible. Naturally, the cost of the transport depends on the distance between the factory and a consumer: the farther they are, the bigger the cost. In this sense, we will use this method to minimize the dissipation of the energy of the system - we will try to keep it as low as possible.

1. Preliminaries and Notations

After the non-formal introduction to the topic of this work, we start by introducing notation and presenting objects and mathematical models that will be studied by giving their formal definitions together with an overview of well-known results in the area of research. The literature used includes [IW89], [Dur96], [LP17], [Zit15], [SV06], [GS79], [MT93a], [MT93b].

All objects we consider will be connected to some probability space. Therefore, we first define the setting that we will work around.

Fix a complete probability space $(\Omega, \mathscr{F}, \mathbb{P})$ with a right-continuous increasing family $\{\mathscr{F}_t\}_{t \geq 0}$ of σ-fields of $\mathscr{F}$. We call the space $(\Omega, \mathscr{F}, \{\mathscr{F}_t\}_{t \geq 0}, \mathbb{P})$ a **stochastic basis** that satisfies the usual conditions, and $\{\mathscr{F}_t\}_{t \geq 0}$ a **filtration** on $(\Omega, \mathscr{F}, \mathbb{P})$.

Let I be some time space of interest (for example, some discrete set such as $\mathbb{N}$, $[0, +\infty\rangle$ or some finite interval $[0, T]$). A **stochastic process** is a family of random variables $X = \{X_t(\omega) : t \in I\}$ with values in $\mathbb{R}^d$. We call X a *discrete* or *continuous* stochastic process depending on set I being discrete or continuous set. Next, we list some properties a stochastic process X might posses: X is called

- **adapted** to $\{\mathscr{F}_t\}_{t \geq 0}$ if X_t is $\mathscr{F}_t$-measurable for all $t \in I$;
- **predictable** if the mapping $(t, \omega) \mapsto X_t(\omega)$ is $(\mathscr{S}, \mathscr{B}(\mathbb{R}^d))$-measurable, where
  - $\mathscr{S}$ is σ-algebra generated by all left-continuous adapted processes;
- **integrable** if it is adapted to $\{\mathscr{F}_t\}_{t \geq 0}$ and $\mathbb{E}[X_t] < +\infty$ for all $t \in I$;
  - Similarly, we call X square integrable if it is adapted to $\{\mathscr{F}_t\}_{t \geq 0}$ and $\mathbb{E}[|X_t|^2] < +\infty$ for all $t \in I$.

- **increasing** if $X_0 = 0$ and $t \mapsto X_t$ is right-continuous and increasing (that is, non-decreasing) function a.s.;

  ↬ It follows that $X_t \geq 0$ a.s.

- **continuous** if $t \mapsto X_t$ is continuous function a.s.;

- **of bounded variation** on a finite interval $[a,b]$ if

$$\sup \left\{ \sum_{i=0}^{n-1} |X_{t_{i+1}} - X_{t_i}| : a = t_0 < t_1 < \ldots t_n = b, n \in \mathbb{N} \right\} < +\infty, \quad \text{a.s.}$$

We can consider two random processes equal in several ways. In one case we identify them if their trajectories agree up to a $\mathbb{P}$-null set. It is also possible to identify them on the basis of their properties.

Definition 1.0.1.

- A random process $Y = \{Y_t\}_{t \in I}$ is called a **modification** of a process $X = \{X_t\}_{t \in I}$ if, for all $t \geq 0$,

$$\mathbb{P}(\{\omega : X_t(\omega) = Y_t(\omega)\}) = 1.$$

- Two random processes $X = \{X_t\}_{t \in I}$ and $Y = \{Y_t\}_{t \in I}$ are equal, or we call Y **indistinguishable** from process X, if almost all of their trajectories are equal, that is, if

$$\mathbb{P}(\{\omega : t \in I \mapsto X_t(\omega) \text{ coincide to } t \in I \mapsto Y_t(\omega)\}) = 1.$$

- Two random processes $X = \{X_t\}_{t \in I}$ and $Y = \{Y_t\}_{t \in I}$ are **equal in distribution** if they have the same distribution, that is,

$$\mathbb{P}(X \in A) = \mathbb{P}(Y \in A), \quad A \in \mathscr{C},$$

  where $(C, \mathscr{C})$ is a measurable space defined by the functions $\omega \mapsto X_t(\omega)$ and use the filtration generated by $\omega(s)$, for $s \leq t$.

  ↬ *Remark:* if X and Y are continuous, and Y is a modification of X, then they are indistinguishable (because then they are determined by their values on a countable dense subset of $[0, +\infty\rangle$).

A standard situation in applications is to observe the value of the process at some random time point of interest in the future. Naturally, in order to analyse them within mathematical framework, we need to impose certain assumptions on those random times.

Definition 1.0.2. A random variable $\tau : \Omega \to [0, +\infty]$ is called a **stopping time** (with respect to $\{\mathscr{F}_t\}_{t\geq 0}$) if for all $t > 0$, $\{\tau < t\} \in \mathscr{F}_t$.

- *Remark:* since $\mathscr{F}$ is right-continuous and increasing family of σ-algebras, it is equivalent to require that $\{\tau \leq t\} \in \mathscr{F}_t$ for all $t > 0$.

In this work we discuss properties of some special types of stochastic processes. In order to get there, we need to develop the theory for certain broad classes of processes with some nice and useful properties and gradually extend it more and more. Therefore, we continue by naming some classes of stochastic processes that will be of interest to us.

Definition 1.0.3. A real stochastic process $\{X_t\}_{t\in I}$ is called a **martingale / submartingale / supermartingale** with respect to $\{\mathscr{F}_t\}_{t\geq 0}$ if

(i) X_t is integrable for all $t \in I$,

(ii) X is adapted to $\{\mathscr{F}_t\}_{t\geq 0}$,

(iii) $\mathbb{E}[X_t \mid \mathscr{F}_s] = X_s$ / $\mathbb{E}[X_t \mid \mathscr{F}_s] \leq X_s$ / $\mathbb{E}[X_t \mid \mathscr{F}_s] \geq X_s$ a.s., for all $s, t \in I$ such that $s < t$.

Next, we consider a version of such processes which are stopped at some random time point in the future.

Definition 1.0.4. A stochastic process $X = \{X_t\}_{t\geq 0}$ is called a **local martingale** with respect to $\{\mathscr{F}_t\}_{t\geq 0}$ if it is adapted to $\{\mathscr{F}_t\}_{t\geq 0}$ and if there exists a sequence of $(\mathscr{F}_t)$-stopping times $\{\tau_n\}_{n\geq 0}$ such that $\tau_n < +\infty$ for all $n \in \mathbb{N}$, $\tau_n \uparrow +\infty$ and stopped processes $X^{\tau_n} = \{X_{\tau_n \wedge t}\}_{t\geq 0}$ is a martingale for all $n \in \mathbb{N}$.

If, additinally, X^{τ_n} is a square integrable martingale for all $n \in \mathbb{N}$, then we call X a **locally square integrable martingale** with respect to $\{\mathscr{F}_t\}_{t\geq 0}$.

In many problems the observed motion posses the property that its future behaviour is determined by a present state of the system, not the past ones. For a system whose paths evolve in a random fashion, this means that the current state fully determines the distribution of the path from that point on, that is, it determines with what probability will the system occupy some state from all possible states at all future time points. That property is called the *Markov property* and is very well investigated in the literature since it entails many other useful features. We discuss it for both discrete and continuous system.

Definition 1.0.5. A discrete stochastic process $X = \{X_n\}_{n\geq 0}$ with values in a countable set $\mathbb{S}$ (which we call the **state space** of the chain) is called a **discrete-time Markov chain** if, for any $n \in \mathbb{N}$ and set of states $i, i_0, \dots, i_n \in \mathbb{S}$,

$$\mathbb{P}(X_{n+1} = i \mid X_0 = i_0, \dots X_n = i_n) = \mathbb{P}(X_{n+1} = i \mid X_n = i_n),$$

under assumption that both conditional probabilities are well defined, that is, $\mathbb{P}(X_0 = i_0, \dots X_n = i_n) > 0$.

A Markov chain is called

- **time-homogeneous** if the transition probability is independent of n, that is, for all $n \in \mathbb{N}$ and $i, j \in \mathbb{S}$

  $$\mathbb{P}(X_{n+1} = i \mid X_n = j) = \mathbb{P}(X_n = i \mid X_{n-1} = j);$$

  ↬ Such a Markov chain is characterized by **transition matrix** $P = (P_{i,j})_{i,j\in\mathbb{S}}$, where $P_{i,j} = \mathbb{P}(X_{n+1} = j \mid X_n = i)$, for all $i, j \in \mathbb{S}$ and $n \in \mathbb{N}$.

- **stationary** if for all $n, m \in \mathbb{N}$ and $i_0, \dots, i_n \in \mathbb{S}$

  $$\mathbb{P}(X_0 = i_0, \dots X_n = i_n) = \mathbb{P}(X_m = i_0, \dots X_{n+m} = i_n);$$

- **irreducible** if it is possible to go from every state to every state (possibly in several moves), that is, if for every $i, j \in S$ such that $\mathbb{P}(X_0 = i) > 0$ there exists $k \in \mathbb{N}$ such that $\mathbb{P}(X_k = j \mid X_0 = i) > 0$;

- **aperiodic** if all states $i \in \mathbb{S}$ have a period $d_i = 1$, where

  - $d_i := \gcd\{n \geq 1 : \mathbb{P}(X_n = i \mid X_0 = i) > 0\}$, with the additional definition that $d_i = +\infty$, if $\mathbb{P}(X_n = i \mid X_0 = i) = 0$ for all $n \geq 1$. Here, gcd stands for calculating the greatest common divisor.

- **ergodic** if there is some $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$ and $i, j \in \mathbb{S}$ for which $\mathbb{P}(X_0 = i) > 0$, $\mathbb{P}(X_n = j \mid X_0 = i) > 0$.

  ↬ *Remark:* it is equivalent to demand that X is irreducible and aperiodic Markov chain.

A Markov chain is, thus, a process whose evolution in the future depends only in its present state, and the switch to a following state in the next moment is determined by its transition matrix P. In other words, whatever happened in the past, stays in the past - it is no longer important. To consider such process in a continuous-time setting, we assume the change of the state is again controlled by a transition matrix, but we need to determine when it will happen. Since it can not happen in the next moment (because that means the jump happens instantly, all the time, which makes no sense), we let the chain rest in a present state for some time and then switch it. Of course, waiting times are random and, to support the idea of a Markov property, independent of each other.

Definition 1.0.6. A **continuous-time Markov chain** is a continuous-time process $\{X_t\}_{t\geq 0}$ on a state space $\mathbb{S}$ determined by two components:

(i) a transition matrix P (that corresponds to a discrete-time Markov chain - we call it a **jump chain** or **embedded Markov chain**),

(ii) **holding parameters** $q : \mathbb{S} \to \langle 0, +\infty \rangle$.

The waiting time in state $i \in \mathbb{S}$ is exponentially distributed with parameter $q(i)$. All waiting times are mutually independent. After the waiting time passes by, the jump to the next state is determined by the transition matrix P.

↬ So, we can think of a continuous-time Markov chain as a process that has a Markov chain embedded in it: we first create a set of jumps according to the discrete-time Markov chain (determined by a transition matrix P), and then we transfer it to a continuous-time setting by taking a set of independent exponentially distributed waiting times (in each state $i \in \mathbb{S}$ we let our process wait some time before jumping to the next state - the same state where the Markov chain would jump after being in the state i).

- ↬ To see that these two objects indeed define a stochastic process (therefore, they are also called the **local characteristics** of a Markov chain), we can look at the *integral equation*, namely, it holds that for all $i, j \in \mathbb{S}$ and $t > 0$

$$\mathbb{P}(X_t = j \mid X_0 = i) = \delta_{i,j} e^{-q(i)t} + \int_0^t q(i) e^{-q(i)s} \sum_{k \neq i} P_{i,k} \mathbb{P}(X_{t-s} = j \mid X_0 = k)\, ds.$$

A continuous-time Markov chain is called

- **regular** or **non-explosive** if its explosion time is infinite. More precisely, if we denote the *final state* (sometimes called a cemetary) by ∂ and a time it reaches that state (the so-called **explosion time**) by ζ, that is it holds that $X_t = \partial$ for $t \geq \zeta$, we have

$$\mathbb{P}(\zeta = +\infty \mid X_0 = i) = 1, \quad i \in \mathbb{S}.$$

  - ↬ A non-explosive Markov chain indeed satisfies the Markov property, defined below.

- **irreducible** if we can go with positive probability to any state, starting from any state, that is, for all $i, j \in \mathbb{S}$ such that $\mathbb{P}(X_0 = i) > 0$, there is $t_0 > 0$ such that

$$\mathbb{P}(X_{t_0} = j \mid X_0 = i) > 0.$$

  - ↬ Irreducibility implies that $\mathbb{P}(X_t = j \mid X_0 = i) > 0$ for all $t > 0$, so this property is also called **aperiodicity**.

- **ergodic** if it is irreducible (or aperiodic) and it possesses a stationary distribution, that is, a probability distribution π such that

$$\lim_{t \to +\infty} \mathbb{P}(X_t = j \mid X_0 = i) = \pi(j), \quad i, j \in \mathbb{S}.$$

  - ↬ Irreducible Markov chain has at most one stationary distribution.
  - ↬ A continuous-time Markov chain on a finite state space $\mathbb{S}$ has at least one stationary distribution, so, by the previous remark, it follows that it is ergodic if, and only if, it is irreducible.

Since continuous-time Markov chains will have a very important role later in the this work, we proceed by presenting some of their properties.

Definition 1.0.7. With the help of the local characteristics of a Markov chain, we define the **generator** matrix $\mathscr{Q} = (\mathrm{q}_{i,j})_{i,j\in\mathbb{S}}$ by

$$\mathrm{q}_{i,j} = \begin{cases} -q(i),\ j = i, \\ q(i)P_{i,j},\ j \neq i. \end{cases}$$

- *Remarks:* $\mathrm{q}_{i,i} < 0$ and $\mathrm{q}_{i,j} \geq 0$ for all $i, j \in \mathbb{S}$.

- $\sum_{j\in\mathbb{S}} \mathrm{q}_{i,j} = 0$ for all $i \in \mathbb{S}$.

- A generator matrix $\mathscr{Q}$ can be expressed in the form of two equations: if we denote by $P(t)$ the matrix $(\mathbb{P}(X_t = j \mid X_0 = i))_{i,j\in\mathbb{S}}$, then for all $i, j \in \mathbb{S}$ the function $t \mapsto \mathbb{P}(X_t = j \mid X_0 = i)$ is continuously differentiable and

  – *differential backward equation:*

$$P'(t) = \mathscr{Q}P(t), \quad t \geq 0.$$

  – *differential forward equation:*

$$P'(t) = P(t)\mathscr{Q}, \quad t \geq 0.$$

  A special case is for $t = 0$: we get that $P'(0) = Q$, that is,

$$\mathrm{q}_{i,j} = \lim_{h\to 0} \frac{\mathbb{P}(X_h = j \mid X_0 = i)}{h}, \quad i, j \in \mathbb{S}.$$

- The Markov chain can also be obtained from a generator matrix, because we also have to following equality: for $h > 0$ and $i, j \in \mathbb{S}$,

$$\mathbb{P}(X_h = j \mid X_0 = i) = \begin{cases} 1 + \mathrm{q}_{i,j}h + o(h) = 1 - q(i)h + o(h), & i = j, \\ \mathrm{q}_{i,j}h + o(h), & i \neq j, \end{cases}$$

  so $\mathrm{q}_{i,j}$ is also called the **transition rate** between states i and j from $\mathbb{S}$.

- An invariant measure π for X satisfies:

$$\pi Q = 0.$$

  If π is a probability measure, we call it an *invariant distribution*.

- If $\mu = (\mu_i)_{i \in \mathbb{S}}$ is an invariant measure for an embedded Markov chain, that is, $\mu = \mu P$, then $\pi = (\pi_i)_{i \in \mathbb{S}}$ defined by

$$\mu_i = \pi_i q(i), \quad i \in \mathbb{S},$$

is an invariant measure for continuous-time Markov chain X.

Further generalisation of the notion of a Markov chain leads us to consider a process that possesses a similar property of not depending on the past, but without assuming that it changes values only at some discrete set of time points - so, we allow it moves around all the time, and let the movement be determined only by a present state.

Definition 1.0.8. Let $(S, \mathscr{S})$ be a measurable space. An adapted stochastic process $X = \{X_t\}_{t \geq 0}$ with values in S is called the **continuous-time Markov process** (or a continuous-time process possessing **the Markov property**) with respect to $\{\mathscr{F}_t\}_{t \geq 0}$ if for all $A \in \mathscr{S}$ and $0 \leq s < t$

$$\mathbb{P}(X_t \in A \mid \mathscr{F}_s) = \mathbb{P}(X_t \in A \mid X_s).$$

- *Remarks:* alternatively, the Markov property can be stated in the following way: for all $f : S \to \mathbb{R}$, $f \in \mathscr{B}_b(\mathbb{R}^d)$, and all $0 \leq s \leq t$

$$\mathbb{E}[f(X_t) \mid \mathscr{F}_s] = \mathbb{E}[f(X_t) \mid \sigma(X_s)].$$

  - Here, $\mathscr{B}_b(\mathbb{R}^d)$ stands for all Borel-measurable and bounded functions.

If there is a function $p(s,x,t,A)$, where $x \in S$, $A \in \mathscr{S}$, $0 \leq s \leq t$, which satisfies

$$p(s, X_s, t, A) = \mathbb{P}(X_t \in A \mid X_s)$$

and possesses the following properties:

(i) $x \mapsto p(s,x,t,A)$ is a non-negative measurable function;

(ii) $A \mapsto p(s,x,t,A)$ is a positive probability measure on $\mathscr{S}$;

(iii) for all $x \in S$, $A \in \mathscr{S}$ and $0 \leq s < t < T$,

$$\int_S p(s,x,t,dy)\, p(t,y,T,A) = p(s,x,T,A);$$

we say that a Markov process X has a **transition probability function** or a **transition kernel** $p(s,x,t,A)$.

↬ *Remark:* transition probability plays the role of the transition matrix of a Markov chain in a continuous setting.

↬ Property (iii) is called a **Chapman-Kolmogorov equation**.

A Markov process is called **time-homogeneous** if it satisfies additional assumption that its transition probability depends on s and t only through their difference $t-s$. Namely, $p(s,x,t,A) = p(0,x,t-s,A)$ for all $0 \leq s \leq t$, $x \in S$ and $A \in \mathscr{S}$, so we can write p as a function of three variables: $p(t,x,dy) = p(0,x,t,dy)$.

As before, there is a version of these properties concerning a stopped process.

Definition 1.0.9. Take a stochastic process $X = \{X_t\}_{t\geq 0}$ with its natural filtration $\{\mathscr{F}_t\}_{t\geq 0}$ and a stopping time τ.

- $\mathscr{F}_\tau := \{A \in \mathscr{F} : \forall t \geq 0, \{\tau \leq t\} \cap A \in \mathscr{F}_t\}$
- X is called a **strong Markov process** (or said to posses a **strong Markov property**) if, conditional on $\{\tau < +\infty\}$ and X_τ, $X_{\tau+t}$ is independent of $\mathscr{F}_\tau$.

  ↬ This definition is naturally transferred to a discrete setting: a Markov chain $X = \{X_n\}_{n\geq 0}$ has a strong Markov property if for $i \in S$, conditional on $\{\tau < +\infty\}$ and $X_\tau = i$, $\{X_{\tau+n}\}_{n\geq 0}$ is a Markov chain starting from point i that is independent of $X_0, \dots, X_\tau$.

A very important example of a stochastic process, one of the most commonly used in applications and researched in the literature, is certainly Brownian motion. Unintentionally discovered by the Scottish botanist Robert Brown in 1827, while observing movement of pollen grains suspended in water under a microscope, it slowly started to play a key role in most of the mathematical models used to describe phenomena from nature, finance, physics and other science areas. First formal mathematical notions and analysis of Brownian motion date back to the end of 19th century and are due to the Danish astronomer Thorvald N. Thiele and the French mathematician Louis Bachelier who was modelling stock and option markets.

Definition 1.0.10. A stochastic process $\{X_t\}_{t\geq 0}$ is called a (one-dimensional) **Brownian motion** with the initial distribution μ if

(i) X_0 follows the probability distribution μ,

(ii) (*independent increments*) for every $n \in \mathbb{N}$ and all $0 = t_0 < t_1 < \ldots < t_n$ the variables $X_0, X_{t_1} - X_{t_0}, \ldots, X_{t_n} - X_{t_{n-1}}$ are mutually independent,

(iii) (*normal increments*) for all $0 \leq s < t$, $X_t - X_s \sim N(0, t-s)$,

(iv) (*continuity*) trajectories are a.s. continuous, i.e. $t \mapsto X_t(\omega)$ is continuous for almost every $\omega \in \Omega$.

If $\mu \equiv 0$, we call X a **standard Brownian motion**.

Furthermore, a d-dimensional process whose components are independent one-dimensional Brownian motions is called a d-**dimensional Brownian motion**.

A concept of such a stochastic motion can be generalized - those processes are called Lévy processes and we will mention the two most important representatives of that class. One of them is Brownian motion, and the second will be a Poisson process.

Definition 1.0.11. A stochastic process $\{X_t\}_{t\geq 0}$ that satisfies

(i) $X_0 = 0$,

(ii) (*independent increments*) for every $n \in \mathbb{N}$ and all $0 = t_0 < t_1 < \ldots < t_n$ the variables $X_0, X_{t_1} - X_{t_0}, \ldots, X_{t_n} - X_{t_{n-1}}$ are mutually independent,

(iii) (*stationary increments*) for all $0 < s < t$ it holds that $X_t - X_s \stackrel{d}{=} X_{t-s}$,

(iv) (*continuity in probability*) for all $\varepsilon > 0$ and $s \geq 0$

$$\lim_{t \to s} \mathbb{P}(|X_t - X_s| > \varepsilon) = 0,$$

is called a **Lévy process**.

↬ We can also assume that the trajectories $t \mapsto X_t$ are a.s. right continuous with existing left limits (a property that we call **càdlàg**), since we can always find a modification of process X with that property.

The next type of stochastic process that we will mention will be important to us when we will move from continuous processes to a more general class of processes that include jumps.

Definition 1.0.12. Let $\lambda > 0$.

- A Lévy process M for which $M_t \sim Poi(\lambda t)$ for $t > 0$ is called a **Poisson process with intensity** λ.
  - *Remarks:* variables M_t can be obtained in another way as well - take a sequence of random variables $S_k \sim Exp(\lambda)$ for all $k \in \mathbb{N}$. Then, M_t can be defined as $M_t = \max\{n \in \mathbb{N}_0 : \sum_{k=1}^{n} S_k \leq t\}$ for $t > 0$.
  - Poisson process is one type of a continuous-time Markov chain.
- The process $\tilde{M} = \{\tilde{M}_t\}_{t\geq 0}$ defined by $\tilde{M}_t := M_t - \lambda t$ for $t > 0$ is called a **compensated Poisson process**.
  - *Remark:* this process is a martingale.
- For a sequence of $\mathbb{R}^d$-valued random variables $\{Z_n\}_{n\geq 0}$ which are independent of M, the stochastic process $Y = \{Y_t\}_{t\geq 0}$ defined by $Y_t := \sum_{k=1}^{M_t} Z_k$ for all $t \geq 0$ is called a **compound Poisson process**.

Fix some set $E \subseteq \mathbb{R}^d$ and consider a measurable space $(E, \mathscr{B}(E))$ and a σ-finite measure μ on it with the property that

$$\mu(A) < +\infty, \quad \text{for all } A \in \mathscr{B}_b(E), \tag{1.0.1}$$

where $\mathscr{B}_b(E)$ stands for all bounded elements of $\mathscr{B}(E)$.

Definition 1.0.13.
- A function (random measure) $N : (\Omega, \mathscr{B}(E)) \to \mathbb{N}_0$ is called a **Poisson random measure** or a **Poisson point process** with intensity measure μ if
  - (i) $N(\omega, \cdot)$ is a $\mathbb{N}_0$-valued measure with the property (1.0.1) for a.e. $\omega \in \Omega$,
  - (ii) for all $B \in \mathscr{B}_b(\mathbb{R}^d)$, the random variable $N(B)$ follows a Poisson distribution with parameter $\mu(B)$,

(iii) for all $n \in \mathbb{N}$ and pairwise disjoint sets $B_1, \dots, B_n \in \mathscr{B}(\mathbb{R}^d)$, the random variables $N(B_1), \dots, N(B_n)$ are independent.

- A random measure $\tilde{N}$ is called a **compensated Poisson random measure** if

$$\tilde{N}(\omega, B) = N(\omega, B) - \mu(B), \quad B \in \mathscr{B}_b(\mathbb{R}^d), \omega \in \Omega,$$

and intensity measure μ is sometimes also called a **compensator** and denoted by $\hat{N}$.

↬ *Remarks:* note that this notation does not make sense if $\mu(B) = +\infty$ for some $B \in \mathscr{B}(\mathbb{R}^d)$ (in general, both $N(B)$ and $\mu(B)$ may be infinite). Therefore, we define this measure only for sets B such that $\mu(B) < +\infty$, and integrate with respect to $\tilde{N}$ only for those functions $f \in L^1(E, \mathscr{B}(E), \mu)$ for which the following notation makes sense:

$$\int f \, d\tilde{N} := \int f \, dN - \int f \, d\mu.$$

↬ If intensity measure μ is finite, then the compensated Poisson random measure $\tilde{N}$ is indeed a signed measure.

The reason why a Poisson random measure is also called Poisson point process is the fact that it can be represented as counting measure for sequence of random points. Namely, for a Poisson random measure N with intensity μ, we can find a sequence of i.i.d. random variables $\{Z_n\}_{n \geq 0}$ on E and a random variable $K \sim Poi(\mu(E))$ that is independent of $\{Z_n\}_{n \geq 0}$ such that

$$N(A) = \sum_{n=1}^{K} \mathbb{1}_A(Z_n) = \sum_{n=1}^{K} \delta_{Z_n}(A), \quad A \in \mathscr{B}(E),$$

where δ_{Z_n} stands for the Dirac measure at point Z_n. Furthermore, by considering E as a product space $\mathbb{R}^d \times [0, T]$ or $[0, +\infty\rangle \times \mathbb{R}^d$, we can observe the function N as a stochastic process (on the time interval $[0, T]$ or $[0, +\infty\rangle$, respectively).

The questions of integration with respect to a Poisson point process is now resolved since, for a.e. $\omega \in \Omega$, $N(\omega, \cdot)$ is a measure on E so we can use standard procedures for integration from measure theory. Namely, for any measurable function $f : E \to \mathbb{R}$ the integral $\int_E f(x)N(\omega,dx)$ is well-defined (for a.e. $\omega\in\Omega$). Moreover, the function $\int_E f dN$ defined by $\omega\mapsto\int_E f(x)N(\omega,dx)$ is a random variable with expectation

$$\mathbb{E}\left[\int_E f dN\right]=\int_E f d\mu.$$

As we will talk about jump processes, we will now construct them starting from a Poisson point process N viewed as a Poisson random measure N on $E=\mathbb{R}^d\backslash\{0\}\times[0,+\infty\rangle$ with intensity μ. As mentioned before, we can write N as

$$N=\sum_{n=1}^{+\infty}\delta_{(Y_n,T_n)},$$

where (Y_n,T_n) are mutually independent random variables. This will represent a jump of size Y_n at time T_n. For any $f\in L^1(E,\mathscr{B}(E),\mu)$ we obtain a stochastic process by integrating f up to time t. More precisely, for any $t\geq 0$, define

$$\begin{aligned}X_t&:=\int_{[0,+\infty\rangle}\int_{\mathbb{R}^d\backslash\{0\}} f(x,s)\mathbb{1}_{[0,t]}(s)N(dx,ds)\\&=\int_0^t\int_{\mathbb{R}^d\backslash\{0\}} f(x,s)N(dx,ds)\\&=\sum_{n\geq 1:T_n\leq t} f(Y_n,T_n).\end{aligned}$$

So, we see that this process indeed jumps at times T_n.

If we integrate with respect to a compensated Poisson random measure $\tilde{N}=N-\mu$, we obtain the following process $\tilde{X}$:

$$\begin{aligned}\tilde{X}_t&:=\int_{[0,+\infty\rangle}\int_{\mathbb{R}^d\backslash\{0\}} f(x,s)\mathbb{1}_{[0,t]}(s)\tilde{N}(dx,ds)\\&=\int_0^t\int_{\mathbb{R}^d\backslash\{0\}} f\,dN-\int_0^t\int_{\mathbb{R}^d\backslash\{0\}} f\,d\mu,\quad t\geq 0.\end{aligned}$$

Such a process is now a martingale with respect to the filtration $\tilde{\mathscr{F}}$, where

$$\tilde{\mathscr{F}}_t:=\sigma\{N(A):A\in\mathscr{E}\},\quad t>0.$$

It turns out that there is a connection between continuous-time Markov processes and integrals with respect to a Poisson random measure. Let $\Lambda=\{\Lambda_t\}_{t\geq 0}$ be the continuous-time Markov chain with values in $\mathbb{S}=\{1,2\ldots\}$ (that is, it has countable state space) and with the generator matrix $\mathscr{Q}=(\mathrm{q}_{i,j})_{\mathbb{N}\times\mathbb{N}}$. As discussed in [MY06, p. 46], any continuous-time Markov chain can be written in the form of the stochastic integral with respect to the

Poisson random measure (details can be found in [GS82, p. 226-227] and [GAM97]). Hence, the $\{\Lambda_t\}_{t\geq 0}$ has a following martingale form:

$$\Lambda_t = \int_0^t \int_{\mathbb{R}^d} h(\Lambda_{s-}, y) N(dx, ds), \qquad t \geq 0, \tag{1.0.2}$$

or, equivalently written, satisfies the following stochastic differential equation (we will formally introduce the notion of a SDE later):

$$d\Lambda_t = \int_{\mathbb{R}^d} h(\Lambda_{t-}, x) N(dx, dt), \qquad t \geq 0 \tag{1.0.3}$$

where N is a Poisson random measure with intensity $dt \times \lambda(dx)$ (in which λ is the Lebesque mesure on $\mathbb{R}^d$) and $h : \mathbb{S} \times \mathbb{R}^d \to \mathbb{R}$ is defined by:

$$h(i,x) := \begin{cases} j-i, & x \in \Delta_{ij} \text{ for some } j \in \mathbb{S}, \\ 0, & \text{otherwise,} \end{cases}$$

for all $i \in \mathbb{S}$ and $x \in \mathbb{R}^d$. Here $\{\Delta_{i,j} : i, j \in S\}$ is a family of disjoint intervals defined on a non-negative part of the real line as follows

$$\begin{aligned}
\Delta_{12} &:= [0, \mathrm{q}_{12}\rangle\,, \\
\Delta_{13} &:= [\mathrm{q}_{12}, \mathrm{q}_{12} + \mathrm{q}_{13}, \rangle \\
&\ \ \vdots \\
\Delta_{21} &:= [\mathrm{q}(1), \mathrm{q}(1) + \mathrm{q}_{21}\rangle\,, \\
\Delta_{23} &:= [\mathrm{q}(1) + \mathrm{q}_{21}, \mathrm{q}(1) + \mathrm{q}_{21} + \mathrm{q}_{23}\rangle\,, \\
&\ \ \vdots
\end{aligned}$$

where $\mathrm{q}(i) = \sum_{j \in \mathbb{S} \setminus \{i\}} \mathrm{q}_{ij} \in \langle 0, +\infty\rangle$. If $\mathrm{q}_{ij} = 0$ for some $i \neq j \in \mathbb{S}$, we define $\Delta_{ij} = \emptyset$.

As mentioned, jump processes will be important to us as we will develop the theory of ergodicity for them. So we proceed by formally defining jumps and their features.

Recall that a càdlàg process X has right-continuous sample paths with left limits, namely, for any $t > 0$ the following limit exists:

$$X_{t-} = \lim_{h \downarrow 0} X_{t-h}.$$

Definition 1.0.14. • We say a càdlàg and adapted process $X = \{X_t\}_{t \geq 0}$ **jumps** (or has a discontinuity) at time $t > 0$ if

$$\Delta X_t := X_t - X_{t-} \neq 0.$$

↬ *Remarks:* X can admit only countably many jumps, so **jump times** make a countable set and we denote them by $(T_n)_{n\geq 1}$, that is, $\{t \geq 0 : \Delta X_t \neq 0\} = T_n : n \geq 1$.

↬ T_n for $n \geq 1$ are stopping times.

↬ **Jump sizes** are denoted by $(Y_n)_{n\geq 1}$, that is,

$$Y_n := \Delta X_{T_n} = X_{T_n} - X_{T_{n-}} \in \mathbb{R}^d \backslash \{0\}, \quad n \geq 1.$$

Y_n is $\mathscr{F}_{T_n}$-measurable random variable for all $n \geq 1$.

- The sequence of jump times and jump sizes defines a **jump measure** J_X of X by

$$J_X(\omega, A \times B) := \sum_{n\geq 1} \delta_{(T_n(\omega), Y_n(\omega))}(A \times B) = \text{card}\{(t, \Delta X_t(\omega)) \in A \times B\}, \quad (1.0.4)$$

for $\omega \in \Omega, A \times B \in \mathscr{B}([0, +\infty\rangle \times \mathbb{R}^d)$.

↬ *Remarks:* a jump measure is a random measure on $[0, +\infty\rangle \times \mathbb{R}^d$ and for any $t \geq 0$ and $B \in \mathscr{B}(\mathbb{R}^d)$ the random variable $J_X([0,t] \times B)$ counts the number of jumps of X until time t such that their size is in B.

Since Lévy processes will be of our special interest, we investigate them further.

Definition 1.0.15. Let $X = \{X_t\}_{t\geq 0}$ be a Lévy process. The **Lévy measure** or a **Lévy kernel** of X is a measure ν on $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$ defined by:

$$\nu(\{0\}) = 0$$
$$\nu(A) = \mathbb{E}[J_X([0,1] \times B)] = \mathbb{E}[\text{card}\{t \in [0,1] : \Delta X_t \neq 0, \Delta X_t \in B\}], \quad B \in \mathscr{B}(\mathbb{R}^d).$$

↬ *Remarks:* if $B \in \mathscr{B}(\mathbb{R}^d)$ is a compact set such that $0 \notin B$, then $\nu(B)$ is finite, since there is $\varepsilon > 0$ such that B is contained in the complement of the ball around origin of radius ε B_ε^c, and the monotonicity of measure ν then implies

$$\nu(B) \leq \nu(B_\varepsilon^c) = \mathbb{E}[J_X([0,1] \times B_\varepsilon^c)] < +\infty,$$

beacuse the right hand side is just a number of jumps in one time unit that are larger in size than ε. As X is càdlàg, we have only finitely many such jumps. This means that ν is a Radon measure on $(\mathbb{R}^d \backslash \{0\}, \mathscr{B}(\mathbb{R}^d \backslash \{0\}))$.

- It holds that

$$\int_{\mathbb{R}^d} (1 \wedge |x|^2)\nu(dx) < +\infty \iff \int_{\mathbb{R}^d} \frac{|x|^2}{1+|x|^2}\nu(dx) < +\infty.$$

  This condition implies that X might have infinitely many small jumps

- ν does not need to be a finite measure on $\mathbb{R}^d$: there might be infinitely many small jumps on time interval $[0,1]$, resulting in a measure ν having a singularity at the origin.

- The jump measure J_X is a Poisson random measure on $[0,+\infty\rangle \times \mathbb{R}^d \backslash \{0\}$ with intensity $dt\,\nu(dx)$.

- For some $\varepsilon > 0$, the number of jumps until some time $t \geq 0$ is finite, so we can define for $t \geq 0$

$$N_t := \operatorname{card}\{0 \leq s \leq t : |\Delta X_s| > \varepsilon\}.$$

  Then, $N = \{N_t\}_{t\geq 0}$ is a Poisson process. Particularly, $N_1 = J_X([0,1] \times B_\varepsilon^c)$ follows a Poisson distribution with a finite parameter.

We are now ready to state the most important result regarding Lévy processes: a theorem that says that any Lévy process is actually a linear combination of a Brownian motion with drift and a countable sum of independent compound Poisson processes. A reverse of the result states that such a decomposition is unique, hence, a Lévy process can be characterized by a triplet of coefficients.

Theorem 1.0.16 (Lévy-Itô decomposition). Let $X = \{X_t\}_{t\geq 0}$ be a Lévy process with Lévy measure ν. Then there exists a vector $\beta \in \mathbb{R}^d$, a positive, semi-definite, symmetric matrix γ with all positive elements (i.e. $\gamma > 0$) and a standard Brownian motion $B = \{B_t\}_{t\geq 0}$ in $\mathbb{R}^d$ such that, for all $t \geq 0$, X_t can be decomposed in the following way:

$$X_t = \beta t + \gamma^{1/2} B_t + X_t^l + \lim_{\varepsilon \downarrow 0} \tilde{X}_t^\varepsilon, \tag{1.0.5}$$

where

$$X_t^l := \int_0^t \int_{|x|>1} x J_X(ds,dx),$$
$$\tilde{X}_t^\varepsilon := \int_0^t \int_{\varepsilon<|x|\leq 1} x \tilde{J}_X(ds,dx) = \int_0^t \int_{\varepsilon<|x|\leq 1} x(J_X(ds,dx) - ds\nu(dx)).$$

All terms in the decomposition are mutually independent. Convergence in the last term is a.s. and uniform on any finite $[0,T]$. The triplet (β,γ,ν) is uniquely determined by X and is called a **Lévy triplet** or a **characteristic triplet** .

- *Remarks:* if X is a continuous Lévy process, then its Lévy-Itô decomposition suggests it can be decomposed as a sum of a drift and a Brownian motion - we have first two terms of the decomposition in (1.0.5), since the last two terms encode jumps of the process.

- As mentioned in the previous remark, the process N_t that counts the jumps until time t, for all $t \geq 0$, is a Poisson process. Next, one can show that the sizes of jumps are i.i.d. variables (see [CT04, Proposition 3.3]). This, together with (1.0.4), implies that $X^l = (X_t^l)_{t\geq 0}$ is a compound Poisson process, and it follows from (1.0.4) that it is actually summing only large jumps (more precisely, jumps larger than 1 in size). So, X^l is a compound Poisson process with a.s. finitely many jumps. This fact will then imply that J_X is a Poisson random measure, which we stated above.

- Similary, X^ε is a compound Poisson process, too, taking into account small jumps of sizes between ε and 1. Since X might have infinitely many small jumps, we cannot simply integrate with respect to J_X for such jumps. Hence, the limit $\lim_{\varepsilon\downarrow 0} \tilde{X}_t^\varepsilon$ might not exist too. So, in this case, we consider the compensated compound Poisson process and integrate

- The threshold 1 that separates small and large jumps is arbitrary - any other threshold could be used.

- Observe that the decomposition implies that X_t can be written as a sum of a Brownian motion with drift and a, possibly infinite, sum of independent compound Poisson processes, since we can write the limit as a countable sum of $X^{\varepsilon_{n+1}} - X^{\varepsilon_n}$ (which are all independent compound Poisson processes), for some sequence $(\varepsilon_n)_{n\geq 1}$ such that $\varepsilon_n \downarrow 0$ as $n \uparrow +\infty$. Hence, we will be able to derive a diffusion process by using some Lévy triplet.

- There is 1-1 correspondence between Lévy triplets and Lévy processes, as the reverse of the theorem holds true. Namely, if we are given a vector $\beta \in \mathbb{R}^d$, a positive, semi-definite symmetric matrix $\gamma \in \mathbb{R}^d \times \mathbb{R}^d$ such that $\gamma > 0$, and a Radon measure ν on $\mathbb{R}^d \backslash \{0\}$ such that $\nu(\{0\}) = 0$ and

$$\int_{\mathbb{R}^d} (1 \wedge |x|^2)\nu(dx) < +\infty,$$

then we can find a Lévy process X with the characteristic triplet (β, γ, ν) that satisfies (1.0.5).

- If $\{X_t\}_{t\geq 0}$ is a $\mathbb{R}^n$-valued Lévy process with Lévy triplet (β, γ, ν), the linear transformation for some matrix $\sigma \in \mathbb{R}^d \times \mathbb{R}^n$, that is, the process $(\sigma X_t)_{t\geq 0}$, is again a Lévy process, and the Lévy triplet $(\beta_\sigma, \gamma_\sigma, \nu_\sigma)$ of it is given by (see e.g. [Sat13, Proposition 11.10])

$$(\beta_\sigma, \gamma_\sigma, \nu_\sigma) = (\sigma\beta + \int_{\mathbb{R}^n} \sigma y(\mathbb{1}_{B_1(0)}(\sigma y) - \mathbb{1}_{B_1'(0)}(y))\nu(dy), \sigma\gamma\sigma^T, \nu_\sigma),$$

where $\nu_\sigma(B) = \nu(\{x \in \mathbb{R}^n : \sigma x \in B\})$ for $B \in \mathscr{B}(\mathbb{R}^d)$, $B_1'(0) = \{y \in \mathbb{R}^n : |y| \leq 1\}$ and $B_1(0) = \{x \in \mathbb{R}^d : |x| \leq 1\}$.

1.1. Itô's Calculus

The problem that motivated the construction of the stochastic integral was the question of calculating integral $\int_0^t dx_s$ in a case that x_t is nowhere differentiable curve (trajectory). Since this situation did not fit into the standard framework for integration, Itô introducted a concept of integration that worked for stochastic processes that we call *infinitely divisible processes*. Later on, the theory was extended to a class of stochastic processes called *semimartingales*.

We observe stochastic process on time interval $[0, +\infty\rangle$ and if X is an martingale, we assume that $t \mapsto X_t$ is right continuous a.s. (since we can always find a modification of X such that this holds).

Definition 1.1.1. • Define $\mathscr{L}_2$ to be the space of all real measurable processes $X = \{X_t\}_{t\geq 0}$ which are adapted to $\{\mathscr{F}_t\}_{t\geq 0}$ and such that

$$\|X\|_{2,T}^2 := \mathbb{E}\left[\int_0^T X_s^2 ds\right] < +\infty, \quad \forall T > 0.$$

- Define also $\mathscr{L}_0$ to be a subspace of $\mathscr{L}_2$ of such processes $\{X_t\}_{t\geq 0}$ for which there is a sequence of real numbers $0 = t_0 < t_1 < t_2 < \ldots \to +\infty$ and a sequence of random variables $\{Z_n\}_{n\geq 0}$ such that Z_n is $\mathscr{F}_{t_n}$-measurable for all $n \in \mathbb{N}$, $\sup_{n\in\mathbb{N}} \|Z_n\|_\infty < +\infty$ and

$$X_t(\omega) = \begin{cases} Z_0(\omega), & \text{if } t = 0, \\ Z_n(\omega), & \text{if } t \in \langle t_n, t_{n+1}], \text{ for } n = 0, 1, \ldots \end{cases}$$

↬ *Remark:* as a result, obviously, $X_t(\omega)$ can be expressed in the following form:

$$X_t(\omega) = Z_0(\omega)\mathbb{1}_{\{t=0\}\}}(t) + \sum_{n=0}^{+\infty} Z_n(\omega)\mathbb{1}_{\langle t_n, t_{n+1}]}(t), \quad t \geq 0.$$

For $X \in \mathscr{L}_2$, let

$$\|X\|_{\mathscr{L}_2} := \sum_{n=1}^{+\infty} 2^{-n}(\|X\|_{2,n} \wedge 1).$$

Then, together with a definition that two processes X and $Y \in \mathscr{L}_2$ are equal if

$$\|X - Y\|_{2,T} = 0, \quad \text{for all } T > 0,$$

the function $(X,Y) \mapsto \|X - Y\|_{\mathscr{L}_2}$ becomes a metric on space $\mathscr{L}_2$. It holds that for any $X \in \mathscr{L}_2$ there is another $X' \in \mathscr{L}_2$ which is predictable and which is identical to X. Furthermore, $\mathscr{L}_2$ is a complete metric space and $\mathscr{L}_0$ is dense in $\mathscr{L}_2$ in this metric.

Definition 1.1.2.

- $\mathscr{M}_2 := \{X = \{X_t\}_{t\geq 0}$ is a square integrable martingale with respect to $\{\mathscr{F}_t\}_{t\geq 0}$ and $X_0 = 0$ $a.s.\}$

- $\mathscr{M}_2^c := \{X \in \mathscr{M}_2 : t \mapsto X_t$ is continuous $a.s.\}$

We consider two processes X and Y from $\mathscr{M}_2$ to be equal if

$$t \mapsto X_t \text{ and } t \mapsto Y_t \text{ are equal } a.s.$$

Furthermore, for $X \in \mathscr{M}_2$, set

$$\|X\|_T := \mathbb{E}[X_T^2]^{\frac{1}{2}}, \quad \text{for all } T > 0,$$

and

$$\|X\|_{\mathscr{M}_2} := \sum_{n=1}^{+\infty} 2^{-n}(\|X\|_n \wedge 1).$$

Now the function $(X,Y) \mapsto \|X-Y\|_{\mathscr{M}_2}$ makes $\mathscr{M}_2$ a complete metric space. Moreover, $\mathscr{M}_2^c$ is a closed subspace of $\mathscr{M}_2$.

The first type of stochastic integration that we discuss will be integration with respect to a Brownian motion $\{B_t\}_{t\geq 0}$. Fix a standard Brownian motion $\{B_t\}_{t\geq 0}$ on our probability space $(\Omega, \mathscr{F}, \mathbb{P})$.

Definition 1.1.3. Take stochastic process $X \in \mathscr{L}_0$. Then, we can write $X_t(\omega) = Z_0(\omega) \cdot \cdot \mathbb{1}_{\{t=0\}\}}(t) + \sum_{n=0}^{+\infty} Z_n(\omega)\mathbb{1}_{\langle t_n, t_{n+1}]}(t), t \geq 0$. We define a **stochastic integral of** X **with respect to Brownian motion** B to be

$$\int_0^t X_s(\omega)dB_s(\omega) := \sum_{i=0}^{n-1} Z_i(\omega)(B_{t_{i+1}}(\omega) - B_{t_i}(\omega)) + Z_n(\omega)(B_t(\omega) - B_{t_n}(\omega)),$$

if $t_n \leq t \leq t_{n+1}, n \in \mathbb{N}$, and we use the notation $\int_0^t X_s dB_s = \int_0^t X_s(\omega)dB_s(\omega)$.

- *Remark:* note that the integral can be expressed in the following form as well:

$$\int_0^t X_s dB_s = \sum_{i=0}^{+\infty} Z_i(\omega)(B_{t\wedge t_{i+1}}(\omega) - B_{t\wedge t_i}(\omega)).$$

- Furthermore, it holds that the set of integrals $I(X)_t := \left(\int_0^t X_s dB_s\right)_{t\geq 0}$ viewed as a stochastic process (denote it by $I(X)$) is a continuous $\mathscr{F}_t$-martingale, that is, it is an element of $\mathscr{M}_2^c$, and is independent of the choice of $\{t_n\}_{n\geq 0}$. It also holds that, for all $T > 0$,

$$\|I(X)\|_T = \|X\|_{2,T} \quad \text{and} \quad \|I(X)\|_{\mathscr{M}_2} = \|X\|_{\mathscr{L}_2}.$$

After defining the integral for processes in $\mathscr{L}_0$, we extend the class of integrands first to space $\mathscr{L}_2$, and then to all adapted processes.

First, since $\mathscr{L}_0$ is dense in $\mathscr{L}_2$, for a process $X \in \mathscr{L}_2$ we can take an approximating sequence $\{X_n\}_{n\geq 0}$ from space $\mathscr{L}_0$, that is, the sequence satisfies

$$\|X - X_n\|_{\mathscr{L}_2} \to 0, \quad n \to +\infty.$$

Note that $\|I(X_n) - I(X_m)\|_{\mathscr{M}_2} = \|X_n - X_m\|_{\mathscr{L}_2}$, so $\{X_n\}_{n\geq 0}$ is a Cauchy sequence in $\mathscr{L}_2$, which implies that $(I(X_n))_{n\geq 0}$ is a Cauchy sequence in $\mathscr{M}_2$. Since $I(X_n) \in \mathscr{M}_2^c$, and $\mathscr{M}_2^c$ is a closed subspace of $\mathscr{M}_2$, it follows that $I(X_n)$ converges to a unique element $Z = \{Z_t\}_{t\geq 0}$ from the space $\mathscr{M}_2^c$. Note that Z does not depend on the choice of $\{X_n\}_{n\geq 0}$. We denote Z by $I(X)$ since it is clearly determined uniquely from the process X.

Definition 1.1.4. Take a stochastic process $X \in \mathscr{L}_2$. The process $I(X) \in \mathscr{M}_2^c$ defined above is called the **stochastic integral of X with respect to the Brownian motion B**. We use the following notation for $I(X)_t$: $\int_0^t X_s dB_s$ or $\int_0^t X_s(\omega) dB_s(\omega)$.

↬ *Remark*: the integral satisfies all properties of a linear functional.

We continue to expand the class of integrands to the following set.

Definition 1.1.5.
- $\mathscr{L}_2^{loc} = \{X = \{X_t\}_{t\geq 0} : X$ is a real measurable process adapted to $\{\mathscr{F}_t\}_{t\geq 0}$ such that for all $T > 0$, $\int_0^T X_t^2(\omega)dt < +\infty \; a.s\}$.

  ↬ *Remark*: similarly as before, we will need to identify two processes X and Y from $\mathscr{L}_2^{loc}$. This time, we write $X = Y$ if for all $T > 0$ we have that

  $$\int_0^T |X_t(\omega) - Y_t(\omega)|^2 dt = 0 \; a.s.$$

  ↬ Furthermore, for any $X \in \mathscr{L}_2^{loc}$ we can find $X' \in \mathscr{L}_2^{loc}$ which is also predictable and such that $X = X'$.

- $\mathscr{M}_2^{loc} := \{X = \{X_t\}_{t\geq 0}$ is a locally square integrable martingale with respect to $\{\mathscr{F}_t\}_{t\geq 0}$ and $X_0 = 0 \; a.s.\}$

- $\mathscr{M}_2^{c,loc} := \{X \in \mathscr{M}_2^{loc} : t \mapsto X_t$ is continuous $a.s.\}$

Fix any $X \in \mathscr{L}_2^{loc}$. Now, define the sequence of stopping times

$$\tau_n(\omega) := \inf\{t \geq 0 : \int_0^t X_s^2(\omega) ds \geq n\} \wedge n, \quad n \in \mathbb{N}.$$

Then $\{\tau_n\}_{n\geq 0}$ obviously satisfies properties that $\tau_n < +\infty$ and $\tau_n \uparrow +\infty$, as $n \to \infty$. Let now $X_s^{(n)}(\omega) = X_s(\omega)\mathbb{1}_{\{s \leq \tau_n(\omega)\}}$. It holds that

$$\int_0^{+\infty} \left(X_s^{(n)}(\omega)\right)^2 ds = \int_0^{\tau_n} \left(X_s^{(n)}(\omega)\right)^2 ds \leq n,$$

so it follows that $X_n \in \mathscr{L}_2$, for all $\in \mathbb{N}$, and we have defined integrals of processes in $\mathscr{L}_2$. So, we have $I(X^{(n)}) \in \mathscr{M}_2^c$. Furthermore, since $I(X^{(n)})_{t\wedge\tau_m} = I(X^{(m)})_t$, for all $m < n$, it follows that the process $I(X)$ defined by $I(X)_t := I(X^{(n)})_t$, for $t \leq \tau_n$, is well-defined, it is a continuous process and belongs to the space $\mathscr{M}_2^{c,loc}$.

Definition 1.1.6. For a stochastic process $X \in \mathscr{L}_2^{loc}$, the process $I(X) \in \mathscr{M}_2^{c,loc}$ defined above is called the **stochastic integral of X with respect to the Brownian motion B** and $I(X)_t$ is denoted by $\int_0^t X_s dB_s$ or $\int_0^t X_s(\omega) dB_s(\omega)$, as before.

We have successfully defined integration with respect to a Brownian motion for a wide range of integrands. Now, further generalization leads us to consider a more general set with respect to which we can integrate. Namely, we want to integrate with respect to martingales and semi-martingales.

First, fix a martingale $M \in \mathscr{M}_2$. Then, there exists (by the Doob-Meyer's decomposition) a unique (up to a modification) integrable increasing process $A = \{A_t\}_{t\geq 0}$ such that $(M_t^2 - A_t)_{t\geq 0}$ is a martingale.

Similarly, if M and N are in $\mathscr{M}_2$, then there exists a process $\{A'_t\}_{t\geq 0}$ which can be expressed as a difference of two integrable increasing processes such that $(M_t N_t - A'_t)_{t\geq 0}$ is a martingale.

Definition 1.1.7.

- Denote a unique integrable increasing process A defined above by $\langle M \rangle = \{\langle M \rangle_t\}_{t\geq 0}$.
- A process A' defined above is called **quadratic variational process** corresponding to M and N and denoted by $\langle M, N \rangle = \{\langle M, N \rangle_t\}_{t\geq 0}$.
  - *Remarks:* $\langle M, N \rangle$ is the unique adapted and continuous process of bounded variation on any finite interval such that $\langle M, N \rangle_0 = 0$ and $MN - \langle M, N \rangle$ is a martingale.
  - It holds that $\langle M \rangle = \langle M, M \rangle$.

Definition of integral with respect to a martingale follows the same procedure as in the case of a Brownian motion. We just need to adapt the definition of spaces and corresponding integrals.

Definition 1.1.8. Let $\mathscr{L}_2(M)$ to be the space of all real predictable processes $X = \{X_t\}_{t\geq 0}$ such that for all

$$(\|X\|_{2,T}^M)^2 := \mathbb{E}\left[\int_0^T X_s^2 d\langle M \rangle_s\right] < +\infty, \quad \forall T > 0.$$

- *Remark:* if M is a Brownian motion, then $\mathscr{L}_2(M) = \mathscr{L}_2$ defined before.

- since every process from the space $\mathscr{L}_0$ is left-continuous, therefore also predictable, it holds that $\mathscr{L}_0 \subseteq \mathscr{L}_2(M)$.

For $X \in \mathscr{L}_2(M)$, let

$$\|X\|_{\mathscr{L}_2(M)} := \sum_{n=1}^{+\infty} 2^{-n}(\|X\|_{2,n}^M \wedge 1).$$

As before, we consider two processes X and Y from $\mathscr{L}_2(M)$ to be equal, and write $X = Y$, if their difference is 0 in the corresponding norm, that is, if

$$\|X - Y\|_{2,T}^M = 0, \quad \text{for all } T > 0.$$

Now, we see that the function $(X,Y) \mapsto \|X - Y\|_{\mathscr{L}_2(M)}$ makes the space $\mathscr{L}_2(M)$ a complete metric space and it holds that $\mathscr{L}_0$ is dense in $\mathscr{L}_2(M)$ with respect to this metric function.

First we define integral for integrands in $X \in \mathscr{L}_0$. Since, for $t \geq 0$, X can be written in the following form: $X_t(\omega) = Z_0(\omega) \cdot \mathbb{1}_{\{t=0\}\}}(t) + \sum_{n=0}^{+\infty} Z_n(\omega) \mathbb{1}_{\langle t_n, t_{n+1}]}(t)$, we define

$$I^M(X)_t := \sum_{i=0}^{n-1} Z_i(\omega)(M_{t_{i+1}}(\omega) - M_{t_i}(\omega)) + Z_n(\omega)(M_t(\omega) - M_{t_n}(\omega)),$$

for $t_n \leq t \leq t_{n+1}, n \in \mathbb{N}$. As before, the process $I^M(X) := (I^M(X)_t)$ is an element of $\mathscr{M}_2$ and it holds that $\|I^M(X)\|_{\mathscr{M}_2} = \|X\|_{\mathscr{L}_2(M)}$. By using this equality, we extend the definition of the process $I^M(X)$ for all $X \in \mathscr{L}_2(M)$.

Definition 1.1.9. For $M \in \mathscr{M}_2$ and a stochastic process $X \in \mathscr{L}_2(M)$, the process $I^M(X)$ is called the **stochastic integral of** X **with respect to the martingale** M and $I^M(X)_t$ is denoted by $\int_0^t X_s dM_s$ or $\int_0^t X_s(\omega) dM_s(\omega)$.

- *Remark:* if M is a Brownian motion, $I(X)$ defined above coincides with $I^M(X)$.
- If $M \in \mathscr{M}_2^c$, then $I^M(X) \in \mathscr{M}_2^c$.

We further extend our definition of a stochastic integral to the case of local martingales. The procedure starts by defining quadratic variation for such processes.

Take two processes $M, N \in \mathscr{M}_2^{loc}$. Then we can take a sequence of finite stopping times $\{\tau_n\}_{n\geq 0}$ such that $\tau_n \uparrow +\infty$ as $n \to +\infty$ a.s. and stopped processes M^{τ_n} and N^{τ_n} are elements of $\mathscr{M}_2$. Since the quadratic variational process is unique, it follows that, for $m < n$,

$$\langle M^{\tau_n}, N^{\tau_n} \rangle_{t\wedge\tau_m} = \langle M^{\tau_m}, N^{\tau_m} \rangle.$$

Hence, there exists a unique predictable process $\langle M,N\rangle$ such that $\langle M,N\rangle_{t\wedge\tau_n} = \langle M^{\tau_n},N^{\tau_n}\rangle_t$ for all $n\in\mathbb{N}$ and $t\geq 0$. If $M=N$, we write $\langle M,M\rangle = \langle M\rangle$.

Definition 1.1.10. Let $M\in\mathscr{M}_2^{loc}$. $\mathscr{L}_2^{loc}(M) = \{X=\{X_t\}_{t\geq 0} : X$ is a real predictable process such that there exists a sequence of stopping times $\{\tau_n\}_{n\geq 0}$ such that $\tau_n\uparrow+\infty$ as $n\to+\infty$ a.s. and $\mathbb{E}\left[\int_0^{T\wedge\tau_n} X_t^2(\omega)d\langle M\rangle_t\right] < +\infty$ for all $T>0$ and $n\in\mathbb{N}\}$.

- *Remark*: $\mathscr{L}_2^{loc}(M) = \mathscr{L}_2^{loc}$ if M is a Brownian motion, since the condition on the expectation is equivalent to that $\int_0^T X_t^2(\omega)d\langle M\rangle_t < +\infty$ for all $T>0$ a.s.

We now proceed to define the integral in this case: fix $M\in\mathscr{M}_2^{loc}$ and $X\in\mathscr{L}_2^{loc}(M)$. Then we can find a sequence of stopping times $\{\tau_n\}_{n\geq 0}$ such that $\tau_n\uparrow+\infty$ as $n\to+\infty$ a.s., the stopped process M^{τ_n} is in $\mathscr{M}_2$ and $\mathbb{E}\left[\int_0^{T\wedge\tau_n} X_t^2(\omega)d\langle M\rangle_t\right] < +\infty$ for all $T>0$ and $n\in\mathbb{N}$. Define $X_t^{(n)}(\omega) := X_t(\omega)\mathbb{1}_{\{t\leq\tau_n(\omega)\}}$ and $M^{(n)} := M^{\tau_n}$. It holds that $M^{(n)}\in\mathscr{M}_2$ and $X^{(n)} = (X_t^{(n)})_{t\geq 0}\in\mathscr{L}_2(M^{(n)})$, so by Definition (1.1.9) we can find $I^{M^{(n)}}(X^{(n)})$ and it is easy to see that for $m<n$

$$I^{M^{(n)}}(X^{(n)})_{t\wedge\tau_m} = I^{M^{(m)}}(X^{(m)})_t.$$

Therefore, there exists a unique process $I^M(X) = \{I^M(X)_t\}_{t\geq 0}$ for which it holds that $I^M(X)_{t\wedge\tau_n} = I^{M^{(n)}}(X^{(n)})_t$ for all $n\in\mathbb{N}$ and $t\geq 0$. Obviously, $I^M(X)\in\mathscr{M}_2^{loc}$.

Definition 1.1.11. For $M\in\mathscr{M}_2^{loc}$ and $X\in\mathscr{L}_2^{loc}(M)$, the process $I^M(X)$ defined above is called the **stochastic integral of** X **with respect to the local martingale** M. As before, $I^M(X)_t$ is denoted by $\int_0^t X_s dM_s$ or $\int_0^t X_s(\omega)dM_s(\omega)$.

1.2. SEMI-MARTINGALES

As was discussed before, our goal is to create a mathematical framework which will allow us to model some physical motions and phenomena in nature and real life. Here, we are interested in situations where motions can not be described by a deterministic model (some version of a differential equation) since certain level of randomness is incorporated in their movement. One can think of modeling a random process as creating differential equation in a random environment. Informally, that type of model can be written as an equation of the following form:

$$dX_t = b(X_t)dt + \sigma(X_t)dB_t, \quad t \geq 0,$$

that is, the motion is a sum of a *drift* (trend or a mean motion) and a *diffusion* (fluctuation) away from that trend. Randomness is incorporated in the movement of such a process via a Brownian motion B, so such a process will be continuous. If we want to model a motion that is not continuous by its nature, we would take a model that is moved by a random process with jumps, for example, a Lévy process:

$$dX_t = b(X_t)dt + \sigma(X_t)dL_t, \quad t \geq 0,$$

where L is an underlying Lévy process. We then call such an SDE a **Lévy-driven** SDE.

To bring this notation to life, we introduce processes that will have such properties.

Definition 1.2.1. A real-valued stochastic process $\{X_t\}_{t\geq 0}$ is called a **on-dimensional semi-martingale** if X_t can be expressed in the following form: for $t \geq 0$

$$X_t = X_0 + A_t + M_t + \int_0^{t+}\int_{\mathbb{R}^d} f(x,s,\cdot)N(dx,ds) + \int_0^{t+}\int_{\mathbb{R}^d} g(x,s,\cdot)\tilde{N}(dx,ds), \quad (1.2.1)$$

where

- X_0 is an $\mathscr{F}_0$-measurable random variable,
- $M \in \mathscr{M}_2^{c,loc}$,
- the process $A = \{A_t\}_{t\geq 0}$ is a continuous process adapted to $\mathscr{F}$ of bounded variation (that is, its trajectories $t \mapsto A_t$ are of bounded variation on each finite interval) such that $A_0 = 0$,

- N is a Poisson random measure with intensity μ and $\tilde{N}$ is its compensated version,

- $f \in L^1(E, \mathscr{B}(E), \mu)$ for $E = \mathbb{R}^d \times [0,t\rangle$ and $g \in L^{2,loc}_{\mu}$, where

  - $L^{2,loc}_{\mu}$ is the set of all (random) functions $g(x,t,\omega)$ which are predictable and for which there is a sequence of stopping times τ_n such that $\tau_n \uparrow +\infty$ and for all $n \in \mathbb{N}$ and $t > 0$

$$\mathbb{E}\left[\int_0^{t\wedge\tau_n}\int_{\mathbb{R}^d} |g(x,s,\cdot)|^2 \mu(dx,ds)\right] < +\infty,$$

and it holds that

$$fg = 0. \tag{1.2.2}$$

If X_t can be expressed only by first three terms in (1.2.1), we call it a **continuous semi-martingale**.

- *Remarks:* since there are no non-trivial continuous local martingales of bounded variation, it follows that the **decomposition** of a continuous semi-martingale in (1.2.1) into the initial value, continuous local martingale and a continuous adapted process of bounded variation is unique (and process M is called a *continuous martingale part* of X).

- Continuous semi-martingale is a continuous process.

- Any discontinuity that a semi-martingale X might posses comes from last two terms in the decomposition (corresponding to the Poisson random measure).

- The property (1.2.2) means that last two terms in the decomposition do not have any common discontinuities.

The generalisation to a multi-dimensional case is straightforward: a stochastic process $X = (X^1, \ldots, X^d) = (X^1_t, \ldots, X^d_t)_{t\geq 0}$ is called a d**-dimensional semi-martingale** if, for $t \geq 0$, X_t can be written as

$$X^i_t = X^i_0 + A^i_t + M^i_t + \int_0^{t+}\int_{\mathbb{R}^d} f^i(x,s,\cdot)N(dx,ds) + \int_0^{t+}\int_{\mathbb{R}^d} g^i(x,s,\cdot)\tilde{N}(dx,ds), \tag{1.2.3}$$

$$i = 1, \ldots, d, \tag{1.2.4}$$

where

- $X_0^i, A_t^i, M_t^i, f^i, g^i, N$ and $\tilde{N}$ satisfy all conditions as corresponding objects above for all $i = 1, \dots, d$,

- (1.2.2) holds for f^i and g^j, for every $i, j = 1, \dots, d$.

In previous sections we discussed integration with respect to, first, Poisson random measures and compensated Poisson random measures, and, finally, local martingales. This included adapted, continuous processes of bounded variation. Actually, it can be shown that the linear combination of those processes form the largest class of integrators that we can use, that is, that we know how to integrate with respect to.

In order to do that, we first define the quadratic variational process of a continuous semi-martingale $X = X_0 + M + A$ to be

$$\langle X \rangle_t = X_0^2 + \langle M \rangle_t, \quad t > 0.$$

Definition 1.2.2. Let $\mathscr{L}_2(X)$ be the space of all real predictable processes $H = \{H_t\}_{t \geq 0}$ such that for all $T > 0$

$$\int_0^T |H_s(\omega)| d|A|_s + \int_0^t H_s^2 d\langle M \rangle_s < +\infty \ a.s.$$

Then, for $H \in \mathscr{L}_2(X)$, we define the **stochastic integral of H with respect to the continuous semi-martingale X** to be the process $I(H;X)$ defined by

$$I(H;X)_t = \int_0^t H_s dA_s + \int_0^t H_s dM_s, \quad t \geq 0, \tag{1.2.5}$$

and we use the notation for $I(H;X)_t$ to be as before: $\int_0^t H_s dX_s$ or $\int_0^t H_S(\omega) dX_s(\omega)$.

- *Remarks:* note that the integral is well-defined since it is a sum of two integrals that have been defined before: for $H \in \mathscr{L}_2(X)$, $\int_0^t H_s dA_s$ is a standard Lebesgue-Stieltjes integral and $\int_0^t H_s dM_s$ is a stochastic integral with respect to a local martingale.

- $\int_0^t H_s dA_s$ is an adapted process of bounded variation on finite intervals.

- $\int_0^t H_s dM_s \in \mathscr{M}_2^{loc}$.

- Therefore, $I(H;X)$ is a continuous semi-martingale, and its decomposition is given by (1.2.5) (it consists of two processes).

We proceed by mentioning one of the most important tools one needs when analysing semi-martingales - the famous Itô lemma, which says that a transformation of a semi-martingale is again a semi-martingale and provides a way to calculate it (that is, its integro-differential form). We give it in a multidimensional case.

Theorem 1.2.3. ([IW89, Theorem 2.5.1]) Take $\{X_t\}_{t\geq 0}$ to be a d-dimensional semi-martingale and denote $f = (f^1, \ldots, f^d)$ and $g = (g^1, \ldots, g^d)$. Assume that the function g is bounded, that is,

$$g^i(x,t,\omega) \leq M, \quad \forall i = 1, \ldots, d,\, t \geq 0, x \in \mathbb{R}^d,\, \omega \in \Omega.$$

Take a function $F \in C^2(\mathbb{R}^d)$. Then, $F(X) = (F(X_t))_{t\geq 0}$ is again a semi-martingale and we can calculate the value of $F(X_t)$ for $t \geq 0$ via the following integro-differential formula:

$$\begin{aligned}
F(X_t) - F(X_0) = & \sum_{i=1}^{d} \int_0^t \frac{\partial F}{\partial x_i}(X_s) dA_s^i + \sum_{i=1}^{d} \int_0^t \frac{\partial F}{\partial x_i}(X_s) dM_s^i \\
& + \frac{1}{2} \sum_{i,j=1}^{d} \int_0^t \frac{\partial F}{\partial x_i \partial x_j}(X_s) d\left\langle M^i, M^j \right\rangle_s \\
& + \int_0^{t+} \int_{\mathbb{R}^d} (F(X_{s-} + f(x,s,\cdot)) - F(X_{s-})) N(dx,ds) \\
& + \int_0^{t+} \int_{\mathbb{R}^d} (F(X_{s-} + g(x,s,\cdot)) - F(X_{s-})) \tilde{N}(dx,ds) \\
& + \int_0^{t} \int_{\mathbb{R}^d} \left(F(X_s + g(x,s,\cdot)) - F(X_s) - \sum_{i=1}^{d} g^i(x,s,\cdot) \frac{\partial F}{\partial x_i}(X_s) \right) \hat{N}(dx,ds).
\end{aligned}$$

1.3. Stochastic Differential Equations

We are now ready to formalise the problem we are discussing in this work. In the previous section, we discussed that we are looking for a model that will describe movement of a random process through time. Informally, one type of random movement will be described by the equation of the following form:

$$dX_t = b(X_t)dt + \sigma(X_t)dB_t\,, \quad t \geq 0, \tag{1.3.1}$$

where B is a standard Brownian motion and the process starts from some (possibly random) point X_0. Thus, the above equation assumes that our process follows some drift function through time, which is given by a coefficient b, and it is randomly fluctuating around it, which is governed by an independent path of Brownian motion that is additionally transformed by a coefficient σ.

If we want to observe a d-dimensional stochastic process, the underlying random motion B can be a standard n-dimensional Brownian motion, where n is some other dimension than d. To make sense of the equation (1.3.1), we will consider X_t and B_t for all $t \geq 0$ and b to be column vectors, and σ a $d \times n$ matrix.

If the starting position X_0 is some fixed point $x \in \mathbb{R}^d$, we denote the process by X^x and write the equation (1.3.1) in the following way:

$$dX_t^x = b(X_t^x)dt + \sigma(X_t^x)dB_t\,, \quad X_0^x = x \in \mathbb{R}^d\,, \tag{1.3.2}$$

Definition 1.3.1. The model (1.3.1) is called the **stochastic differential equation** (in short we will write **SDE**) and a process X defined by (if the following integrals exist)

$$X_t := X_0 + \int_0^t b(X_s)ds + \int_0^t \sigma(X_s)dB_s, \quad t \geq 0,$$

is called the **strong solution** of that equation Here, X_0 is an $\mathscr{F}_0$-measurable variable, B is a standard n-dimensional Brownian motion, $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times n}$, $b : \mathbb{R}^d \to \mathbb{R}^d$.

- *Remark:* the term *strong solution* assumes that the process X is constructed from a given (fixed) Brownian motion B. If we are allowed to construct the Brownian motion and the solution at the same time, we would call X **the weak solution**.

If $X_0 = x \in \mathbb{R}^d$, the solution of the SDE is a process X^x starting from the point x at time 0 and (if the integrals do exist) it is given by

$$X_t^x := x + \int_0^t b(X_s^x)ds + \int_0^t \sigma(X_s^x)dB_s, \quad t \geq 0.$$

Definition 1.3.2. We call a solution of the SDE to be **pathwise unique** if, for any two solutions X^x and $\tilde{X}^x$ starting from the same point $x \in \mathbb{R}^d$ (that is, $X_0 = \tilde{X}_0 = x$) and driven by the same Brownian motion B, it holds that their trajectories are a.s. equal, that is,

$$\mathbb{P}(X_t = \tilde{X}_t, \forall t \geq 0) = 1$$

(that is, X^x and $\tilde{X}^x$ are indistinguishable).

Naturally, the solution might or might not exist, depending on the existence of the integrals from the definition, and in the case that it does exist, it can be more or less nice. The theory that discusses the existence and uniqueness of the solution and its properties is extensive and is a topic in itself. It is very well-studied in the literature, so there are many references. However, it is beyond the scope of this work so we will just mention the most common results in this area and restrict ourselves to the special case of our interest - the so-called *diffusion processes*.

Definition 1.3.3. A **diffusion process** is a family of solutions of the SDE (1.3.1) (for each starting point $X_0 = x_0 \in \mathbb{R}^d$ there is one process which solves the SDE) which is also a strong Markov process with continuous paths.

- *Remark:* we will see that a diffusion process can be obtained in another way, by considering a set of the so-called *transition probabilities*.

Standard assumptions that ensure existence and (pathwise) uniqueness of the solution are local Lipschitz property together with the (at most) linear growth of coefficients b and σ. Formally, assume the following:

(C1) *local boundedness:* for any $r > 0$,

$$\sup_{x \in B_r(0)} \left(|b(x)| + \|\sigma(x)\|_{\mathrm{HS}}\right) < \infty;$$

(C2) *local Lipschitz property:* for any $r > 0$ there is $\Gamma_r > 0$ such that for all $x, y \in B_r(0)$,

$$2\langle x-y, b(x)-b(y)\rangle + \|\sigma(x)-\sigma(y)\|_{\text{HS}}^2 \leq \Gamma_r |x-y|^2 \,;$$

(C3) *linear growth:* there is $\Gamma > 0$ such that for all $x \in \mathbb{R}^d$,

$$2\langle x, b(x)\rangle + \|\sigma(x)\|_{\text{HS}}^2 \leq \Gamma(1+|x|^2)\,,$$

where $B_r(x)$ denotes the open ball with radius $r > 0$ around $x \in \mathbb{R}^d$, and $\|M\|_{\text{HS}}^2 := \operatorname{Tr} MM^T$ is the Hilbert-Schmidt norm of a real matrix M. Under **(C1)-(C3)**, the solution to the SDE is a diffusion process. More precisely, from these conditions we conclude that:

- **(C2)** & **(C3)** $\Longrightarrow$ the SDE in (1.3.2) admits a strong solution that is pathwise unique (see, for example, [Dur96, Theorems 5.3.1 and 5.3.2]),
- pathwise uniqueness $\Longrightarrow$ uniqueness in distribution $\Longrightarrow$ X^x has a transition kernel $p(t,x,dy) = \mathbb{P}(X_t^x \in dy)$, $t \geq 0$, $x \in \mathbb{R}^d$ (see the proof of [Dur96, Theorems 5.4.1]),
- **(C3)** $\Longrightarrow$ the solution does not explode,
- construction of the solution (the so-called Picard's iteration) ensures that sample paths are a.s. continuous, that is, the solution is a continuous process,
- **(C1)** & **(C3)** $\Longrightarrow$ the solution is a strong Markov process (see [Dur96, Theorems 5.4.5, 5.4.6 and 5.6.1]),
- the solution X is also a C_b-**Feller process**, that is, the corresponding **semigroup**, the set of linear operators defined by the kernel $p(t,x,dy)$ with

$$P_t f(x) := \mathbb{E}^x[f(X_t)] = \int_{\mathbb{R}^d} f(y) p(t,x,dy)\,, \qquad t \geq 0\,,\ x \in \mathbb{R}^d\,,\ f \in B_b(\mathbb{R}^d)\,,$$

  satisfies $P_t(C_b(\mathbb{R}^d)) \subseteq C_b(\mathbb{R}^d)$ ([Maj16, Lemma 2.5]).

  - Here, $B_b(\mathbb{R}^d)$ and $C_b(\mathbb{R}^d)$ denote the spaces of bounded Borel measurable functions and bounded continuous functions, respectively.
  - This property automatically implies that $\{X_t\}_{t\geq 0}$ is a strong Markov process with respect to the right-continuous and completed version of the underlying natural filtration.

- *Remark:* in the above-mentioned lemma the author assumes that $b(x)$ is continuous, but the assertion of the lemma also holds true in the case when $b(x)$ is locally bounded (condition **(C1)**). In particular, this automatically implies that $\{X_t\}_{t\geq 0}$ is a strong Markov process with respect to the right-continuous and completed version of the underlying natural filtration.

- If $b(x)$ and $\sigma(x)$ are Lipschitz continuous then $\{X_t\}_{t\geq 0}$ is a C_∞-Feller process, that is, $P_t(C_\infty(\mathbb{R}^d)) \subseteq C_\infty(\mathbb{R}^d)$ for all $t \geq 0$ (see [RW00, page 164]), where $C_\infty(\mathbb{R}^d)$ stands for the space of continuous functions vanishing at infinity.

A transition kernel of a Markov process will play a key role in determining a properties of a solution of the SDE as it describes the movement of a process. Therefore, we list some properties it possesses:

(i) there exists $\delta > 0$ such that for all $x \in \mathbb{R}^d$

$$\lim_{t\to 0} \frac{1}{t} \int_{\mathbb{R}^d} |x-y|^{2+\delta} p(t,x,dy) = 0,$$

(ii) denote $c(x) := \sigma(x)\sigma(x)^T$, then for all $x \in \mathbb{R}^d$

(a)

$$\lim_{t\to 0} \frac{1}{t} \int_{\mathbb{R}^d} (y-x) p(t,x,dy) = b(x),$$

(b)

$$\lim_{t\to 0} \frac{1}{t} \int_{\mathbb{R}^d} (y-x)^2 p(t,x,dy) = c(x).$$

- *Remark:* these properties can be used to take an alternative approach to a diffusion process. Namely, a diffusion process can equivalently be defined as a Markov process whose transition kernel satisfies (i) and (ii) for some functions $c(x)$ and $b(x)$, called the **diffusion** and **drift** (or *displacement*) coefficient. One then shows that conditions **(C1)-(C3)** ensure that a solution of the SDE is exactly a diffusion process with diffusion coefficient being the function $c(x) = \sigma(x)\sigma(x)^T$ and drift coefficient being the function $b(x)$ given in the equation (1.3.2) (see [GS79, Theorem 3.10.2]).

- Another alternative way of looking at a diffusion process is via the theory of partial differential equations ([SV06]). Namely, we consider the fundamental solution

$p(s,x,t,y)$ of the equation

$$\frac{\partial p}{\partial s} + \sum_{j=1}^{d} b_j(s,x)\frac{\partial p}{\partial x_j} + \frac{1}{2}\sum_{i,j=1}^{d} a_{ij}(s,x)\frac{\partial^2 p}{\partial x_i \partial x_j} = 0.$$

Under **(C1)-(C3)**, the fundamental solution exists. Take this function $p(s,x,t,y)$ to be the transition probability of a Markov process and call this process a diffusion process with diffusion coefficent a and drift b. One can check that such a process also coincides with the solution of the SDE with those coefficients.

↬ Finally, a last equivalent way to obtain a diffusion process is based on posing a so-called **martingale problem**. In order to do that, we need to define a probability space and a filtration on it. Since we will only be interested in the process X, and not the underlying space, we take the *canonical* (sample-path) space $\Omega^c = C([0,+\infty\rangle;\mathbb{R}^d) := \{f : [0,+\infty\rangle \to \mathbb{R}^d : f$ is a continuous function $\}$. Then, define a process $X = \{X_t\}_{t\geq 0}$ on Ω^c by $X_t(\omega) = \omega(t)$. We call X a projection process. Let filtration $\{\mathscr{F}_t^c\}_{t\geq 0}$ be its *natural filtration*, that is, the filtration generated by X_t for $0 \leq t < +\infty$. Then Ω^c is a complete separable metric space with the topology defined by uniform convergence on bounded intervals (that is, take $\mathscr{F}^c$ to be the completion of the natural filtration). For a set of coefficients $c(x) \in \mathbb{R}^{d\times d}$ and $b(x) \in \mathbb{R}^d$ for $x \in \mathbb{R}^d$ such that $c(x)$ is a positive semi-definite matrix, define an operator L on functions $f \in C^2(\mathbb{R}^d)$ by

$$\mathscr{L}(f)(x) = \mathscr{L}f(x) := \sum_{j=1}^{d} b_j(x)\frac{\partial f}{\partial x_j} + \frac{1}{2}\sum_{i,j=1}^{d} a_{ij}(x)\frac{\partial^2 f}{\partial x_i \partial x_j}.$$

We say that a measure $\mathbb{P}^x$ is a **solution to the martingale problem** for b and a starting at the point $x \in \mathbb{R}^d$ if

(i) $\mathbb{P}^x$ is a probability measure on $(\Omega^c, \mathscr{F}^c)$ such that $\mathbb{P}^x(X_0 = x) = 1$,

(ii) for all $f \in C^2(\mathbb{R}^d)$, $f(X_t) - f(x) - \int_0^t \mathscr{L}f(X_s)\,ds$ is a local martingale with respect to the space $(\Omega^c, \mathscr{F}^c, \mathbb{P}^x)$.

One can show that, under **(C1)-(C3)**, the solution to the martingale problem for given coefficients a and b and a starting position x_0 does exist, it is unique, and it is equal to the solution of the SDE (1.3.2) with that same coefficients b and σ (where

σ is a square root of a) in a sense that a solution to the martingale problem is the distribution of the solution to the SDE ([Dur96, Theorem 5.4.5]). Furthermore, the process X is a strong Markov process with respect to the filtration $\{\mathscr{F}_t^c\}_{t\geq 0}$ and it has continuous sample paths.

Based on these different views over a diffusion process (that is, equivalences of the definition of a diffusion process), we can deduce many useful properties of such a process. Since some of them will be important for us later on, we briefly summarise the most important ones:

- conditions **(C1)-(C3)** ensure that the solution to the SDE (1.3.2) exists, is unique and possesses a strong Markov property. We denote it by $\{X_t^x\}_{t\geq 0}$. This implies that we can obtain a transition kernel for our $p(t,x,dy) = \mathbb{P}(X_t^x \in dy)$, $t \geq 0$, $x \in \mathbb{R}^d$. Since there is 1-1 correspondence between distributions of solutions to the SDE and solutions of the Martingale problem (which are distributions on the canonical space $(\Omega^c, \mathscr{F}^c)$, we conclude that our transition kernel defines a unique probability measure $\mathbb{P}^x$ on the canonical space such that the projection process, denoted by $\{X_t\}_{t\geq 0}$, is a strong Markov process (with respect to the completion of the corresponding natural filtration), it has continuous sample paths, and the same finite-dimensional distributions (with respect to $\mathbb{P}^x$) as $\{X_t^x\}_{t\geq 0}$ (with respect to $\mathbb{P}$).
- Also, it holds that the martingale problem for b and $a := \sigma\sigma^T$ and starting position x is well-posed, that is

$$f(X_t) - f(X_0) - \int_0^t \mathscr{L} f(X_s)\, ds\,, \qquad t \geq 0\,,$$

is a $\mathbb{P}^x$-local martingale for every $x \in \mathbb{R}^d$ and every $f \in C^2(\mathbb{R}^d)$, where

$$\mathscr{L} f(x) \;:=\; \langle b(x), \nabla f(x)\rangle + \frac{1}{2}\mathrm{Tr}\,\sigma(x)\sigma(x)^T\nabla^2 f(x)\,.$$

As it turns out, the operator $\mathscr{L}$ defined above will play a key role in determining stability properties of a stochastic process. It has connections with an infinitesimal generator.

Definition 1.3.4. Let $\{X_t\}_{t\geq 0}$ be an $\mathbb{R}^d$-valued Markov process with semigroup $\{P_t\}_{t\geq 0}$ (defined as above). The **infinitesimal generator** (with respect to $(\|\cdot\|_\infty, B_b(\mathbb{R}^d))$) is a linear operator $\mathscr{A} : \mathscr{D}_{\mathscr{A}} \to B_b(\mathbb{R}^d)$ defined by

$$\mathscr{A} f := \lim_{t\to 0} \frac{P_t f - f}{t}, \qquad f \in \mathscr{D}_{\mathscr{A}} := \left\{ f \in B_b(\mathbb{R}^d) : \lim_{t\to 0} \frac{P_t f - f}{t} \text{ exists in } \|\cdot\|_\infty \right\}.$$

- If $b(x)$ and $\sigma(x)$ are continuous, then the infinitesimal generator $(\mathscr{A}, \mathscr{D}_{\mathscr{A}})$ of the solution of the SDE $\{X_t\}_{t\geq 0}$ (with respect to the Banach space $(B_b(\mathbb{R}^d), \|\cdot\|_\infty)$) satisfies $C_c^2(\mathbb{R}^d) \subseteq \mathscr{D}_{\mathscr{A}}$ and

$$\mathscr{A}|_{\mathscr{D}_{\mathscr{A}}} = \mathscr{L}.$$

  - $\|\cdot\|_\infty$ and $C_c^2(\mathbb{R}^d)$ denote the supremum norm and the space of twice continuously differentiable functions with compact support, respectively.
  - Obviously, the domain of the operator $\mathscr{L}$ (denoted by $\mathscr{D}_{\mathscr{L}}$) is larger than the domain of $\mathscr{A}$, so we call $\mathscr{L}$ the **extended generator** of the process X (see [MT93b, Section 1] for more details).

1.4. ERGODICITY

In previous sections, by defining objects and mathematical models of our interest, we set the basis for formalising the problem we want to tackle and methods that we will use. Doing that will be the final step we will take before presenting all our results.

We want to discuss *stochastic stability* of a diffusion process X. Recall that such a process has the transition probability $p(t,x,dy) = \mathbb{P}^x(X_t \in dy)$ for $t \geq 0$ and $x \in \mathbb{R}^d$ which defines the set of linear operators $\{P_t\}_{t\geq 0}$. Our **problem** is to investigate a possible convergence of the process to some stable state - equilibrium. Mathematically, we can write it in the following form:

$$r(t)\|\mathbb{P}^x(X_t \in \cdot) - \pi(\cdot)\| \xrightarrow{\textbf{?}} 0,$$

where

- $\mathbb{P}^x(X_t \in \cdot)$ is the distribution of the process at time t which started from point x,
- $\pi(\cdot)$ is its invariant distribution (the equilibrium),
- $r(t)$ is the rate of the convergence,
- $\|\cdot\|$ is an appropriate norm.

A process that satisfies that will be called **ergodic**.

Namely, this problem raises the following **questions**:

$\rightsquigarrow$ under which conditions on the coefficients does the diffusion process admit a unique invariant probability measures?

$\rightsquigarrow$ do the marginals of the process converge to the invariant measure, in some norm?

$\rightsquigarrow$ at which rate does the convergence occur?

$\rightsquigarrow$ do we get the same results for non-local operators (that is, when we add jumps to our process)?

$\rightsquigarrow$ can we extend this analysis to some more general stochastic processes?

From an above formula we see that for an ergodic process the distribution at time t is starting to look more and more like some fixed distribustion π that we call an *invariant* distribution. This distribution itself has some nice properties - invariance and finiteness. To define this, we will need the following notation: for a positive measure μ on a measurable space $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$ we write

$$\mu P_t(A) = \int_{\mathbb{R}^d} p(t,x,A)\,\mu(dx)\,, \quad t \geq 0, A \in \mathscr{B}(\mathbb{R}^d).$$

Definition 1.4.1. Let X be a Markov process. A (positive) measure π on $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$ is called

- **finite** if $\pi(\mathbb{R}^d) < +\infty$, otherwise, it called **non-finite**,
- **probability measure** if $\pi(\mathbb{R}^d) = 1$,
- **invariant** or **stationary measure** for X if

$$\pi P_t = \pi, \quad t \geq 0.$$

Obviously, in order to check whether our process approaches to some equilibrium state, we need to be able to measure how far away from it this process is at some point. More precisely, we need some function that measures the distance between two distributions. In mathematics, such a distance function is called the **norm**. Therefore, ergodicity properties of our process need to be discussed with respect to a certain norm function: maybe convergence does not occur in one norm, but it does in some other. Hence, we proceed by naming several possible norms that will be of our interest. For a Borel-measurable function $g : \mathbb{R}^d \to \mathbb{R}$ and a signed measure μ on $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$ denote

$$\mu(g) := \int_{\mathbb{R}^d} g(x)\mu(dx).$$

Definition 1.4.2. Let $f : \mathbb{R}^d \to [1, +\infty\rangle$ be a Borel-measurable function. For a signed measure μ on $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$ define the following norms:

- f-**norm** is $\|\mu\|_f := \sup_{g:|g|\leq f} |\mu(g)|$,
- the **total variation norm** is $\|\mu\|_{\mathrm{TV}} := \sup_{A\in\mathbb{R}^d} \mu(A) - \inf_{A\in\mathbb{R}^d} \mu(A)$.

↬ *Remark:* for $f(x) = 1$ for all $x \in \mathbb{R}^d$, that is, for a constant function equal to 1, the f-norm (i.e. 1-norm) is nothing more than the total variation norm.

↬ $\|\mu\|_{\mathrm{TV}} = \sup_{B\in\mathscr{B}(\mathbb{R}^d)} |\mu(B)|$ for a signed measure μ on $\mathscr{B}(\mathbb{R}^d)$.

As it turns out, in order to obtain convergence in f-norm (for a Borel function $f \geq 1$) one needs assume a certain level of regularity in the behaviour of a process X. Namely, we will need to ensure the process is *irreducible* and *aperiodic*. This will usually be provided by imposing certain regularity assumptions (like *uniform ellipticity*) on a diffusion coefficient $c(x)$. In the case when $c(x)$ is not regular enough, the topology induced by the total variation distance becomes too "rough", that is, it cannot completely capture the singular behaviour of $\{X_t\}_{t\geq 0}$. Formally, $p(t,x,dy)$ cannot converge to the underlying invariant probability measure (if it exists) in this topology, but it still might converge in some other (see [San17] and the references therein). If we find it, this other topology will then provide convergence in some weaker sense. In this situation, we naturally resort to Wasserstein distances which, in a certain sense, do induce a finer topology, because, as we will see, convergence with respect to a Wasserstein distance implies the weak convergence of probability measures (see [Vil09, Theorems 6.9 and 6.15]).

Definition 1.4.3. Let ρ be a metric on $\mathbb{R}^d$ and $p \geq 0$.

- Let $\mathbb{R}^d_\rho$ denotes the topology induced by ρ and $\mathscr{B}(\mathbb{R}^d_\rho)$ be the corresponding Borel σ-algebra. So, $(\mathbb{R}^d_\rho, \mathscr{B}(\mathbb{R}^d_\rho))$ is a measurable space.
  - An important example of a metric is the standard d-dimensional **Euclidean metric**, denote it by ρ_E. Then $\mathbb{R}^d_\rho$ is the set of all open subsets of $\mathbb{R}^d$, in classical sense.
- Define $\mathscr{P}_{\rho,p}$ to be the space of all probability measures μ on $\mathscr{B}(\mathbb{R}^d_\rho)$ having finite p-th moment, that is,

$$\int_{\mathbb{R}^d} \rho(x_0,x)^p \mu(dx) < \infty, \quad \text{for some (and then any) } x_0 \in \mathbb{R}^d.$$

  - $\mathscr{P}_{\rho,0} := \mathscr{P}_\rho$ - this is just the space of all probability measures on $\mathscr{B}(\mathbb{R}^d_\rho)$
  - $\mathscr{P}_{\rho_E,0} := \mathscr{P}_p$

  - $\mathscr{P}_{\rho_E} := \mathscr{P}$

- If $p \geq 1$ and $\mu, \nu \in \mathscr{P}_\rho$, the $\mathscr{L}^p$-**Wasserstein distance** between μ and ν is defined as
$$\mathscr{W}_{\rho,p}(\mu,\nu) := \inf_{\Pi \in \mathscr{C}(\mu,\nu)} \left(\int_{\mathbb{R}^d \times \mathbb{R}^d} \rho(x,y)^p \, \Pi(dx,dy) \right)^{1/p},$$
where $\mathscr{C}(\mu,\nu)$ is the family of couplings of μ and ν, that is, $\Pi \in \mathscr{C}(\mu,\nu)$ if and only if Π is a probability measure on $\mathbb{R}^d \times \mathbb{R}^d$ having μ and ν as its marginals.

↬ *Remarks:* $\mathscr{W}_{\rho,p}$ satisfies the axioms of a (not necessarily finite) distance on $\mathscr{P}_\rho$.

  - However, if we observe a restriction of $\mathscr{W}_{\rho,p}$ to $\mathscr{P}_{\rho,p}$, such a distance function is finite ([Vil09, Theorem 6.4]).

↬ If $(\mathbb{R}^d, \rho)$ is a Polish space, then $(\mathscr{P}_{\rho,p}, \mathscr{W}_{\rho,p})$ is also a Polish space (see [Vil09, Theorem 6.18]).

↬ Of our special interest will be the situation when ρ takes the form $\rho(x,y) = f(|x-y|)$ for $x,y \in \mathbb{R}^d$ (that is, $\rho = f \circ \rho_E$), where $f : [0,+\infty\rangle \to [0,+\infty\rangle$ is a non-decreasing concave function satisfying $f(t) = 0$ if and only if $t = 0$. In this situation, the corresponding Wasserstein space does not have to be a Polish space and is denoted by
$$(\mathscr{P}_{\rho,p}, \mathscr{W}_{\rho,p}) := (\mathscr{P}_{f,p}, \mathscr{W}_{f,p}).$$

  - If $f(t) = \mathbb{1}_{\langle 0,+\infty\rangle}(t)$, then $\mathscr{W}_{f,p}(\mu,\nu) = \|\mu - \nu\|_{\mathrm{TV}}$ for all $p \geq 1$.
  - If$f(t) = t$, then ρ is an Euclidean metric so $\mathscr{P}_{\rho,p} = \mathscr{P}_p$ and the corresponding Wasserstein space is denoted just by $(\mathscr{P}_p, \mathscr{W}_p)$ (which is always a Polish space).

↬ Wasserstein distances metrize weak convergence. Namely, this means that, for a sequence of measures $\{\mu_n\}_{n\geq 0}$ in $\mathscr{P}_{\rho,p}$ and a measure μ in $\mathscr{P}_\rho$, then the convergence with respect to a Wasserstein distance implies the weak convergence of probability measures. Mathematically, it holds that
$$\mathscr{W}_{\rho,p}(\mu_n,\mu) \longrightarrow 0 \implies \mu_n \xrightarrow{\mathrm{w}} \mu, \quad n \to +\infty.$$

– The weak convergence of μ_n to μ, denoted by $\mu_n \xrightarrow{\text{w}} \mu$, means that for any $g \in C_b(\mathbb{R}^d)$, $\mu(g) \longrightarrow \mu(g)$.

↬ We mentioned that the control in Wasserstein distance is somewhat weaker then the control in total variation distance. However, it holds that if $(\mathbb{R}^d, \rho)$ is a Polish space, then for all $x_0 \in \mathbb{R}^d$

$$\mathscr{W}_{\rho,p} \leq 2^{\frac{1}{p'}} \left(\int_{\mathbb{R}^d \times \mathbb{R}^d} \rho(x_0, x)^p \right)^{1/p}, \quad \frac{1}{p} + \frac{1}{p'} = 1,$$

so in a special case that $p = 1$ and $\rho(\mathbb{R}^d) < +\infty$, we have that for all $\mu, \nu \in \mathscr{P}_\rho$

$$\mathscr{W}_{\rho,1}(\mu, \nu) \leq D\|\mu - \nu\|_{\mathrm{TV}}, \quad \text{for some } D > 0.$$

Now that we have discussed all aspects of the convergence, it is time to formally define the convergence to an equilibrium itself.

Definition 1.4.4. We call a Markov process $\{X_t\}_{t\geq 0}$ r-**ergodic with respect to the norm** $\|\cdot\|$ if it possesses an invariant probability measure π and there exists a non-decreasing function $r : [0, +\infty\rangle \to [1, +\infty\rangle$ such that

$$\lim_{t\to\infty} r(t)\|p(t, x, dy) - \pi(dy)\| = 0, \qquad x \in \mathbb{R}^d.$$

An ergodic process X is called

- **geometrically** (or **exponentially**) ergodic if

$$r(t) = \mathrm{e}^{\kappa t}, \quad \text{for some } \kappa > 0,$$

- **sub-geometrically** ergodic if

$$\lim_{t\to\infty} \frac{\ln r(t)}{t} = 0.$$

↬ *Remarks:* the speed of the convergence of a sub-geometrically ergodic process is less than exponential.

In the discussion above we already mentioned that the type and speed of the convergence we can obtain will depend on the structural properties of the process. The aim of this work is to detect those conditions and conlude what kind of a convergence they imply. So, in order to do that, we will recall some of the features of Markov processes together with the most import results concerning them. Our main references are [MT93a], [Twe94], [BG68]) and [FR05].

As always, we have a stochastic basis $(\Omega, \mathscr{F}, \{\mathscr{F}_t\}_{t\geq 0}, \{\mathbb{P}^x\}_{x\in\mathbb{R}^d})$, where $p(t,x,dy) := \mathbb{P}^x(X_t \in dy)$, for $t \geq 0$ and $x \in \mathbb{R}^d$, correspond to a Markov process X with càdlàg sample paths and state space $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$. Also, assume that $p(t,x,dy)$ is a probability measure, that is, $\{X_t\}_{t\geq 0}$ does not admit a cemetery point in the sense of [BG68]. Observe that this is not a restriction since, as we have already commented, our assumptions on the coefficients will imply $\{X_t\}_{t\geq 0}$ is non-explosive.

Definition 1.4.5. A Markov process X is called

- **ϕ-irreducible** if there exists a σ-finite measure ϕ on $\mathscr{B}(\mathbb{R}^d)$ such that for all $B \in \mathscr{B}(\mathbb{R}^d)$

$$\phi(B) > 0 \implies \int_0^{+\infty} p(t,x,B)dt > 0, \quad \text{for all } x \in \mathbb{R}^d;$$

  - *Remarks:* the condition for ϕ-irreducibility is equivalent to asking that whenever $\phi(B) > 0$, we have that

$$\mathbb{E}^x\left[\int_0^{+\infty} \mathbb{1}_B(X_t)dt\right] > 0, \quad \text{for all } x \in \mathbb{R}^d,$$

    which can be interpreted as a condition that we expect our process to stay sufficiently long (or positive amount of time) in any large-enough set (in a sense that it is a set of positive measure).

  - If X is ϕ-irreducible, then the irreducibility measure ϕ can be maximized (in the sense of absolute continuity). More precisely, this means that there exists a unique "maximal" irreducibility measure ψ such that for any measure $\bar{\phi}$, $\{X_t\}_{t\geq 0}$ is $\bar{\phi}$-irreducible if and only if $\bar{\phi}$ is absolutely continuous with respect to ψ (see [Twe94, Theorem 2.1]). In view to this, when we refer to an irreducibility measure we actually refer to the maximal irreducibility measure.

  - If π a an invariant measure, then π is a maximal irreducibility measure.

For a ϕ-irreducible process, we define a set $C \in \mathscr{B}(\mathbb{R}^d)$ to be

– **accessible** if $\psi(C) > 0$,

– a **petite** set if there exists a probability measure η_C on $\mathscr{B}([0,+\infty\rangle)$ and a non-trivial σ-finite measure ν_C on $\mathscr{B}(\mathbb{R}^d)$ such that

$$\int_0^{+\infty} p(t,x,B)\eta_C(dt) \geq \nu_C(B), \quad x \in C, B \in \mathscr{B}(\mathbb{R}^d).$$

↬ *Remarks:* a meaning of a petite set is that it is small enough so that we can guarantee that, starting from it, for any fixed set, the expected amount of time our process stays in that fixed set is longer than some positive constant (which depends on the size of that set).

↬ a ϕ-irreducible process always possesses an accessible closed petite set.

↬ if X is ϕ-irreducible and C_b-Feller process, then every compact set is petite.

A ϕ-irreducible process is further called

– **transient** if there exists a countable covering of $\mathbb{R}^d$ with sets $\{B_j\}_{j\in\mathbb{N}} \subseteq \mathscr{B}(\mathbb{R}^d)$, and for each $j \in \mathbb{N}$ there exists a finite constant $\gamma_j \geq 0$ such that $\int_0^\infty p(t,x,B_j)\,dt \leq \gamma_j$ holds for all $x \in \mathbb{R}^d$;

  ↬ *Remark:* note that a transient Markov process cannot have a finite invariant measure. Indeed, assume that $\{X_t\}_{t\geq 0}$ is transient and that it admits a finite invariant measure π, and fix some $t > 0$. Then, for each $j \in \mathbb{N}$, with γ_j and B_j as above, we have

  $$t\pi(B_j) = \int_0^t \pi P_s(B_j)ds \leq \gamma_j \pi(\mathbb{R}^d).$$

  Now, by letting $t \to \infty$ we obtain $\pi(B_j) = 0$ for all $j \in \mathbb{N}$, which is impossible.

– **recurrent** if for all $B \in \mathscr{B}(\mathbb{R}^d)$

$$\phi(B) > 0 \implies \int_0^{+\infty} p(t,x,B)dt = +\infty, \quad \text{for all } x \in \mathbb{R}^d;$$

  ↬ *Remarks:* similarly as discussed above, the process will be transient if we can cover the state space with sets which have the property that the process cannot stay in them forever, that is, we expect the process to visit them only limited amount of time. On the other hand, the process is recurrent if the process is expected to spend infinite amount of time in any large-enough set (so, it must keep returning to any such set).

↬ Every ψ-irreducible Markov process is either transient or recurrent (see [Twe94, Theorem 2.3]).

↬ If X is recurrent, then it possesses a unique (up to constant multiples) invariant measure π (see [Twe94, Theorem 2.6]).

A recurrent process is called

- **positive recurrent** if its invariant measure is finite.
  - ↬ *Remark:* in that case, the invariant measure may be normalized to a probability measure.
- **null-recurrent** otherwise.

- **open-set irreducible** if the support of its maximal irreducibility measure ψ,

$$\operatorname{supp}\psi = \{x \in \mathbb{R}^d : \psi(O) > 0 \text{ for every open neighborhood } O \text{ of } x\},$$

has a non-empty interior;

↬ *Remark:* open-set irreducibility and C_b-Feller property ensure that every compact set is petite (see [Twe94, Theorems 5.1 and 7.1]).

- **aperiodic** if it admits an irreducible **skeleton chain**, that is, there exist $t_0 > 0$ and a σ-finite measure ϕ on $\mathscr{B}(\mathbb{R}^d)$, such that for all $B \in \mathscr{B}(\mathbb{R}^d)$

$$\phi(B) > 0 \implies \sum_{n=0}^{\infty} p(nt_0, x, B) > 0, \quad \text{for all } x \in \mathbb{R}^d.$$

↬ *Remarks:* this actually means that we can extract a Markov chain $(X_{nt_0})_{n\in\mathbb{N}}$ out of our process X such that it is a ϕ-irreducible Markov chain (that is, its expected time of staying in any large-enough set is positive).

↬ a sufficient condition, perhaps more intuitive to think about because it brings the meaning of aperiodicity to life, is the following: X is ϕ-irreducible and there exist an accessible petite set C and $t_0 > 0$ such that $p(t,x,C) > 0$ for all $x \in C$ and $t \geq t_0$. Therefore, in this sense, the aperiodic process excludes any cyclic behaviour (which is precisely the definition of aperiodicity in a discrete setting). Furthermore, [MT93a, Proposition 6.1] states that positive Harris-recurrent and aperiodic process satisfies this condition.

- **Harris-recurrent** if there exist a σ-finite measure ϕ such that for all $B \in \mathscr{B}(\mathbb{R}^d)$

$$\phi(B) > 0 \implies \mathbb{P}^x \left(\int_0^{+\infty} \mathbb{1}_A(X_t)dt = +\infty \right) = 1, \quad \text{for all } x \in \mathbb{R}^d.$$

  ↬ *Remarks:* equivalently, we call X Harris-recurrent if there exist a σ-finite measure ϕ such that for all $B \in \mathscr{B}(\mathbb{R}^d)$

$$\phi(B) > 0 \implies \mathbb{P}^x(\tau_B < +\infty) = 1, \quad \text{for all } x \in \mathbb{R}^d,$$

  where $\tau_B := \inf\{t \geq 0 : X_t \in B\}$ is the hitting time of set B.

  ↬ Obviously, Harris-recurrence implies ϕ-irreducibility and, hence, also recurrence as defined above.

  ↬ A Harris-recurrent (right-continuous) process does admit a (possibly infinite) invariant measure.

A Harris-recurrent process is further called

  – **positive Harris-recurrent** if its invariant measure is finite.

- C_b**-Feller** (or sometimes, weak Feller or just Feller) process if

$$P_t(C_b(\mathbb{R}^d)) \subseteq C_b(\mathbb{R}^d).$$

- **strong Feller** process if

$$P_t(B_b(\mathbb{R}^d)) \subseteq C_b(\mathbb{R}^d).$$

With this definition we close the discussion on the properties of a Markov process. It remains to explain methods that we will adopt in order to obtain ergodic properties of our process. In general, there are two basic methods that we will use: one is the so-called *Lyapunov energy method* (and it will be used to obtain convergence with respect to a total variation distance) and other is called the *coupling method* (and it will be used in the case when we need to consider convergence in a weaker sence then with respect to a total variation distance).

Lyapunov energy method The idea is to observe the generator of our process $\mathscr{L}$ and construct the appropriate energy function $\mathscr{V}$ (the so-called *Lyapunov energy function*). Then, $\mathscr{L}\mathscr{V}$ describes the total dissipated energy of the system. So, if

- $\mathscr{L}\mathscr{V} \geq 0$, then we have an unstable system;
- $\mathscr{L}\mathscr{V} \leq 0$, then we have a stable system.

  Furthermore, the better the bound is, the better our control over the dissipation of the energy is. So, if we can obtain better results like

  - $\mathscr{L}\mathscr{V} \leq -\mathscr{V}$, we then know that the speed of the stabilization if exponential,
  - $\mathscr{L}\mathscr{V} \leq -\varphi(\mathscr{V})$, for some concave function φ, we can conclude that the stabilization is happening, but at some slower pace, depending on the function φ.

Mathematical object that decribes $\mathscr{L}$ could be

- ”derivative of the distribution P_t”, that is, an infinitesimal generator $\mathscr{A}$ from the previous section. However, this approach will narrow down the domain of $\mathscr{L}$ so it will be harder to find a function $\mathscr{V}$.
- Therefore, we adopt the probabilistic approach to the generator $\mathscr{L}$ and consider instead the extended generator defined by the martingale problem.

Such an technique has been used for determining stability of first, Markov chains (see [Hai16]). It was called the Foster - Lyapunov method. Later, it was extended to continuous-time Markov processes. The version for sub-geometric ergodicity was developed in [DFG09]. The method itself consists of finding an appropriate petite (recurrent) set $C \in \mathscr{B}(\mathbb{R}^d)$, and constructing an appropriate function $\mathscr{V} : \mathbb{R}^d \to [1,+\infty\rangle$ (the Lyapunov function) contained in the domain of the extended generator $\mathscr{A}$ of the underlying Markov process $\{X_t\}_{t\geq 0}$, such that the Lyapunov equation (also called the **drift inequality**)

$$\mathscr{L}\mathscr{V}(x) \leq -\varphi(\mathscr{V}(x)) + \beta \mathbb{1}_C(x), \qquad x \in \mathbb{R}^d, \tag{1.4.1}$$

holds for some $\beta \in \mathbb{R}$ (see [DFG09, Theorem 3.4]). Recall that for a diffusion process we know how the extended generator looks like: its domain $C^2(\mathbb{R}^d) \subseteq \mathscr{D}_{\mathscr{L}}$, and for every $f \in C^2(\mathbb{R}^d)$,

$$\mathscr{L} f(x) := \langle b(x), \nabla f(x) \rangle + \frac{1}{2} \mathrm{Tr} \left[\sigma(x) \sigma(x)^T \nabla^2 f(x) \right] .$$

The second step, after establishing the drift inequality, is to use the following result (which we give here because of its importance):

Theorem 1.4.6. ([DFG09, Theorem 3.2]) Let X be ϕ-irreducible and aperiodic process. Assume that (1.4.1) holds for some $C, \mathscr{V}, \varphi$ such that $\sup_C \mathscr{V} < +\infty$ and $\lim_{t \to +\infty} \varphi'(t) = 0$. Then, for any probability measure λ such that $\lambda(\mathscr{V}) < +\infty$, there is an invariant measure π and it holds

$$\lim_{t \to +\infty} \varphi(\Phi^{-1}(t)) \int_{\mathbb{R}^d} \| \delta_x P_t - \pi \|_{\mathrm{TV}} \lambda(dx) = 0.$$

Furthermore, there exists $D < +\infty$ such that, for all $t \geq 0$ and $x \in \mathbb{R}^d$

$$\varphi(\Phi^{-1}(t)) \| \delta_x P_t - \pi \|_{\mathrm{TV}} \leq D \mathscr{V}(x).$$

- *Remarks:* the idea of the proof is to obtain ergodicity by applying the results for Markov chain (since we have an irreducible skeleton chain).
- In full length, the theorem actually provides results for f-ergodicity for various f-norms, not just the total variation norm. It then follows that the strength of the norm is compensated by the rate of the convergence in a sense that the stronger the norm, the slower the rate. The maximal rate is achieved with the total variation norm, and the minimal with the f^*-norm, where $f^* := \varphi \circ \mathscr{V}$.
- The theorem gives the existence of the invariant distribution π. This is obtained from the drift condition. Namely, the equation in (1.4.1) implies that for any $\delta > 0$ the $\varphi \circ \Phi^{-1}$-moment of the δ-shifted hitting time of set C $\tau_C^\delta := \inf\{t \geq \delta : X_t \in C\}$ of $\{X_t\}_{t \geq 0}$ on $\mathscr{C}$ (with respect to $\mathbb{P}^x$) is finite and controlled by $\mathscr{V}(x)$ (see [DFG09, Theorem 4.1]), that is

  $$\mathbb{E}^x \left[\int_0^{\tau_C^\delta} \varphi \circ \Phi^{-1}(s) \, ds \right] < +\infty.$$

  This implies that the process X is positive Harris-recurrent which then implies that the invariant probability measure exists.

↬ However, it is known that positive Harris-recurrence alone does not immediately imply ergodicity of $\{X_t\}_{t\geq 0}$. Namely, we also need to ensure that a similar property holds for any other "reasonable" set, which will be ensured by assuming ϕ-irreducibility. Together with aperiodicity, or in this case, the property of existence of a irreducible skeleton chain ([MT93a, Proposition 6.1]), the sub-geometric ergodicity will follow. Similar conditions are required to obtain ergodicity in a discrete setting. There, $\{X_t\}_{t\geq 0}$ can also show certain cyclic behaviour which causes ergodicity not to hold (see [MT93a, Section 5] and [MT09, Chapter 5]), so aperiodicity is required (which excludes this type of behaviour). Intuitively, petite sets take a role of singletons for Markov processes on non-discrete state spaces (see [MT93a, Section 4] and [MT09, Chapter 5] for details). To be more precise, the reasoning goes as follows:

- If $\{X_t\}_{t\geq 0}$ is ψ-irreducible and C is a petite set, then indeed for any $\delta > 0$ the $\varphi \circ \Phi^{-1}$-moment of τ_B^δ, for any $B \in \mathscr{B}(\mathbb{R}^d)$ with $\psi(B) > 0$, is again finite and controlled by $\mathscr{V}(x)$ (see [DFG09, the discussion after Theorem 4.1]). Recall also that ψ-irreducibility implies that the state space (in this case $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$ can be covered by a countable union of petite sets (see [MT93a, Propositio 4.1].
- By assuming aperiodicity, the sub-geometric ergodicity of $\{X_t\}_{t\geq 0}$ follows from [FR05, Theorem 1], which states that finiteness of the $\varphi \circ \Phi^{-1}$-moment of τ_C^δ implies sub-geometric ergodicity of $\{X_t\}_{t\geq 0}$ with rate $r(t) = \varphi(\Phi^{-1}(t))$.

Coupling method As mentioned in the introduction, the aim of this method is to construct the optimal coupling strategy so that we are able to control the dissipation of the energy of the system. A usual approach is to control the dissipation by the so-called *asymptotic flatness condition*, which we will state later in the work. In general, this method is more flexible than the Foster - Lypaunov method as it allows a certain level of singularity in the behaviour of the process.

The last note of this section is the reflection over some extensions and other versions of the Foster - Lyapunov method that have appeared in the literature. An analogous approach was used in [Kha12, Chapter 4] to discuss positive recurrence of the process $\{X_t\}_{t\geq 0}$ with globally Lipschitz coefficients and with $c(x)$ being positive definite (hence, according to Theorem 2.2.1, $\{X_t\}_{t\geq 0}$ is open-set irreducible and aperiodic). Based on this result, and analysing polynomial moments of hitting times of compact sets, in [Ver97, Theorem 6] polynomial ergodicity of $\{X_t\}_{t\geq 0}$ has been obtained. In the follow up work, by using analogous techniques the same author established polynomial ergodicity of $\{X_t\}_{t\geq 0}$ without directly assuming ψ-irreducibility and aperiodicity of the process, but basing on a local irreducibility condition which we discuss below (see [Ver99, Theorem 6]).

An alternative and, in a certain sense, more general approach to our problem is based on a **local irreducibility condition**. In this approach, instead of (1.4.1), we assume a slightly more general form of the Lyapunov equation:

$$\mathscr{L}\mathscr{V}(x) \leq -\varphi(\mathscr{V}(x)) + \beta\,, \qquad x \in \mathbb{R}^d\,, \tag{1.4.2}$$

for some $\beta \in \mathbb{R}$, and instead of assuming ψ-irreducibility and aperiodicity of $\{X_t\}_{t\geq 0}$, we assume the so-called (local) **Dobrushin condition** (also known as Markov-Dobrushin condition): the Lyapunov function $\mathscr{V}(x)$ has precompact sub-level sets, and for every $\gamma > 0$ there is $t_\gamma > 0$ such that

$$\sup_{(x,y)\in\{(u,v):\,\mathscr{V}(u)+\mathscr{V}(v)\leq\gamma\}} \|p(t_\gamma, x, dz) - p(t_\gamma, y, dz)\|_{\mathrm{TV}} < 1\,, \tag{1.4.3}$$

see [Hai16, Theorem 4.1] (see also [Kul15, Chapter 1.4] and [Kul18, Chapter 3]). Observe that this condition actually means that for each $(x,y) \in \{(u,v) : \mathscr{V}(u)+\mathscr{V}(v) \leq \gamma\}$ the probability measures $p(t_\gamma, x, dz)$ and $p(t_\gamma, y, dz)$ are not mutually singular. Intuitively, the Dobrushin condition encodes ψ-irreducibility and aperiodicity of $\{X_t\}_{t\geq 0}$, and petiteness of sub-level sets of $\mathscr{V}(x)$. By using a coupling approach with an appropriately chosen Markov coupling of $\{X_t\}_{t\geq 0}$, say $\{X_t^c\}_{t\geq 0}$, the Lyapunov equation and Dobrushin condition, analogously as before, imply that the hitting (that is, coupling) time $\tau_c := \inf\{t \geq 0 : M_t^c \in \mathrm{diag}\}$ of $\{M_t^c\}_{t\geq 0}$ on $\mathrm{diag} := \{(x,x) : x \in \mathbb{R}^d\}$ is a.s. finite (with respect to the probability measure corresponding to $\{M_t^c\}_{t\geq 0}$ with any initial position $(x,y) \in \mathbb{R}^d \times \mathbb{R}^d$). Moreover, it follows that the Φ^{-1}-moment of τ_c is finite and controlled by $\mathscr{V}(x)+\mathscr{V}(y)$. Then from the coupling inequality it follows that $\{X_t\}_{t\geq 0}$ admits a unique invariant $\pi \in \mathscr{P}$, and

$$\sup_{t \geq 0} \varphi(\Phi^{-1}(t)) \| p(t,x,dy) - \pi(dy) \|_{\mathrm{TV}} < \infty, \qquad x \in \mathbb{R}^d,$$

(see [Hai16, Theorem 4.1], or [Kul15, Chapter 1.4] and [Kul18, Chapter 3] for the skeleton chain approach).

Observe that (1.4.2) follows from (1.4.1). Also, ψ-irreducibility and aperiodicity (together with (1.4.1)) imply that the Dobrushin condition holds on the Cartesian product of any petite set with itself. Namely, according to [MT93a, Proposition 6.1], for any petite set C there is $t_C > 0$ such that for the measure η_C (in the definition of petiteness) the Dirac measure in t_C can be taken (with some, possibly different, non-trivial measure ν_C). Thus, $p(t_C, x, B) \geq \nu_C(B)$ for any $x \in C$ and $B \in \mathscr{B}(\mathbb{R}^d)$, which implies

$$\sup_{(x,y) \in C \times C} \| p(t_C, x, dz) - p(t_C, y, dz) \|_{\mathrm{TV}} < 1. \tag{1.4.4}$$

If in addition $\{X_t\}_{t \geq 0}$ is C_b-Feller and open-set irreducible, as we have already commented, every compact set is petite so the above relation holds for any bounded set C, showing that, at least in this particular situation, the approach based on the Dobrushin condition is more general than the approach based on ψ-irreducibility and aperiodicity. In situations that we are interested in and will discuss, where we obtain conditions that ensure these properties, like uniform ellipticity and Lipschitz continuity of $c(x)$ or the ones from Theorem 2.2.1, if we don't make further regularity assumptions on $b(x)$ and $c(x)$, it is not clear how to check the Dobrushin condition. Situations where it shows a clear advantage are discussed in [Kul09] and [AV10]. In the first reference the author considers a Markov process obtained as a solution to a Lévy-driven SDE with highly irregular coefficients and noise term, while in the second a diffusion process with highly irregular (discontinuous) drift function and uniformly elliptic diffusion coefficient has been considered. In these concrete situations it is not clear whether one can obtain ψ-irreducibility and aperiodicity of the processes, whereas the authors obtain (1.4.4) for any compact set C (see [Kul09, Theorem 1.3] and [AV10, Lemma 3]). For more on ergodic properties of Markov processes based on the Dobrushin condition we refer the readers to [Hai16], [Kul15] and [Kul18].

2. Diffusion Processes

The theoretical framework in which we work is set, so we are ready to present our results for the first type of processes that we will discuss - diffusion processes.

To repeat, we are interested in ergodicity of the solution of the following SDE

$$dX_t^x = b(X_t^x)dt + \sigma(X_t^x)dB_t\,, \quad X_0^x = x \in \mathbb{R}^d\,, \tag{2.0.1}$$

where $\{B_t\}_{t\geq 0}$ is a standard n-dimensional Brownian motion (defined on a stochastic basis $(\Omega, \mathscr{F}, \{\mathscr{F}_t\}_{t\geq 0}, \mathbb{P})$ satisfying the usual conditions), and the coefficients $b : \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times n}$ satisfy conditions **(C1)-(C3)** from the previous chapter.

As discussed before, under **(C1)-(C3)**, for any $x \in \mathbb{R}^d$, the stochastic differential equation in (2.0.1) admits a unique strong non-explosive solution $\{X_t^x\}_{t\geq 0}$ which is a strong Markov process with continuous sample paths and transition kernel $p(t,x,dy) = \mathbb{P}(X_t^x \in dy)$, $t \geq 0$, $x \in \mathbb{R}^d$. In the context of Markov processes, it is natural that the underlying probability measure depends on the initial conditions of the process. Using standard arguments (Kolmogorov extension theorem), it is well known that for each $x \in \mathbb{R}^d$ the above defined transition kernel defines a unique probability measure $\mathbb{P}^x$ on the canonical (sample-path) space such that the projection process, denoted by $\{X_t\}_{t\geq 0}$, is a strong Markov process (with respect to the completion of the corresponding natural filtration), it has continuous sample paths, and the same finite-dimensional distributions (with respect to $\mathbb{P}^x$) as $\{X_t^x\}_{t\geq 0}$ (with respect to $\mathbb{P}$). Since we are interested in distributional properties of the solution to (2.0.1) only, in the sequel we rather deal with $\{X_t\}_{t\geq 0}$ than with $\{X_t^x\}_{t\geq 0}$.

Testing the process (or a solution of the SDE) for ergodicity by checking whether conditions in the definition hold or not can be tricky - it would require of us to know the distribution (that is, the transition probability kernel) explicitly. However, the solutions of SDEs are rarely explicitly known. With the help of the Foster - Lypaunov method, this issue was reduced to checking a certain *drift condition* (i.e. (1.4.1)), which is given in terms of the extended generator of our process, together with some distributional properties of the process. Now, since the generator is given in terms of the coefficients of our process (the drift b and diffusive coefficient σ), which are known to us, we will be able to check this drift condition for a certain process. In this way, by assuming our process does possess those nice properties, we would be able to find sufficient conditions on the coefficients of the process that would ensure the ergodicity to hold. The second step could then be to check whether these conditions are sharp in the sense that we check whether they could be strengthened.

2.1. Literature overview

As stochastic models appear as appropriate solutions to model various phenomena in nature and everyday life, and their stability is one of the most interested features they possess, the research on this issue is vast and extensive. Here we present a brief overview of the results concerning ergodicity of diffusion processes with respect to the total variation norm and Wasserstein distances.

There are many books and articles that discuss ergodicity of diffusion processes. Depending on the type of the convergence we are interested in, there are more or less results regarding it.

First, if we consider the qualitative aspect of the problem, we can see that it is very well researched. For example:

- [MT93a], [MT93b]
  - ↬ With the assumption on open-set irreducibility and aperiodicity, they prove that ergodicity is equivalent to positive Harris-recurrence.
- [FR05]
  - ↬ Discussion on the ergodicity with respect to the structural properties of the process like positive recurrence and aperiodicity.

Secondly, from the quantitative aspect, we can divide results based on the obtained rate of the convergence. Most results are for the case of geometric ergodicity.

- **Geometric rate:** most results in this case.

  TV: Very well investigated situation. See [DMT95], [Wan08].

  WASS: This case is considered in recent years. See [Ebe11], [Ebe15], [LW16], [HMS11] (some contractivity results), [vRS05], [Wan16] or [Maj17].

- **Sub-geometric ergodicity:** there are results, but most of them are not optimal (we can find examples of processes that can not be covered by existing results in the literature).

  TV: In this case, the problem ergodicity has been considered in the literature (see [DFG09], [FR05], [Kul15], [Kul18], [San16a], [Ver97] and [Ver99]).

  WASS: This case was considered only recently, so there are only few results (see [DFM16], [But14]).

2.2. Irreducibility and Aperiodicity

After an brief overview of the existing results regarding the ergodicity of diffusion processes with and without jumps, we present our contribution to the topic. As our first step is to consider ergodicity with respect to the total variation distance, and we have discussed in the previous chapter that this will be possible only for processes that behave nicely in a certain way, we start the analysis by looking for sufficient conditions that would ensure these nice properties to hold and that would be easy to check (or at least easier than the properties themselves).

More precisely, we have seen that the secret, but crucial ingredients in the Foster - Lyapunov method that ensured the ergodic property were ϕ-irreducibility and aperiodicity. Both of those conditions require one to know the transition kernel, that is the distribution of the process, to be able to check them. As the explicit formula for the kernel is known in only few special cases, these conditions would in general be hard to verify. Since our problem is presented in the form of an SDE, the only information we do have about our process are its drift coefficient b and diffusive coefficient σ. Therefore, we turn our attention to finding conditions that would be given in terms of the coefficients of the process and would imply irreducibility and aperiodicity.

A usual sufficient conditions for open-irreducibility and aperiodicity that can be found in the literature are Lipschitz continuity and *uniform continuity* of the diffusion coefficient c. The **uniform ellipticity** property holds for c if there is $\gamma \geq 1$ such that

$$\gamma^{-1}|y|^2 \leq \langle y, c(x)y\rangle \leq \gamma|y|^2, \qquad y \in \mathbb{R}^d. \tag{2.2.1}$$

Further requirement is the *at most linear growth* of b and σ, that is, the existence of $\Gamma > 0$ such that

$$|b(x)|^2 + \|c(x)\|^2_{\mathrm{HS}} \leq \Gamma(1+|x|^2), \qquad x \in \mathbb{R}^d. \tag{2.2.2}$$

Now, assumptions (2.2.1) and (2.2.2), together with Lipschitz continuity of c and **(C1)**-**(C3)**, imply that

- X is a C_b-Feller and strong Feller process ([RW00, Theorem V.24.1] and [Dur96, Theorem 7.3.8]),

- X is open-set irreducible and aperiodic ([ST97, Remark 4.3])).

In the following theorem we show that $\{X_t\}_{t\geq 0}$ will be open-set irreducible and aperiodic if $b(x)$ and $c(x)$ are δ-Hölder continuous (for some $\delta > 0$) and $c(x)$ is uniformly elliptic on an open ball only (not he whole space), without assuming Lipschitz continuity of $c(x)$.

Theorem 2.2.1. Assume **(C1)-(C3)**. Further, assume that there are $x_0 \in \mathbb{R}^d$ and $r_0 > 0$, such that

(i) there are $\delta, \Gamma, \gamma > 0$, such that for all $x, y \in B_{r_0}(x_0)$ we have that

$$|b(x) - b(y)| + \|c(x) - c(y)\|_{\mathrm{HS}} \leq \Gamma|x-y|^{\delta} \quad \text{and} \quad \langle y, c(x)y\rangle \geq \gamma|y|^2\,;$$

(ii) $\mathbb{P}^x(\tau_{B_{r_0}(x_0)} < +\infty) > 0$ for all $x \in \mathbb{R}^d$, where $\tau_B := \inf\{t \geq 0 : X_t \in B\}$ is the first hitting time of a set $B \subseteq \mathbb{R}^d$.

Then, $\{X_t\}_{t\geq 0}$ is open-set irreducible and aperiodic.

Proof. Due to [Dur96, Theorems 7.3.6 and 7.3.7] there is a strictly positive function $q(t,x,y)$ on $(0,\infty) \times \bar{B}_{r_0}(x_0) \times \bar{B}_{r_0}(x_0)$, jointly continuous in t, x and y, and twice continuously differentiable in x on $B_{r_0}(x_0)$, satisfying

$$\mathbb{E}^x(f(X_t), \tau_{\bar{B}^c_{r_0}(x_0)} > t) = \int_{B_{r_0}(x_0)} q(t,x,y) f(y)\, dy\,, \qquad t > 0, x \in B_{r_0}(x_0), f \in C_b(\mathbb{R}^d)\,,$$

where $\tau_{\bar{B}^c_{r_0}(x_0)} := \inf\{t \geq 0 : X_t \in \bar{B}^c_{r_0}(x_0)\}$. Clearly, by employing dominated convergence theorem, the above relation holds also for $\mathbb{1}_O$, for any open set $O \subseteq B_{r_0}(x_0)$. Denote by $\mathscr{D}$ the class of all $B \in \mathscr{B}(B_{r_0}(x_0))$ (the Borel σ-algebra on $B_{r_0}(x_0)$) such that

$$\mathbb{P}^x(X_t \in B,\ \tau_{\bar{B}^c_{r_0}(x_0)} > t) = \int_B q(t,x,y)\, dy\,, \qquad t > 0, x \in B_{r_0}(x_0)\,.$$

Clearly, $\mathscr{D}$ contains the π-system of open rectangles in $B_{r_0}(x_0)$, and forms a λ-system. Hence, by employing Dynkin's π-λ theorem we conclude that $\mathscr{D} = \mathscr{B}(B_{r_0}(x_0))$. Consequently, for any $t > 0$, $x \in B_{r_0}(x_0)$ and $B \in \mathscr{B}(\mathbb{R}^d)$ we have that

$$p(t,x,B) \geq \int_{B \cap B_{r_0}(x_0)} q(t,x,y)\, dy\,.$$

Set now $\phi(\cdot) := \lambda(\cdot \cap B_{r_0}(x_0))$, where λ stands for the Lebesgue measure on $\mathbb{R}^d$. Then, ϕ is a σ-finite measure whose support has a non-empty interior.

Let us now show that $\{X_t\}_{t\geq 0}$ is ϕ-irreducible. Let $x \in B^c_{r_0}(x_0)$ (for $x \in B_{r_0}(x_0)$ the assertion is obvious) and $B \in \mathscr{B}(\mathbb{R}^d)$, $\phi(B) > 0$, be arbitrary. For all $s > 0$ we have

$$\begin{aligned}\int_0^\infty p(t,x,B)\,dt &\geq \int_s^\infty p(t,x,B)\,dt \\ &= \int_s^\infty \int_{\mathbb{R}^d} p(t-s,x,dy)p(s,y,B)\,dt \\ &\geq \int_s^\infty \int_{B_{r_0}(x_0)} p(t-s,x,dy)p(s,y,B)\,dt \\ &= \int_{B_{r_0}(x_0)} p(s,y,B) \int_s^\infty p(t-s,x,dy)\,dt\,,\end{aligned}$$

where in the second equality we used the Chapman-Kolmogorov equation.

The assertion now follows from the fact that $p(s,y,B) > 0$ for $y \in B_{r_0}(x_0)$, and

$$\int_s^\infty p(t-s,x,B_{r_0}(x_0))\,dt = \int_0^\infty p(t,x,B_{r_0}(x_0))\,dt = \mathbb{E}^x\left[\int_0^\infty \mathbb{1}_{\{X_t \in B_{r_0}(x_0)\}}\,dt\right] > 0\,,$$

since $\{X_t\}_{t\geq 0}$ has continuous sample paths, $B_{r_0}(x_0)$ is an open set and, by assumption, $\mathbb{P}^x(\tau_{B_{r_0}(x_0)} < \infty) > 0$ for every $x \in \mathbb{R}^d$.

Finally, let us prove that $\{X_t\}_{t\geq 0}$ is aperiodic. We show that

$$\sum_{n=1}^\infty p(n,x,B) > 0\,, \qquad x \in \mathbb{R}^d\,,$$

whenever $\phi(B) > 0$, $B \in \mathscr{B}(\mathbb{R}^d)$. Again, for $x \in B_{r_0}(x_0)$ the relation obviously holds. For $x \in B^c_{r_0}(x_0)$ and $B \in \mathscr{B}(\mathbb{R}^d)$, $\phi(B) > 0$, we have that

$$\sum_{n=1}^\infty p(n,x,B) \geq \int_{B_{r_0}(x_0)} \sum_{n=1}^\infty p(n-t,x,dy)\,p(t,y,B)\,, \qquad t \in (0,1)\,.$$

Since $p(t,y,B) > 0$ for $y \in B_{r_0}(x_0)$, it suffices to show that

$$\sum_{n=1}^\infty p(n-t,x,B_{r_0}(x_0)) \geq \mathbb{P}^x\left(\bigcup_{n=1}^\infty \{X_{n-t} \in B_{r_0}(x_0)\}\right) > 0$$

for some $t \in (0,1)$. Assume this is not the case, that is,

$$\mathbb{P}^x\left(\bigcup_{n=1}^\infty \{X_{n-t} \in B_{r_0}(x_0)\}\right) = 0\,, \qquad t \in (0,1)\,.$$

This, in particular, implies that

$$\mathbb{P}^x\left(\bigcup_{q\in\mathbb{Q}_+\setminus\mathbb{Z}_+} \{X_q \in B_{r_0}(x_0)\}\right) = 0\,,$$

which is impossible since $\{X_t\}_{t\geq 0}$ has continuous sample paths, $B_{r_0}(x_0)$ is an open set and $\mathbb{P}^x(\tau_{B_{r_0}(x_0)} < \infty) > 0$ for every $x \in \mathbb{R}^d$. Thus,

$$\sum_{n=1}^{\infty} p(n,x,B) > 0\,, \qquad x \in \mathbb{R}^d\,,$$

whenever $\phi(B) > 0$, which concludes the proof. ■

The second assumption in Theorem 2.2.1 again requires computing the distribution of the process X. So, we also provide a sufficient condition for this assumption to hold.

Proposition 2.2.2. Assume **(C1)-(C3)**. Then for any $x_0 \in \mathbb{R}^d$ and $r_0 > 0$, provided that $c(x)$ is positive definite for all $x \in \mathbb{R}^d$, $|x-x_0| \geq r_0$, it holds that

$$\mathbb{P}^x(\tau_{B_{r_0}(x_0)} < +\infty) > 0\,, \qquad x \in \mathbb{R}^d\,.$$

Proof. Let $0 < \varepsilon < r_0$, and let

$$\bar{\mathscr{V}}(r) := \int_{r_0-\varepsilon}^{r} \mathrm{e}^{-I_{x_0}(u)}du\,, \qquad r \geq r_0 - \varepsilon\,.$$

Then, for $r > r_0 - \varepsilon$ we have

$$\bar{\mathscr{V}}'(r) = \mathrm{e}^{-I_{x_0}(r)} > 0 \qquad \text{and} \qquad \bar{\mathscr{V}}''(r) = -\frac{\bar{\mathscr{V}}'(r)}{r}\imath_{x_0}(r)\,.$$

Further, let $\mathscr{V} : \mathbb{R}^d \to [0,\infty)$, $\mathscr{V} \in C^2(\mathbb{R}^d)$, be such that $\mathscr{V}(x) = \bar{\mathscr{V}}(|x-x_0|)$ for $x \in \mathbb{R}^d$, $|x-x_0| \geq r_0$. Now, for $x \in \mathbb{R}^d$, $|x-x_0| \geq r_0$, we have

$$\begin{aligned} 2\mathscr{L}\mathscr{V}(x) &= C_{x_0}(x)\bar{\mathscr{V}}''(|x-x_0|) + \frac{\bar{\mathscr{V}}'(|x-x_0|)}{|x-x_0|}(2A(x) - C_{x_0}(x) + 2B_{x_0}(x)) \\ &= \frac{\bar{\mathscr{V}}'(|x-x_0|)}{|x-x_0|}\left(2A(x) - C_{x_0}(x) + 2B_{x_0}(x) - C_{x_0}(x)\imath(|x-x_0|)\right) \\ &\leq 0\,. \end{aligned}$$

Further, as we have already discussed, for every $x \in \mathbb{R}^d$ the process

$$\mathscr{V}(X_t) - \mathscr{V}(X_0) - \int_0^t \mathscr{L}\mathscr{V}(X_s)\,ds\,, \qquad t \geq 0\,,$$

is a local $\mathbb{P}^x$-martingale. For $n \in \mathbb{N}$, define $\tau_n := \tau_{B_n^c(x_0)}$. Clearly, τ_n, $n \in \mathbb{N}$, are stopping times such that (due to non-explosivity of $\{X_t\}_{t\geq 0}$) $\tau_n \to \infty$ $\mathbb{P}^x$-a.s. as $n \to \infty$ for all $x \in \mathbb{R}^d$. Hence, the processes

$$\mathscr{V}(X_{t\wedge\tau_n}) - \mathscr{V}(X_0) - \int_0^{t\wedge\tau_n} \mathscr{L}\mathscr{V}(X_s)\,ds\,, \qquad t \geq 0, n \in \mathbb{N}\,,$$

are $\mathbb{P}^x$-martingales. Now, for $x \in \mathbb{R}^d$, $|x - x_0| \geq r_0$, we have

$$\begin{aligned} 2\mathbb{E}^x[\bar{\mathscr{V}}(|X_{t\wedge\tau_n\wedge\tau_{B_{r_0}(x_0)}} - x_0|)] - 2\bar{\mathscr{V}}(|x - x_0|) &= 2\mathbb{E}^x[\mathscr{V}(X_{t\wedge\tau_n\wedge\tau_{B_{r_0}(x_0)}})] - 2\mathbb{E}^x[\mathscr{V}(X_0)] \\ &= \mathbb{E}^x \int_0^{t\wedge\tau_n\wedge\tau_{B_{r_0}(x_0)}} 2\mathscr{L}\mathscr{V}(X_s)\, ds \\ &\leq 0\,, \end{aligned}$$

that is,

$$\mathbb{E}^x[\bar{\mathscr{V}}(|X_{t\wedge\tau_n\wedge\tau_{B_{r_0}(x_0)}} - x_0|)] \leq \bar{\mathscr{V}}(|x - x_0|)\,.$$

Thus,

$$\mathbb{E}^x[\bar{\mathscr{V}}(|X_{t\wedge\tau_n} - x_0|)\mathbb{1}_{\{\tau_{B_{r_0}(x_0)} > \tau_n\}}] \leq \bar{\mathscr{V}}(|x - x_0|)\,, \qquad x \in \mathbb{R}^d,\ |x - x_0| \geq r_0\,.$$

By letting $t \to \infty$ Fatou's lemma implies

$$\bar{\mathscr{V}}(n)\mathbb{P}^x(\tau_{B_{r_0}(x_0)} > \tau_n) \leq \bar{\mathscr{V}}(|x - x_0|)\,, \qquad x \in \mathbb{R}^d, |x - x_0| \geq r_0\,.$$

Consequently, by letting $n \to \infty$, we conclude

$$\mathbb{P}^x(\tau_{B_{r_0}(x_0)} = \infty) \leq \frac{\bar{\mathscr{V}}(|x - x_0|)}{\bar{\mathscr{V}}(\infty)} < 1\,, \qquad x \in \mathbb{R}^d, |x - x_0| \geq r_0\,,$$

that is, $\mathbb{P}^x(\tau_{B_{r_0}(x_0)} < \infty) > 0$ for all $x \in \mathbb{R}^d$. ■

2.3. ERGODICITY IN THE TOTAL VARIATION DISTANCE

We now continue by discussing the property of a stochastic process that we are most interested in: ergodicity. In this sequel we obtain sub-geometric ergodicity of classical diffusions and diffusions with jumps with respect to the total variation norm. Before stating the main results, we introduce some notation we need in the sequel.

Fix $x_0 \in \mathbb{R}^d$ and $r_0 \geq 0$, and put

$$\begin{aligned}
c(x) &:= \sigma(x)\sigma(x)^T\,,\\
A(x) &:= \frac{1}{2}\operatorname{Tr} c(x)\,, \qquad x \in \mathbb{R}^d\,,\\
B_{x_0}(x) &:= \langle x - x_0, b(x)\rangle\,, \qquad x \in \mathbb{R}^d\,,\\
C_{x_0}(x) &:= \frac{\langle x - x_0, c(x)(x - x_0)\rangle}{|x - x_0|^2}\,, \qquad x \in \mathbb{R}^d \setminus \{x_0\}\,,\\
\gamma_{x_0}(r) &:= \inf_{|x - x_0| = r} C_{x_0}(x)\,, \qquad r > 0\,,\\
\iota_{x_0}(r) &:= \sup_{|x - x_0| = r} \frac{2A(x) - C_{x_0}(x) + 2B_{x_0}(x)}{C_{x_0}(x)}\,, \qquad r > 0\,,\\
I_{x_0}(r) &:= \int_{r_0}^{r} \frac{\iota_{x_0}(s)}{s} ds\,, \qquad r \geq r_0\,.
\end{aligned}$$

The notation given here might seem complicated and puzzling. However, note that all functions are defined by coefficients b and σ, they are some transformation of the coefficients. Therefore, they somehow capture and describe the behaviour of our process. The relation (2.3.1) that we will introduce below will summarize all this information about the movement of the process in one single coefficient: Λ.

Theorem 2.3.1. Assume **(C1)-(C3)**, and assume that $\{X_t\}_{t\geq 0}$ is open-set irreducible and aperiodic. Further, let $\varphi : [1,\infty) \to (0,\infty)$ be a non-decreasing, differentiable and concave function satisfying $\lim_{t\to\infty} \varphi'(t) = 0$ and

$$\Lambda := \int_{r_0}^{\infty} \varphi\left(\int_{r_0}^{u} \mathrm{e}^{-I_{x_0}(v)} dv + 1\right) \frac{\mathrm{e}^{I_{x_0}(u)}}{\gamma_{x_0}(u)} du < \infty \tag{2.3.1}$$

for some $x_0 \in \mathbb{R}^d$ and $r_0 \geq 0$, and assume that $c(x)$ is positive definite for all $x \in \mathbb{R}^d$, $|x - x_0| \geq r_0$ (hence, the above functions and the relation in (2.3.1) are well defined).

Then, $\{X_t\}_{t\geq 0}$ admits a unique invariant $\pi \in \mathscr{P}$ satisfying

$$\lim_{t\to\infty} \varphi(\Phi^{-1}(t))\|\delta_x P_t - \pi\|_{\mathrm{TV}} = 0\,, \qquad x \in \mathbb{R}^d\,,$$

where

$$\Phi(t) := \int_1^t \frac{ds}{\varphi(s)}\,, \qquad t \geq 1\,.$$

Proof. ∗*Idea*∗ We prove this result by using the Foster - Lyapunov method explained in the previous chapter. There are two steps in this method:

- finding appropriate function $\mathscr{V}$ and petite set C so that the drift-inequality (1.4.1) holds,
- applying Theorem 1.4.6 that would provide existence of the invariant probability measure π and establish the convergence of the marginal distributions of the process to π in total variation norm.

• Set $\varphi_\Lambda(t) = \varphi(t)/\Lambda$, where Λ is given in (2.3.1), and observe that $\varphi_\Lambda(t)$ has the same properties as $\varphi(t)$. Next, define

$$\bar{\mathscr{V}}(r) := \int_{r_0}^r \mathrm{e}^{-I_{x_0}(u)} \int_u^\infty \varphi_\Lambda\left(\int_{r_0}^v \mathrm{e}^{-I_{x_0}(w)}dw + 1\right)\frac{\mathrm{e}^{I_{x_0}(v)}}{\gamma_{x_0}(v)}dv\,du\,, \qquad r \geq r_0\,.$$

Clearly, for $r \geq r_0$ it holds that

$$\bar{\mathscr{V}}(r) \leq \int_{r_0}^r \mathrm{e}^{-I_{x_0}(u)}du\,, \tag{2.3.2}$$

and

$$\bar{\mathscr{V}}'(r) = \mathrm{e}^{-I_{x_0}(r)} \int_r^\infty \varphi_\Lambda\left(\int_{r_0}^u \mathrm{e}^{-I_{x_0}(v)}dv + 1\right)\frac{\mathrm{e}^{I_{x_0}(u)}}{\gamma_{x_0}(u)}du$$

$$\bar{\mathscr{V}}''(r) = -\frac{\iota_{x_0}(r)}{r}\mathrm{e}^{-I_{x_0}(r)} \int_r^\infty \varphi_\Lambda\left(\int_{r_0}^u \mathrm{e}^{-I_{x_0}(v)}dv + 1\right)\frac{\mathrm{e}^{I_{x_0}(u)}}{\gamma_{x_0}(u)}du - \frac{\varphi_\Lambda\left(\int_{r_0}^r \mathrm{e}^{-I_{x_0}(u)}du + 1\right)}{\gamma_{x_0}(r)}\,.$$

Further, fix $r_1 > r_0$ and let $\mathscr{V} : \mathbb{R}^d \to [0, +\infty\rangle$, $\mathscr{V} \in C^2(\mathbb{R}^d)$, be such that $\mathscr{V}(x) = \bar{\mathscr{V}}(|x - x_0|) + 1$ for $x \in \mathbb{R}^d$, $|x - x_0| \geq r_1$. Now, for $x \in \mathbb{R}^d$, $|x - x_0| \geq r_1$, we have

$$\begin{aligned}\mathscr{L}\mathscr{V}(x) &= \frac{1}{2}C_{x_0}(x)\bar{\mathscr{V}}''(|x-x_0|) + \frac{\bar{\mathscr{V}}'(|x-x_0|)}{2|x-x_0|}(2A(x) - C_{x_0}(x) + 2B_{x_0}(x))\\ &\leq -\frac{1}{2}\varphi_\Lambda\left(\int_{r_0}^{|x-x_0|} \mathrm{e}^{-I_{x_0}(u)}du + 1\right)\end{aligned}$$

$$\leq -\frac{1}{2}\varphi_\Lambda(\mathscr{V}(x)),$$

where in the final step we employed the fact that $\varphi(t)$ (that is, $\varphi_\Lambda(t)$) is non-decreasing and (2.3.2).

Define $C = \bar{B}_{r_1}(x_0)$ (the topological closure of the open ball $B_{r_1}(x_0)$). It is a closed set, so ψ-irreducibility and C_b-Feller property imply that it is also petite. Also, $\sup_C \mathscr{V} < +\infty$ because $\mathscr{V}$ is a continuous function.

Moreover, we have obtained the relation in (3.11) in [DFG09, Theorem 3.4 (i)] with $\phi(t) = \varphi_\Lambda(t)$, C, and $b = \sup_{x\in C} |\mathscr{L}V(x)|$. Now, [Twe94, Theorems 5.1 and 7.1], together with open-set irreducibility, aperiodicity and C_b-Feller property of $\{X_t\}_{t\geq 0}$, imply that $\{X_t\}_{t\geq 0}$ meets the conditions of [DFG09, Theorem 3.2] with $\Psi_1(t) = t$ and $\Psi_2(t) = 1$, which concludes the proof. ■

- *Remarks:* as mentioned before, the drift condition implies the existence of a unique invariant probability measure.

- **(C1)-(C3)** imply existence of the solution that is a time-homogeneous and non-explosive strong Markov process that also satisfies the C_b-Feller property.

- The drift inequality (for a function φ) together with aperiodicity imply sub-geometric ergodicity. Namely, as mentioned before, the sub-geometric ergodicity with rate $r(t) = \varphi(\Phi^{-1}(t))$ comes from finiteness of the $\varphi \circ \Phi^{-1}$-moment of δ-hitting time of petite set C, which we denote by τ_C^δ, and aperiodicity, which says that the process cannot exhibit any type of cyclic behaviour ([FR05, Theorem 1]). Now, since ϕ-irreducibility says that we can cover the state space with the countable union of petite sets, we obtain the finiteness of the $\varphi \circ \Phi^{-1}$-moment of δ-hitting time of any petite set, and thus conclude sub-geometric ergodicity.

- In our case, the role of petite set C is played by a closed ball around the origin with large enough radius. Thus, in view of the previous remark, to obtain sub-geometric ergodicity we need to control $\varphi \circ \Phi^{-1}$-moment of its δ-hitting time τ_C^δ by function $\mathscr{V}$. We do this through the relation (2.3.1), since this relation is crucial in the construction of the appropriate Lyapunov function $\mathscr{V}$ (actually, it appears as a part of it).

- Observe that in the proof of Theorem 2.3.1 we did not use the fact that $\{X_t\}_{t\geq 0}$ is a unique strong solution to (2.0.1). All that we needed is that the martingale problem for (b,c) is well posed, which is equivalent to that (2.0.1) admits a unique (in distribution) weak solution (see [RW00, Theorem V.20.1]).

- Even though the steps of the proof are precise and clear, note that the appropriate Lyapunov function was presented out of the blue, without any clear motivation. It is shown that it satisfies all conditions required and that it works for our purpose, but it is not clear how one comes up with this formula. Actually, versions of this formula did appear in the literature before and were used to obtain various types of drift inequalities: $\mathscr{L}\mathscr{V} \leq 0$ (which implies recurrency), $\mathscr{L}\mathscr{V} \geq 0$ (which implies transience), $\mathscr{L}\mathscr{V} \leq -\mathscr{V}$ (to obtain geometric rate of the convergence) and so on (see [Dur96, Section 6.6.4], [Bha78, Theorem 3.3], [MT93b, Theorem 9.1]). The idea comes from solving a partial differential equation of order 2. For example, take a solution to the SDE X and calculate its extended generator $\mathscr{L}$ - it is a second-order partial differential operator. If f is twice-differentiable function, it is in the domain of the extended generator and $f(X_t) - f(X_0) - \int_0^t \mathscr{L}f(X_s)\,ds$ is a local martingale. If f further satisfies $\mathscr{L}f = 0$, then we know that $Y_t := f(X_t)$ is a local martingale, it solves a martingale problem for drift coefficient 0 and some diffusive coefficients h, so it follows that we can calculate Y_t by time changing Brownian motion. Then, X_t can be obtained as $X_t := f^{-1}(Y_t)$ and it will have some nice properties like recurrence. Similarly, if we are interested in geometric ergodicity, we need to solve the partial differential equation (in short, PDE) $\mathscr{L}f = -f$. By using techniques of PDEs, one can obtain the solution in the form of the integral, as we have in our proof.

From Theorem 2.3.1 we see that the rate of the convergence is equal to

$$r(t) = \varphi(\Phi^{-1}(t)).$$

So, by changing the the function φ, we get different rates: the convergence can be slower or faster. As there are various similar results concerning sub-geometric ergodicity, our result might coincide with some of those in some special cases. As mentioned before, it is important to position our result among existing ones and, also, check whether it is sharp enough. In this case, it turns out that our result can be related to some other similar ones, but some of them are far from being optimal and our result improves them. Here is a list of some special cases:

(i) $\varphi \equiv 1 \implies r(t) = 1$, if $\int_{r_0}^{\infty} \frac{e^{I_{x_0}(u)}}{\gamma_{x_0}(u)} du < +\infty$

- The same would follow for any bounded function φ.
- This is exactly the condition for (strong) ergodicity established in the paper [Bha78, Theorem 3.5].
- In one-dimensional case, similar results can be found in [Wan08, Theorem 1.2] and [Man68, Chapter IV].

(ii) $\varphi(t) = t \implies$ assumptions of the Theorem 2.3.1 not satisfied since $\lim_{t\to+\infty} \varphi'(t) = 1 \neq 0$.

- For such φ, in one-dimensional case exponential ergodicity was proven in [Wan08, Theorem 1.3] under (2.3.1) (with $\varphi(t) = t$). In Proposition 2.3.6 we extend this to the multi-dimensional case.

(iii) $\varphi(t) = t^{\alpha}$ for $0 < \alpha < 1 \implies r(t) = t^{\alpha/(1-\alpha)}$ (if the corresponding $\Lambda < +\infty$)

- Condition for sub-geometric ergodicity, precisely, the polynomial ergodicity.
- Similar results do exist already, see for example [DFG09, Theorem 5.4], [FR05, page 1581], [Kul15, Theorem 1.30], [Kul18, Theorem 3.3.6], [San16a, Theorem 3.3 (iv)], [Ver97, Theorem 6] and [Ver99, Theorem 6]. But they are not sharp enough, as we discuss below.

The case of a polynomial rate that follows directly from Theorem 2.3.1 can be related to [San16a, Theorem 3.3 (iv)]. There, the author has shown that $\{X_t\}_{t\geq 0}$ will be sub-geometrically ergodic with rate $t^{\alpha/(1-\alpha)}$ (that is, $\varphi(t) = t^{\alpha}$), $\alpha \in (0,1)$, if there exist $\gamma > 0$, $\Gamma > 0$ and $r_0 \geq 0$, such that

$$A(x) - \left(1 - \frac{\gamma}{2}\right) C_0(x) + B_0(x) \leq -\Gamma |x|^{\gamma\alpha-\gamma+2}, \qquad |x| \geq r_0. \tag{2.3.3}$$

However, this results is not sufficiently sharp as there are examples that cannot be covered by it, and they are covered by our result. Namely, in Proposition 2.3.3 we show that (2.3.3) implies (2.3.1), and in Example 2.3.2 we give an example of a diffusion process satisfying conditions from Theorem 2.3.1, but not the condition in (2.3.3).

Example 2.3.2. Let $\sigma(x)\equiv 1$, and let $b(x)$ be locally Lipschitz continuous and such that $b(x)=-\mathrm{sgn}(x)(\cos x+1)$ for all $|x|$ large enough, where

$$\mathrm{sgn}(x) := \begin{cases} 1, & x\geq 0\,, \\ -1, & x<0\,. \end{cases}$$

Clearly, $b(x)$ and $\sigma(x)$ satisfy **(C1)-(C3)** and define, through (2.0.1), an open-set irreducible and aperiodic diffusion process $\{X_t\}_{t\geq 0}$. The condition in (2.3.1) now reduces to showing that there is $r_0\geq 0$ such that

$$\int_{r_0}^{\infty}\left(\int_{r_0}^{u}\mathrm{e}^{2\sin v+2v}+1\right)^{\alpha}\mathrm{e}^{-2\sin u-2u}du\,<\,\infty\,,$$

which can be obviously obtained for any $0<\alpha<1$. On the other hand, the condition in (2.3.3) is equivalent to showing that there are $\gamma>0$, $\Gamma>0$ and $r_0\geq 0$, such that

$$\frac{\gamma-1}{2}-x\,\mathrm{sgn}(x)(\cos x+1)\;\leq\;-\Gamma|x|^{\gamma\alpha-\gamma+2}\,,\qquad |x|\geq r_0\,.$$

However, observe that in the points of the form $x=(2k+1)\pi$, $k\in\mathbb{Z}$, the second term on the left-hand side in the above inequality vanishes. Thus, we conclude that it is necessary that $0<\gamma<1$ and $\gamma\alpha-\gamma+2<0$, which is impossible. Note also that if we take $b(x)$ to be locally Lipschitz continuous and such that $b(x)=-\mathrm{sgn}(x)(\cos x+\rho)$ for all $|x|$ large enough, where $\rho>0$, then we again easily conclude that (2.3.1) holds for any $0<\alpha<1$. On the other hand, by the same reasoning as above, (2.3.3) can never hold. Observe that for $0<\rho<1$ the drift function generates a region in which the process is "pushed towards infinity" (set of points for which $\mathrm{sgn}(x)b(x)>0$). The condition in (2.3.1) says that this region is small compared to the region in which the process is "pushed towards the center of the state space" (set of points for which $\mathrm{sgn}(x)b(x)<0$) and which is responsible for the ergodic behavior.

Proposition 2.3.3. Assume **(C1)-(C3)**. Further, assume that $\gamma<2/(1-\alpha)$ and there are $r_0\geq 0$ and $\Delta\geq 1$, such that $\Delta^{-1}\leq C_0(x)\leq\Delta$ for all $|x|\geq r_0$. Then, (2.3.1) (with $x_0=0$) is a consequence of (2.3.3).

Proof. We have that

$$\iota_0(r) = \sup_{|x|=r} \frac{2\left(A(x) - \left(1-\frac{\gamma}{2}\right)C_0(x) + B_0(x)\right) + (1-\gamma)C_0(x)}{C_0(x)} \le -\frac{2\Gamma}{\Delta} r^{\gamma\alpha-\gamma+2} + 1 - \gamma$$

for all $r \ge r_1$, for some $r_1 \ge r_0$ large enough. Thus, there are $\Gamma_1 > 0$ and $r_2 \ge r_1$, such that

$$\iota_0(r) \le -\Gamma_1 r^{\gamma\alpha-\gamma+2}, \qquad r \ge r_2 .$$

This automatically implies that there are $\Gamma_2 > 0$ and $r_3 \ge r_2$, such that

$$I_0(r) \le -\Gamma_2 r^{\gamma\alpha-\gamma+2}, \qquad r \ge r_3 .$$

Now, by employing L'Hospital's rule (here we use the assumption $\gamma < 2/(1-\alpha)$), we have that

$$\lim_{u\to\infty} \frac{\left(\int_{r_3}^{u} \mathrm{e}^{-I_0(v)} dv + 1\right)}{\mathrm{e}^{-I_0(u)}} = 0 .$$

Hence, there is $r_4 \ge r_3$ such that

$$\int_{r_3}^{u} \mathrm{e}^{-I_0(v)} dv + 1 \le \mathrm{e}^{-I_0(u)} \qquad u \ge r_4 .$$

Finally, we conclude

$$\int_{r_4}^{\infty} \left(\int_{r_4}^{u} \mathrm{e}^{-I_0(v)} dv + 1\right)^{\alpha} \mathrm{e}^{I_0(u)} du \le \int_{r_4}^{\infty} \mathrm{e}^{(1-\alpha)I_0(u)} du < \infty,$$

which proves the assertion. ∎

Condition in (2.3.1) might be hard to check since the formula is somewhat complicated. Sometimes, another version of this condition might be easier to verify. In order to obtain it, we first present an auxiliary result (the following Proposition 2.3.4, which actually generalizes [Che00, Lemma 1.2] to the sub-geometric case) and then state the sufficient condition in Corollary 2.3.5.

Proposition 2.3.4. Let $c \ge 0$, and let $\rho(t)$ be a non-negative and non-decreasing differentiable function defined on $[0,\infty)$. Further, let $f(r)$ and $g(r)$ be non-negative Borel measurable functions, also defined on $[0,\infty)$, satisfying

$$\Delta := \sup_{r\ge r_0} \rho\left(\int_{r_0}^{r} g(u)du + c\right)^{1+\beta} \int_{r}^{\infty} f(u)\,du < \infty \tag{2.3.4}$$

for some $r_0 \ge 0$ and $\beta \ge 0$. Then,

(i) if $\beta > 0$,

$$\int_r^\infty \rho\left(\int_{r_0}^u g(v)\,dv + c\right) f(u)\,du \le \frac{\Delta(1+\beta)}{\beta}\rho\left(\int_{r_0}^r g(u)\,du + c\right)^{-\beta}, \qquad r \ge r_0 .$$

(ii) if $\beta = 0$, and $\int_{r_0}^\infty g(r)\,dr < \infty$ or $\rho(t)$ is bounded,

$$\int_r^\infty \rho\left(\int_{r_0}^u g(v)\,dv + c\right) f(u)\,du \le \Delta + \Delta \ln \frac{\rho\left(\int_{r_0}^\infty g(u)\,du + c\right)}{\rho\left(\int_{r_0}^r g(u)\,du + c\right)}, \qquad r \ge r_0 .$$

Proof. Set $F(r) = \int_r^\infty f(u)\,du$, $r \ge r_0$. Then, by assumption,

$$F(r) \le \Delta \rho\left(\int_{r_0}^r g(u)\,du + c\right)^{-1-\beta}, \qquad r \ge r_0 .$$

Consequently, for $r \ge r_0$, we have that

$$\begin{aligned}
&\int_r^\infty \rho\left(\int_{r_0}^u g(v)\,dv + c\right) f(u)\,du \\
&= -\int_r^\infty \rho\left(\int_{r_0}^u g(v)\,dv + c\right) dF(u) \\
&\le \rho\left(\int_{r_0}^r g(u)\,du + c\right) F(r) + \int_r^\infty \rho'\left(\int_{r_0}^u g(v)\,dv + c\right) g(u) F(u)\,du \\
&\le \Delta\rho\left(\int_{r_0}^r g(u)\,du + c\right)^{-\beta} + \Delta \int_r^\infty \rho'\left(\int_{r_0}^u g(v)\,dv + c\right) g(u) \rho\left(\int_{r_0}^u g(v)\,dv + c\right)^{-1-\beta} du .
\end{aligned}$$

Now, under the assumption in (i) we have that

$$\begin{aligned}
&\int_r^\infty \rho\left(\int_{r_0}^u g(v)\,dv + c\right) f(u)\,du \\
&\le \Delta\rho\left(\int_{r_0}^r g(u)\,du + c\right)^{-\beta} - \frac{\Delta}{\beta}\int_r^\infty d\rho\left(\int_{r_0}^u g(v)\,dv + c\right)^{-\beta} \\
&\le \Delta\rho\left(\int_{r_0}^r g(u)\,du + c\right)^{-\beta} + \frac{\Delta}{\beta}\rho\left(\int_{r_0}^r g(u)\,du + c\right)^{-\beta} \\
&= \frac{\Delta(1+\beta)}{\beta}\rho\left(\int_{r_0}^r g(u)\,du + c\right)^{-\beta},
\end{aligned}$$

where in the second step we employed integration by parts formula. On the other hand, under the assumptions in (ii),

$$\begin{aligned}
\int_r^\infty \rho\left(\int_{r_0}^u g(v)\,dv + c\right) f(u)\,du &\le \Delta + \Delta\int_r^\infty d\ln\left(\rho\left(\int_{r_0}^u g(v)\,dv + c\right)\right) \\
&= \Delta + \Delta \ln \frac{\rho\left(\int_{r_0}^\infty g(u)\,du + c\right)}{\rho\left(\int_{r_0}^r g(u)\,du + c\right)},
\end{aligned}$$

which concludes the proof. ■

With the help of Proposition 2.3.4, we now have a sufficient condition for the relation (2.3.1).

Corollary 2.3.5. (2.3.1) holds true if for some $\beta > 0$:

$$\sup_{r\geq r_0} \varphi\left(\int_{r_0}^{r} \mathrm{e}^{-I_{x_0}(u)}du+1\right)^{1+\beta}\int_{r}^{\infty}\frac{\mathrm{e}^{I_{x_0}(u)}}{\gamma_{x_0}(u)}du < +\infty.$$

As a final discussion regarding continuous diffusion, we present the extension of the result from J. Wang from 2008 to the multi-dimensional process. We obtain the geometric ergodicity for function φ that do not satisfy condition $\lim_{t\to\infty}\varphi'(t)\neq 0$.

Proposition 2.3.6. If in Theorem 2.3.1 $\liminf_{t\to\infty}\varphi'(t) > 0$, then $\{X_t\}_{t\geq 0}$ is geometrically ergodic.

Proof. First, observe that since $\varphi(t)$ is differentiable and concave, $t\mapsto\varphi'(t)$ is non-increasing. Thus, since $\varphi(t)$ is also non-decreasing, there are constants $\Gamma\geq\gamma>0$ such that

$$\gamma t-\gamma+\varphi(1)\leq\varphi(t)\leq\Gamma t-\Gamma+\varphi(1)\,,\qquad t\geq 1\,.$$

Consequently, the condition in (2.3.1) is equivalent to

$$\int_{r_0}^{\infty}\left(\int_{r_0}^{u}\mathrm{e}^{-I_{x_0}(v)}dv+1\right)\frac{\mathrm{e}^{I_{x_0}(u)}}{\gamma_{x_0}(u)}du<\infty$$

(recall that $\varphi(1)>0$). Denote this constant again by Λ. Analogously as in the proof of Theorem 2.3.1, let

$$\bar{\mathscr{V}}(r):=\frac{1}{\Lambda}\int_{r_0}^{r}\mathrm{e}^{-I_{x_0}(u)}\int_{u}^{\infty}\left(\int_{r_0}^{v}\mathrm{e}^{-I_{x_0}(w)}dw+1\right)\frac{\mathrm{e}^{I_{x_0}(v)}}{\gamma_{x_0}(v)}dv\,du\,,\qquad r\geq r_0\,,$$

and, for arbitrary but fixed $r_1>r_0$, let $\mathscr{V}:\mathbb{R}^d\to[0,\infty)$, $\mathscr{V}\in C^2(\mathbb{R}^d)$, be such that $\mathscr{V}(x)=\bar{\mathscr{V}}(|x-x_0|)+1$ for $x\in\mathbb{R}^d$, $|x-x_0|\geq r_1$. Then, for all $x\in\mathbb{R}^d$, $|x-x_0|\geq r_1$, it holds that

$$\mathscr{L}\mathscr{V}(x)\leq-\frac{1}{2\Lambda}\mathscr{V}(x)\,,\tag{2.3.5}$$

which is exactly the Lyapunov equation on [MT93b, page 529] with $c=1/2\Lambda$, $f(x)=\mathscr{V}(x)$, $C=\bar{B}_{r_1}(x_0)$ and $b=\sup_{x\in C}|\mathscr{L}V(x)|$. The fact that C is a petite set follows from [Twe94, Theorems 5.1 and 7.1], together with open-set irreducibility and C_b-Feller property of $\{X_t\}_{t\geq 0}$. Next, from [MT93a, Proposition 6.1], [MT93b, Theorem 4.2] and aperiodicity it follows now that the are a petite set $\mathscr{C} \in \mathscr{B}(\mathbb{R}^d)$, $T > 0$ and a non-trivial measure $\nu_{\mathscr{C}}$ on $\mathscr{B}(\mathbb{R}^d)$, such that $\nu_{\mathscr{C}}(\mathscr{C}) > 0$ and

$$p(t,x,B) \geq \nu_{\mathscr{C}}(B)\,, \qquad x \in \mathscr{C}\,,\; t \geq T\,,\; B \in \mathscr{B}(\mathbb{R}^d)\,.$$

In particular,

$$p(t,x,\mathscr{C}) > 0\,, \qquad x \in \mathscr{C}\,,\; t \geq T\,,$$

which is exactly the definition of aperiodicity used on [DMT95, page 1675]. Finally, observe that (2.3.5) is also the Lyapunov equation used on [DMT95, page 1679] with $c = 1/2\Lambda$, $C = \bar{B}_{r_1}(x_0)$ and $b = \sup_{x \in C} |\mathscr{L}V(x)|$. The assertion now follows from [DMT95, Theorem 5.2]. ■

2.3.1. Diffusion processes with jumps

Whats happens with ergodicity when we add jumps to our process? After establishing results for classical diffusions, which are continuous solutions of the SDE with coefficients b and σ, this question naturally appears. In order to see whether we can extend our results to this case, we first need to define a jump-process. Actually, we can create a jump-process in many ways, depending on the times of jumps and their size. We will mention two types of jump-diffusions. In both cases, the results for ergodicity will come up as an application of Theorem 2.3.1.

Recall that one way of defining a process is by through the martingale approach, that is, by setting its extended generator. So, in the first case, we create jumps by adding a "jump measure" to the generator of our process. This measure will determine the probability of any jump size, depending on the position of the process at the time of the jump. We call it a Lévy measure, as it will correspond to a jump measure of some Lévy process with jumps. Hence, we consider a jump-diffusion processes generated by the operator of the form

$$\begin{aligned} \mathscr{L}f(x) = {} & \langle b(x), \nabla f(x) \rangle + \frac{1}{2} \mathrm{Tr}\, c(x) \nabla^2 f(x) \\ & + \int_{\mathbb{R}^d} \Big(f(y+x) - f(x) - \langle y, \nabla f(x) \rangle \mathbb{1}_{B_1(0)}(y) \Big)\, \nu(x,dy)\,, \end{aligned} \tag{2.3.6}$$

where $b(x)$ is an $\mathbb{R}^d$-valued Borel measurable function, $c(x)$ is a symmetric non-negative definite $d \times d$ matrix-valued Borel measurable function, and $\nu(x,dy)$ is a non-negative

Borel kernel on $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$ (or a Radon measure on $\mathbb{R}^d \times \{0\}$), called the Lévy kernel, satisfying

$$\nu(x,\{0\}) = 0\,, \quad \text{and} \quad \int_{\mathbb{R}^d} (1 \wedge |y|^2)\nu(x,dy) < \infty\,, \qquad x \in \mathbb{R}^d\,.$$

↬ *Remark:* if $\nu(x,dy)$ is a null-measure, then $\mathscr{L}$ becomes a diffusion operator.

As before, we will need certain assumptions that will ensure the process is regular enough to be a candidate for an ergodic process with respect to the total variation norm. So, assume that

(A1) there is a càdlàg Markov process $(\Omega, \mathscr{F}, \{\mathscr{F}_t\}_{t\geq 0}, \{\theta_t\}_{t\geq 0}, \{X_t\}_{t\geq 0}, \{\mathbb{P}^x\}_{x\in\mathbb{R}^d})$, denoted by $\{X_t\}_{t\geq 0}$ in the sequel, which we call jump-diffusion process, such that for every $f \in C^2(\mathbb{R}^d)$ the process

$$f(X_t) - f(X_0) - \int_0^t \mathscr{L} f(X_s)\,ds\,, \qquad t \geq 0\,,$$

is a $\mathbb{P}^x$ local martingale for all $x \in \mathbb{R}^d$ under the natural filtration;

(A2) the process $\{X_t\}_{t\geq 0}$ satisfies the C_b-Feller property;

(A3) the process $\{X_t\}_{t\geq 0}$ is open-set irreducible and aperiodic;

(A4) there is $\rho > 0$ such that $\nu(x, B^c_{|x|}(-x)) = 0$ and $\int_{B_1(0)} |y|\,\nu(x,dy) < \infty$ for all $x \in \mathbb{R}^d$, $|x| \geq \rho$;

(A5) the functions $b(x)$, $c(x)$ and $x \mapsto \int_{B_1(0)} y\,\nu(x,dy)$ are continuous on $B^c_\rho(0)$.

- $C^2_b(\mathbb{R}^d)$ denotes the space of twice continuously differentiable functions with bounded derivatives.

↬ *Remarks:* **(A1)** always holds for the infinitesimal generator $(\mathscr{A}, \mathscr{D}_{\mathscr{A}})$ of $\{X_t\}_{t\geq 0}$ (see [EK86, Theorem 2.2.13 and Proposition 4.1.7]).

↬ Sufficient conditions, in terms of $b(x)$, $c(x)$ and $\nu(x,dy)$, ensuring **(A1)** and **(A2)** are (see [BSW13, Theorems 2.37, 3.23, 3.24, 3.25] and [Str75, Remark after Theorem 4.3]): $b(x)$ is continuous and bounded, $c(x)$ continuous, bounded and positive definite, $x \mapsto \int_B (1 \wedge |y|^2)\nu(x,dy)$ continuous and bounded for any $B \in \mathscr{B}(\mathbb{R}^d)$, and

$$(x,\xi) \mapsto i\langle \xi, b(x)\rangle + \frac{1}{2}\langle \xi, c(x)\xi\rangle + \int_{\mathbb{R}^d} \left(1 - e^{i\langle \xi, y\rangle} + i\langle \xi, y\rangle \mathbb{1}_{B_1(0)}(y)\right)\nu(x,dy)$$

is continuous. Namely, this condition imply that

(i) there is a unique non-explosive strong Markov process $\{X_t\}_{t\geq 0}$ with infinitesimal generator $(\mathscr{A}, \mathscr{D}_{\mathscr{A}})$ such that $C_c^\infty(\mathbb{R}^d) \subseteq \mathscr{D}_{\mathscr{A}}$, and $\mathscr{A}|_{C_c^\infty(\mathbb{R}^d)}$ takes the form in (2.3.6), where $C_c^\infty(\mathbb{R}^d)$ stands for the space of smooth functions with compact support;

(ii) the operator $\mathscr{L} := \mathscr{A}|_{C_c^\infty(\mathbb{R}^d)}$ satisfies **(A1)**;

(iii) the semigroup of $\{X_t\}_{t\geq 0}$ satisfies the Feller and strong Feller property ,

- There are many conditions ensuring **(A3)**, that is, open-set irreducibility and aperiodicity of jump-diffusion processes. References, depending on the type of the process, are:
  - the so-called stable-like processes: [Kol00] and [Kol11],
  - jump-diffusion processes with bounded coefficients: [KS12], [KS13], [KC99], [PS16, Remark 3.3] [San16b, Theorem 2.6] and [Str75],
  - special case of a class of jump-diffusion processes obtained as a solution to certain jump-type SDEs: [APS19], [BC86], [Ish01], [KK18], [Mas07, Mas09] and [Pic96, Pic10],
  - [Twe94, Theorem 3.2]: open-set irreducibility and aperiodicity are proven in the case the process X is strong Feller (actually it suffices to assume that $\{X_t\}_{t\geq 0}$ is a T-model in the sense of [Twe94], which is a certain weak version of the strong Feller property) and $\mathbb{P}^x(X_t \in O) > 0$ for every $t > 0$, $x \in \mathbb{R}^d$ and non-empty open set $O \subseteq \mathbb{R}^d$.
- **(A4)** means that when $\{X_t\}_{t\geq 0}$ is far away from the center of the state space, it admits bounded jumps only, with maximal intensity equal twice the distance to the origin. Also, with each jump, it comes closer to the center of the state space.

We are now ready to state the first result regarding these processes. We will use the same notation as in Theorem 2.3.1, with the exception of

$$B_{x_0}(x) := \left\langle x - x_0, b(x) - \int_{B_1(0)} y\,\nu(x,dy) \right\rangle, \qquad x \in \mathbb{R}^d.$$

Theorem 2.3.7. Let $\{X_t\}_{t\geq 0}$ be an open-set irreducible and aperiodic jump-diffusion process with coefficients $b(x)$, $c(x)$ and $\nu(x,dy)$, satisfying **(A1)-(A5)**. Further, let $\varphi : [1,\infty) \longrightarrow (0,\infty)$ be a non-decreasing, differentiable and concave function satisfying $\lim_{t\to\infty}\varphi'(t) = 0$ and the relation in (2.3.1) for some $x_0 \in \mathbb{R}^d$ and $r_0 \geq \rho + |x_0|$, and assume that $c(x)$ is positive definite for all $x \in \mathbb{R}^d$, $|x-x_0| \geq r_0$. Then, $\{X_t\}_{t\geq 0}$ admits a unique invariant $\pi \in \mathscr{P}$ such that

$$\lim_{t\to\infty} \varphi(\Phi^{-1}(t))\|\delta_x P_t - \pi\|_{\mathrm{TV}} = 0\,, \qquad x \in \mathbb{R}^d\,,$$

where $\Phi(t)$ is as in Theorem 2.3.1.

Proof. We proceed as in the proof of Theorem 2.3.1. Define

$$\bar{\mathscr{V}}(r) := \int_{r_0}^{r} \mathrm{e}^{-I_{x_0}(u)} \int_u^\infty \varphi_\Lambda\left(\int_{r_0}^{v} \mathrm{e}^{-I_{x_0}(w)}dw + 1\right) \frac{\mathrm{e}^{I_{x_0}(v)}}{\gamma_{x_0}(v)} dv\, du\,, \qquad r \geq r_0\,,$$

where $\varphi_\Lambda(t) = \varphi(t)/\Lambda$. Clearly,

$$\bar{\mathscr{V}}(r) \leq \int_{r_0}^{r} \mathrm{e}^{-I_{x_0}(u)} du\,, \qquad r \geq r_0\,, \tag{2.3.7}$$

and, because of **(A5)**, $\bar{\mathscr{V}}(r)$ is twice continuously differentiable on (r_0,∞). Further, for arbitrary, but fixed, $r_1 > r_0$ let $\tilde{\mathscr{V}} : [0,\infty) \to [0,\infty)$ be non-decreasing on $[0,\infty)$, $\tilde{\mathscr{V}}(r) = \bar{\mathscr{V}}(r)$ on $[r_1,\infty)$, and such that $\mathscr{V}(x) := \tilde{\mathscr{V}}(|x-x_0|)+1$ is twice continuously differentiable on $\mathbb{R}^d$. Now, because of **(A1)** and **(A4)**, $\mathscr{L}\mathscr{V}(x)$ is well defined and the process

$$\mathscr{V}(X_t) - \mathscr{V}(X_0) - \int_0^t \mathscr{L}\mathscr{V}(X_s)\,ds \qquad t \geq 0\,,$$

is a local martingale. For $x \in \mathbb{R}^d$, $|x| \geq r_1$, we have that

$$\begin{aligned}
\mathscr{L}\mathscr{V}(x) &= \frac{1}{2}C_{x_0}(x)\bar{\mathscr{V}}''(|x-x_0|) + \frac{\bar{\mathscr{V}}'(|x-x_0|)}{2|x-x_0|}(2A(x) - C_{x_0}(x) + 2\langle x-x_0, b(x)\rangle) \\
&\quad + \int_{\mathbb{R}^d}\big(\mathscr{V}(y+x) - \mathscr{V}(x) - \langle y, \nabla\mathscr{V}(x)\rangle \mathbb{1}_{B_1(0)}(y)\big)\nu(x,dy) \\
&\leq \frac{1}{2}C_{x_0}(x)\bar{\mathscr{V}}''(|x-x_0|) + \frac{\bar{\mathscr{V}}'(|x-x_0|)}{2|x-x_0|}(2A(x) - C_{x_0}(x) + 2B_{x_0}(x)) \\
&\leq -\frac{1}{2}\varphi_\Lambda\left(\int_{r_0}^{|x-x_0|} \mathrm{e}^{-I_{x_0}(u)}du + 1\right) \\
&\leq -\frac{1}{2}\varphi_\Lambda(\mathscr{V}(x))\,,
\end{aligned}$$

where in the second step we used **(A4)** and properties of $\mathscr{V}(x)$ (that is, $\tilde{\mathscr{V}}(r)$), and the final step follows from (2.3.7). Finally, because of **(A2)** and **(A5)**, as in the proof of

Theorem 2.3.1, we are again in a position to apply [DFG09, Theorems 3.2 and 3.4 (i)] and [Twe94, Theorems 5.1 and 7.1], which concludes the proof. ■

- *Remarks:* If $2A(x) - C_{x_0}(x) + 2B_{x_0}(x) \leq 0$ for all $x \in \mathbb{R}^d$, $|x - x_0| \geq r_0$, then we can replace $\gamma_{x_0}(r)$ and $\iota_{x_0}(r)$ by

$$\begin{aligned}\gamma_{x_0}(r) &= \inf_{|x-x_0|=r} N_{x_0}(x)\,, \qquad r>0\,,\\ \iota_{x_0}(r) &= \sup_{|x-x_0|=r} \frac{2A(x)-C_{x_0}(x)+2B_{x_0}(x)}{N_{x_0}(x)}\,, \qquad r>0\,,\end{aligned}$$

where

$$N_{x_0}(x) = \frac{\langle x-x_0, (c(x)+n(x))(x-x_0)\rangle}{|x-x_0|^2}\,, \qquad x \in \mathbb{R}^d \setminus \{0\}\,,$$

and $n(x) = (n_{ij}(x))_{i,j=1,\dots,d}$ with $n_{ij}(x) = \int_{B_1(0)} y_i y_j \nu(x,dy)$. Also, in this situation, the requirement in Theorem 2.3.7 that $c(x)$ is positive definite for all $x \in \mathbb{R}^d$, $|x - x_0| \geq r_0$, can be replaced by the requirement that $c(x) + n(x)$ is positive definite for all $x \in \mathbb{R}^d$, $|x - x_0| \geq r_0$.

- If $\varphi(t)$ is bounded, then (2.3.1) reads

$$\int_{r_0}^{\infty} \frac{\mathrm{e}^{I_{x_0}(u)}}{\gamma_{x_0}(u)} du < \infty\,,$$

and gives a condition for ergodicity (see [Wan08, Theorem 1.2] for the one-dimensional case).

- If in Theorem 2.3.7 $\liminf_{t\to\infty} \varphi'(t) > 0$ then, as in Proposition 2.3.6, we conclude that $\{X_t\}_{t\geq 0}$ is geometrically ergodic (see also [Wan08, Theorem 1.3] for the one-dimensional case).

Let us now give an example satisfying conditions from Theorem 2.3.7.

Example 2.3.8 (Lévy-driven SDEs)**.** Let $\{Y_t\}_{t\geq 0}$ be an n-dimensional Lévy process, and let $\Phi : \mathbb{R}^d \to \mathbb{R}^{d\times n}$ be bounded and locally Lipschitz continuous. Then, in [SS10, Theorems 3.1 and 3.5, and Corollary 3.3] (see also [BSW13, Theorem 3.8]) it has been shown that the SDE

$$dX_t = \Phi(X_{t-})\,dY_t\,, \qquad X_0 = x \in \mathbb{R}^d\,, \tag{2.3.8}$$

admits a unique strong solution which is a non-explosive strong Markov process whose semigroup satisfies the Feller and C_b-Feller property (thus **(A2)** holds true). Also, it has been shown that $\{X_t\}_{t\geq 0}$ satisfies **(A1)** with certain coefficients $b(x)$, $c(x)$ and $\nu(x,dy)$, which in a special case we give below. Observe that the following SDE is a special case of (2.3.8),

$$dX_t = \Phi_1(X_{t-})\,dt + \Phi_2(X_{t-})\,dB_t + \Phi_3(X_{t-})\,dZ_t\,, \qquad X_0 = x \in \mathbb{R}^d\,, \tag{2.3.9}$$

where $\Phi_1 : \mathbb{R}^d \to \mathbb{R}^d$, $\Phi_2 : \mathbb{R}^d \to \mathbb{R}^{d\times p}$ and $\Phi_3 : \mathbb{R}^d \to \mathbb{R}^{d\times q}$, with $p+q = n-1$, are locally Lipschitz continuous and bounded, $\{B_t\}_{t\geq 0}$ is a p-dimensional Brownian motion, and $\{Z_t\}_{t\geq 0}$ is a q-dimensional pure-jump Lévy process (that is, a Lévy process determined by a Lévy triplet of the form $(0,0,\nu_Z(dy))$) independent of $\{B_t\}_{t\geq 0}$. Namely, set $\Phi(x) = \big(\Phi_1(x),\Phi_2(x),\Phi_3(x)\big)$, and $Y_t = (t, B_t, Z_t)^T$, $t \geq 0$. Assume now that $d = p = q = 1$. Then, from [SS10, Theorem 3.1] we see that the corresponding coefficients read

$$b(x) = \begin{cases} \Phi_1(x)\,, & \Phi_3(x) = 0\,,\\ \Phi_1(x) + \int_{\mathbb{R}} y\left(\mathbb{1}_{B_1(0)}(y) - \mathbb{1}_{B_{|\Phi_3(x)|}(0)}(y)\right)\nu_Z\left(\frac{dy}{\Phi_3(x)}\right), & \Phi_3(x) \neq 0\,,\end{cases}$$

$$c(x) = \Phi_2^2(x)$$

$$\nu(x,dy) = \begin{cases} 0\,, & \Phi_3(x) = 0\,,\\ \nu_Z\left(\frac{dy}{\Phi_3(x)}\right), & \Phi_3(x) \neq 0\,.\end{cases}$$

Take now, for simplicity,

$$\Phi_1(x) = \Phi_3(x) = \begin{cases} -1\,, & x \geq 1\,,\\ -x\,, & -1 \leq x \leq 1\,,\\ 1\,, & x \leq -1\,,\end{cases}$$

$\Phi_2(x) = 1$, and $\nu_Z(dy) = f(y)dy$ with $f(y)$ being the probability density function of the continuous uniform distribution on the segment $[0,1]$. It is straightforward to see that $\{X_t\}_{t\geq 0}$ satisfies **(A4)** and **(A5)**. Open-set irreducibility and aperiodicity of $\{X_t\}_{t\geq 0}$ have been considered on [Mas07, page 43] (see also [KC99, Theorem 3.1]). Finally, since

$$B_0(x) = \begin{cases} -\frac{1}{2}x\,, & x \geq 1\,,\\ \frac{1}{2}x\,, & x \leq -1\,,\end{cases}$$

it is elementary to check that $\{X_t\}_{t\geq 0}$ satisfies (2.3.1) with $x_0 = 0$, $r_0 = 1$ and $\varphi(t) = t^{\alpha}$, $\alpha \in (0,1)$. Thus, $\{X_t\}_{t\geq 0}$ is sub-geometrically ergodic with rate $t^{\alpha/(1-\alpha)}$.

Observe that the same conclusion follows by employing a version of the relation in (2.3.3) including jumps (see [San16a, Theorem 3.3]). However, if we take $\Phi_1(x) = -\mathrm{sgn}(x)(\cos x + 3/2)$ (analogously as in Example 2.3.2), then it is not hard to see that (2.3.3) does not hold. On the other hand, Theorem 2.3.7 (with $x_0 = 0$, $r_0 = 1$ and $\varphi(t) = t^{\alpha}$, $\alpha \in (0,1)$) implies that $\{X_t\}_{t\geq 0}$ is again sub-geometrically ergodic with rate $t^{\alpha/(1-\alpha)}$.

An alternative approach to obtain a class of Markov processes with jumps (from diffusion processes) is through the **Bochner's subordination method**.

Definition 2.3.9.
- A **subordinator** $\{S_t\}_{t\geq 0}$ is a non-decreasing Lévy process on $[0,+\infty\rangle$ with Laplace transform

$$\mathbb{E}\left[\mathrm{e}^{-uS_t}\right] = \mathrm{e}^{-t\phi(u)}, \qquad u > 0, t \geq 0.$$

- A function $\phi : \langle 0,+\infty\rangle \to \langle 0,+\infty\rangle$ is called a **Bernstein function** if it is of class C^{∞} and $(-1)^n \phi^{(n)}(u) \geq 0$ for all $n \in \mathbb{N}$.

↬ *Remarks:* the characteristic (Laplace) exponent ϕ of a subordinator is a Bernstein function.

↬ Every Bernstein function admits a unique (Lévy-Khintchine) representation

$$\phi(u) = bu + \int_{(0,\infty)} (1-\mathrm{e}^{-uy})\,\nu(dy), \qquad u > 0,$$

where $b \geq 0$ is the drift parameter and ν is a Lévy measure, that is, a measure on $\mathscr{B}((0,\infty))$ satisfying $\int_{(0,\infty)} (1 \wedge y)\nu(dy) < \infty$. For more on subordinators and Bernstein functions we refer the readers to the monograph [SSV12].

We will now define the jump process via the Bochner's subordination. We start from a Markov process $\{X_t\}_{t\geq 0}$ with state space $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$ and transition kernel $p(t,x,dy)$.

Definition 2.3.10. Let $\{S_t\}_{t\geq 0}$ be a subordinator with characteristic exponent $\phi(u)$, independent of a Markov process $\{X_t\}_{t\geq 0}$. The process $X^{\phi} = \{X_t^{\phi}\}_{t\geq 0}$ defined by $X_t^{\phi} := X_{S_t}$, for $t \geq 0$, is referred to as the **subordinate process** $\{X_t\}_{t\geq 0}$ with subordinator $\{S_t\}_{t\geq 0}$ in the sense of Bochner.

- *Remarks:* process X^{ϕ} is obtained from $\{X_t\}_{t\geq 0}$ by a random time change through $\{S_t\}_{t\geq 0}$.

- $\{M_t^{\phi}\}_{t\geq 0}$ is again a Markov process with transition kernel

$$p^{\phi}(t,x,dy) = \int_{[0,+\infty\rangle} p(s,x,dy)\,\mu_t(ds)\,,$$

  where $\mu_t(\cdot) = \mathbb{P}(S_t \in \cdot)$ is the transition probability of S_t, $t \geq 0$.

- π measures of both processes coincide. Namely, if π is an invariant probability measure for $\{X_t\}_{t\geq 0}$, then π is also invariant for the subordinate process $\{X_t^{\phi}\}_{t\geq 0}$.

- It has been proven (see [DSS17]) that if $\{X_t\}_{t\geq 0}$ is sub-geometrically ergodic with Borel measurable rate $r(t)$ (with respect to the total variation distance), then $\{X_t^{\phi}\}_{t\geq 0}$ is sub-geometrically ergodic with rate $r_{\phi}(t) = \mathbb{E}[r(S_t)]$.

  - This result implies that, as an direct application of Theorem 2.3.1, we obtain sub-geometric ergodicity results for a class of subordinate diffusion processes.

2.4. Ergodicity with Respect to the Wasserstain Distance

In order to obtain ergodicity in total variation norm (that is, to apply the Foster - Lyapunov method), we had to impose certain regularity assumptions on our process. Namely, we had to assume open-set irreducibility and aperiodicity. As we discussed later on, these properties were sometimes hard to check, so we had found sufficient conditions ensuring them. However, these conditions imposed quite strong regularity and smoothness assumptions on the coefficient $c(x)$. In a case this is not satisfied, for example, when the diffusion coefficient σ is a singular matrix, we would not be able to derive ergodicity with respect to the total variation norm. The case of a singular coefficient σ corresponds to the situation when the diffusive part of the SDE corresponding to the Brownian motion can be projected to a space of lower dimension.

In such cases we turn our attention to search for a different approach that will provide some type of the convergence, typically weaker in some sence. As we discussed, we will consider the Wasserstein distances as our distance functions, and use the so-called synchronous coupling method (see [Che05, Example 2.16] for details) to derive sub-geometric ergodicity.

We start with the following auxiliary result, which will be crucial in the proofs of Theorems 2.4.2 and 2.4.4, and which is a version of non-linear convex Gronwall's inequality.

Lemma 2.4.1. Let $\Gamma > 0$, and let $f : [0,T) \to [0,\infty)$, with $0 < T \leq \infty$, and $\psi : [0,\infty) \to [0,\infty)$ be such that

(i) $f(t)$ is absolutely continuous on $[t_0,t_1]$ for any $0 < t_0 < t_1 < T$;

(ii) $f'(t) \leq -\Gamma\psi(f(t))$ a.e. on $[0,T)$;

(iii) $\psi(f(t)) > 0$ a.e. on $[0,T)$, and $\Psi_{f(0)}(t) := \int_t^{f(0)} \frac{ds}{\psi(s)} < \infty$ for all $t \in (0, f(0)]$.

Then,

$$f(t) \leq \Psi^{-1}_{f(0)}(\Gamma t), \qquad 0 \leq t < \Gamma^{-1}\Psi_{f(0)}(0) \wedge T.$$

In addition, if there is $\kappa \in [f(0),\infty]$ such that $\Psi_\kappa(t) := \int_t^\kappa \frac{ds}{\psi(s)} < \infty$ for $t \in (0,\kappa]$, then

$$f(t) \le \Psi_\kappa^{-1}(\Gamma t), \qquad 0 \le t < \Gamma^{-1}\Psi_{f(0)}(0) \wedge T.$$

Also, if $\psi(t)$ is convex and vanishes at zero, then $\Psi_{f(0)}(0) = \infty$, that is, the above relations hold for all $t \in [0,T)$.

Proof. By assumption,

$$-\Psi_{f(0)}(f(t)) = \int_{f(0)}^{f(t)} \frac{ds}{\psi(s)} = \int_0^t \frac{f'(s)\,ds}{\psi(f(s))} \le -\Gamma t, \qquad t \in [0,T).$$

Now, the first assertion follows.

The second claim follows from the fact that $\Psi_{f(0)}(t) \le \Psi_\kappa(t)$ for all $t \in (0, f(0)]$, while the last part follows from

$$\psi(t) = \psi(t + (1-t)0) \le t\psi(1) + (1-t)\psi(0) = t\psi(1), \qquad t \in [0,1].$$

∎

Theorem 2.4.2. Let $\sigma(x) \equiv \sigma$ be an arbitrary $d \times n$ matrix, and assume **(C1)-(C3)**. Further, let $p \ge 1$ and let $f, \psi : [0,\infty) \to [0,\infty)$ be such that

(i) $f(t)$ is concave, non-decreasing, absolutely continuous on $[t_0,t_1]$ for any $0 < t_0 < t_1 < \infty$, and $f(t) = 0$ if and only if $t = 0$;

(ii) $\psi(t)$ is convex and $\psi(t) = 0$ if and only if $t = 0$;

(iii) there are $\gamma > 0$, $\Gamma > 0$ and $t_0 > 0$, such that $f(t_0) \le \gamma$ and

$$f'(|x-y|)\langle x-y, b(x)-b(y)\rangle \le \begin{cases} -\Gamma|x-y|\,\psi(f(|x-y|)), & f(|x-y|) \le \gamma, \\ 0, & f(|x-y|) > \gamma, \end{cases} \tag{2.4.1}$$

a.e. on $\mathbb{R}^d$.

Then,

(a) for all $x,y \in \mathbb{R}^d$, $f(|x-y|) \le \gamma$, it holds that

$$\mathscr{W}_{f,p}(\delta_x P_t, \delta_y P_t) \le \Psi^{-1}_{f(|x-y|)}(\Gamma t), \qquad t \ge 0, \tag{2.4.2}$$

where $\Psi_\kappa(t) := \int_t^\kappa \frac{ds}{\psi(s)}$ for $\kappa > 0$ and $t \in (0,\kappa]$.

(b) for all $x,y \in \mathbb{R}^d$, $f(|x-y|) \leq \gamma$, and all $\kappa \geq \gamma$ it holds that

$$\mathscr{W}_{f,p}(\delta_x P_t, \delta_y P_t) \leq \Psi_\kappa^{-1}(\Gamma t), \qquad t \geq 0. \tag{2.4.3}$$

In addition, if $\Psi_\infty(t) := \int_t^\infty \frac{ds}{\psi(s)} < \infty$ for $t \in (0,\infty)$, then

$$\mathscr{W}_{f,p}(\delta_x P_t, \delta_y P_t) \leq \Psi_\infty^{-1}(\Gamma t), \qquad t \geq 0. \tag{2.4.4}$$

(c) for any $x,y \in \mathbb{R}^d$ it holds that

$$\mathscr{W}_{f,p}(\delta_x P_t, \delta_y P_t) \leq \lceil \delta |x-y| \rceil \Psi_\gamma^{-1}(\Gamma t), \qquad t \geq 0, \tag{2.4.5}$$

where $\delta := \inf\{t > 0 : f(t^{-1}) \leq \gamma\}$ and $\lceil u \rceil$ denotes the least integer greater than or equal to $u \in \mathbb{R}$. Also, according to (b), $\Psi_\gamma^{-1}(\Gamma t)$ in (2.4.5) can be replaced by $\Psi_\kappa^{-1}(\Gamma t)$ for any $\kappa \geq \gamma$, and by $\Psi_\infty^{-1}(\Gamma t)$ if $\Psi_\infty(t) < \infty$ for $t \in (0,\infty)$.

Proof. $*$*Idea*$*$ The proof of Theorem 2.4.2 is based on the so-called synchronous coupling method (see [Che05, Example 2.16] for details) and the asymptotic flatness condition given in (2.4.1). The idea of the method is to choose two solutions to the SDE starting from different positions (but corresponding to the same Brownian motion B) and observe the time when they meet (the so-called coupling time). Note that we do not know that the coupling time is finite, but it will be of no problem to us.

• Fix $x,y \in \mathbb{R}^d$, $x \neq y$, and let $\{X_t\}_{t\geq 0}$ and $\{Y_t\}_{t\geq 0}$ be solutions to (2.0.1) starting from x and y, respectively. Further, define $\tau := \inf\{t > 0 : X_t = Y_t\}$ and

$$Z_t := \begin{cases} Y_t, & t < \tau, \\ X_t, & t \geq \tau, \end{cases} \qquad t \geq 0.$$

By employing the strong Markov property it is easy to see that $\mathbb{P}^y(Z_t \in \cdot) = \mathbb{P}^y(Y_t \in \cdot)$ for all $t \geq 0$. Consequently,

$$\mathscr{W}_{f,p}(\delta_x P_t, \delta_y P_t) \leq (\mathbb{E}(f(|X_t - Z_t|)^p))^{1/p}, \qquad t \geq 0.$$

Next, since the mapping $t \mapsto |X_t - Z_t|$ is absolutely continuous on $[0,\tau)$, the function $t \mapsto f(|X_t - Z_t|)$ is differentiable a.e. on $[0,\tau)$ and we have that

$$\frac{d}{dt} f(|X_t - Z_t|) = \frac{f'(|X_t - Z_t|)}{|X_t - Z_t|} \langle X_t - Z_t, b(X_t) - b(Z_t) \rangle,$$

a.e. on $[0,\tau)$. Now, by assumption, we get

$$\frac{d}{dt} f(|X_t - Z_t|) \leq 0,$$

a.e. on $[0,\tau)$, which implies that the function $t \mapsto f(|X_t - Z_t|)$ is non-increasing on $[0,\infty)$. Take now $x,y \in \mathbb{R}^d$ such that $0 < f(|x-y|) \leq \gamma$ (which exist by (iii)). Thus, for such starting points, $f(|X_t - Z_t|) \leq \gamma$ on $[0,\infty)$. Now, by assumption,

$$\frac{d}{dt} f(|X_t - Z_t|) \leq -\Gamma \psi(f(|X_t - Z_t|)),$$

a.e. on $[0,\tau)$, which together with Lemma 2.4.1 gives

$$f(|X_t - Z_t|) \leq \Psi^{-1}_{f(|x-y|)}(\Gamma t), \qquad t \geq 0.$$

For $t \geq \tau$ the term on the left-hand side vanishes, and the term on the right-hand side is well defined and strictly positive ($\psi(t)$ is convex and $\psi(t) = 0$ if and only if $t = 0$). Now, by taking the expectation and infimum we conclude

$$\mathscr{W}_{f,p}(\delta_x P_t, \delta_y P_t) \leq \Psi^{-1}_{f(|x-y|)}(\Gamma t), \qquad t \geq 0,$$

which proves (a).

The relations in (b) now follow from (a) and Lemma 2.4.1.

Let us prove (c). If $f(|x-y|) \leq \gamma$ for all $x,y \in \mathbb{R}^d$, then the assertion follows from (a). Assume that there are $x,y \in \mathbb{R}^d$ such that $f(|x-y|) > \gamma$. Observe that, $\delta = 0$ if and only if $f(t) \leq \gamma$ for all $t \in [0,\infty)$. Thus, $\delta > 0$, and we have that

$$f\left(\frac{|x-y|}{\lceil \delta |x-y| \rceil}\right) \leq f(\delta^{-1}) \leq \gamma.$$

Take $z_0, \dots, z_{\lceil \delta|x-y| \rceil} \in \mathbb{R}^d$, such that $z_0 = x$ and

$$z_{i+1} = z_i + \frac{y-x}{\lceil \delta |x-y| \rceil}, \qquad i = 0, \dots, \lceil \delta |x-y| \rceil - 1.$$

By construction, $f(|z_0 - z_1|) = \cdots = f(|z_{\lceil \delta |x-y| \rceil - 1} - z_{\lceil \delta |x-y| \rceil}|) \leq \gamma$. Thus, using (b) we conclude that for $x,y \in \mathbb{R}^d$ such that $f(|x-y|) > \gamma$,

$$\begin{aligned}\mathscr{W}_{f,p}(\delta_x P_t, \delta_y P_t) &\leq \mathscr{W}_{f,p}(\delta_{z_0} P_t, \delta_{z_1} P_t) + \cdots + \mathscr{W}_{f,p}(\delta_{z_{\lceil \delta|x-y|\rceil - 1}} P_t, \delta_{z_{\lceil \delta |x-y| \rceil}} P_t) \\ &\leq \lceil \delta |x-y| \rceil \Psi^{-1}_{\gamma}(\Gamma t), \qquad t \geq 0.\end{aligned}$$

Finally, observe that if $t > 0$ is such that $f(t) \leq \gamma$, then $\delta \leq 1/t$, that is, $\delta t \leq 1$. Hence, for $x,y \in \mathbb{R}^d$ such that $f(|x-y|) \leq \gamma$ we have $\lceil \delta |x-y| \rceil = 1$, which concludes the proof. ■

- *Remarks:* we needed our two solutions to have the same underlying Brownian motion in order that the diffusive component σ is canceled when calculating $\frac{d}{dt}f(|X_t - Z_t|)$.

- Observe that $f(t)$ is $\mathscr{B}([0,+\infty\rangle)$-measurable, implying that the relation in (2.4.2) is well defined.

This result is in general broader than the existing ones, but in some special cases it coincides with some known ones. We proceed by listing some special cases of the theorem and their connection with the existing results:

(i) $\psi(t) = t \Longrightarrow$ geometric rate of convergence with $\Psi^{-1}_{f(|x-y|)}(\Gamma t) = f(|x-y|)\mathrm{e}^{-\Gamma t}$

- This result can be obtained in an alternative way (without Lemma 2.4.1, that is, Grönwall's inequality), by applying Itô's lemma to processes $\{f(|X_t - Z_t|)\}_{t\geq 0}$ and $\{\mathrm{e}^{\Gamma t} f(|X_t - Z_t|)\}_{t\geq 0}$.

(ii) $p = 2$ and $f(t) = \psi(t) = t \Longrightarrow$ again $\Psi^{-1}_{f(|x-y|)}(\Gamma t) = |x-y|\mathrm{e}^{-\Gamma t}$

- In [vRS05] it has been shown that the relation in (2.4.2) is equivalent to the asymptotic flatness condition (in the sense of [ABG12])

$$\langle x-y, b(x)-b(y)\rangle \leq -\Gamma|x-y|^2, \qquad x, y \in \mathbb{R}^d. \tag{2.4.6}$$

- Even though at first sight the condition in (2.4.1) seems to be less restrictive than the condition in (2.4.6), they are actually equivalent. This can be easily observed by taking an equidistant subdivision of the line segment connecting x and y, such that the distance between consecutive points is strictly less than γ, and then applying triangle inequality. On the other hand, in the case when $\psi(t)$ is not the identity function this does not hold in general. Namely, $\psi(t)$ is not sub-additive, but super-additive. A typical example of a drift function (in dimension $d = 1$) satisfying (2.4.1) (and (2.4.8)), but not (2.4.6), is $b(x) = -\mathrm{sgn}(x)|x|^p$, $p > 1$, together with $f(t) = t$ and $\psi(t) = |t|^p$ (see Example 2.4.5). More generally, no drift function that is sub-linear near the origin can satisfy (2.4.6), but it might satisfy (2.4.1).

Let us now give several remarks regarding out result.

Remark 2.4.3. (i) If the condition in (2.4.1) holds for some $\gamma > 0$, then it also holds for any $0 < \bar{\gamma} \leq \gamma$.

(ii) By replacing the condition in (2.4.1) with

$$f(|x-y|)^{p-1} f'(|x-y|)\langle x-y, b(x)-b(y)\rangle \leq \begin{cases} -\frac{\Gamma}{p}|x-y|\psi(f^p(|x-y|)), & f^p(|x-y|) \leq \gamma, \\ 0, & f^p(|x-y|) > \gamma, \end{cases}$$

a.e. on $\mathbb{R}^d$ for $\gamma > 0$ and $\Gamma > 0$, leads to analogous results ($f(t)$ is replaced by $f^p(t)$ in every relation).

(iii) For any $\mu, \nu \in \mathscr{P}$ it holds that

$$\mathscr{W}_{f,p}(\mu P_t, \nu P_t) \leq (\delta \mathscr{W}_p(\mu,\nu)+1)\Psi_\gamma^{-1}(\Gamma t), \qquad t \geq 0.$$

In particular, for $f(t) = t$ we have that

$$\mathscr{W}_p(\mu P_t, \nu P_t) \leq \left(\frac{\mathscr{W}_p(\mu,\nu)}{\gamma}+1\right)\Psi_\gamma^{-1}(\Gamma t), \qquad t \geq 0.$$

(iv) In the case when $f(t) = \psi(t) = t$, according to (2.4.6), we get

$$W_p(\mu P_t, \nu P_t) \leq \mathscr{W}_p(\mu,\nu)\mathrm{e}^{-\Gamma t}, \qquad p \geq 1,\ \mu,\nu \in \mathscr{P},\ t \geq 0, \tag{2.4.7}$$

which is the same results as in [vRS05] (for $p = 2$). Also, by an analogous approach as in the proof of Theorem 2.4.4, from (2.4.7) we see that $\{X_t\}_{t\geq 0}$ admits a unique invariant $\pi \in \cap_{p\geq 1}\mathscr{P}_p$ such that

$$\mathscr{W}_p(\mu P_t, \pi) \leq \mathscr{W}_p(\mu,\pi)\mathrm{e}^{-\Gamma t}, \qquad p \geq 1,\ \mu \in \mathscr{P}_p,\ t \geq 0.$$

(v) From (2.4.7) we see that the mapping $\mathscr{P} \ni \mu \mapsto \mu P_t \in \mathscr{P}$ is a contraction for fixed $t > 0$, that is, the right-hand side in (2.4.7) is strictly smaller than $\mathscr{W}_p(\mu,\nu)$. On the other hand, in the general situation, this is not the case anymore (see (iii)). However, if

$$f'(|x-y|)\langle x-y, b(x)-b(y)\rangle \leq -\Gamma|x-y|\psi(f(|x-y|)), \qquad x,y \in \mathbb{R}^d,$$

then from (2.4.2) we have that for all $x, y \in \mathbb{R}^d$ and all $t \geq 0$,

$$\mathscr{W}_{f,p}(\delta_x P_t, \delta_y P_t) \leq \Psi^{-1}_{f(|x-y|)}(\Gamma t) \leq \Psi^{-1}_{f(|x-y|)}(0) = f(|x-y|),$$

that is,

$$\mathscr{W}_{f,p}(\mu P_t, \nu P_t) \leq \mathscr{W}_{f,p}(\mu, \nu), \qquad p \geq 1,\ \mu, \nu \in \mathscr{P},\ t \geq 0.$$

Thus, the mapping $\mathscr{P} \ni \mu \mapsto \mu P_t \in \mathscr{P}$ is contractive for any fixed $t \geq 0$.

Finally, as a consequence of Theorem 2.4.2 we conclude the following.

Theorem 2.4.4. In addition to the assumptions of Theorem 2.4.2 with $f(t) = t$, assume

$$\langle x-y, b(x)-b(y)\rangle \leq -\Gamma|x-y|\,\psi(|x-y|), \qquad x, y \in \mathbb{R}^d. \tag{2.4.8}$$

Then, the process $\{X_t\}_{t\geq 0}$ admits a unique invariant $\pi \in \cap_{p\geq 1}\mathscr{P}_p$, and for any $\kappa > 0$, $p \geq 1$ and $\mu \in \mathscr{P}_p$,

$$\mathscr{W}_p(\mu P_t, \pi) \leq \left(\frac{\mathscr{W}_p(\mu, \pi)}{\kappa} + 1\right)\Psi^{-1}_\kappa(\Gamma t), \qquad t \geq 0. \tag{2.4.9}$$

Proof of Theorem 2.4.4. First, we prove that $\{X_t\}_{t\geq 0}$ admits an invariant probability measure. According to [MT93a, Theorem 3.1], this will follow if we show that for each $x \in \mathbb{R}^d$ and $0 < \varepsilon < 1$ there is a compact set $C \subset \mathbb{R}^d$ (possibly depending on x and ε) such that

$$\liminf_{t \nearrow \infty} \frac{1}{t}\int_0^t p(s, x, C)\, ds \geq 1 - \varepsilon.$$

By taking $y = 0$ in (2.4.8) we have that

$$\langle x, b(x)\rangle \leq \langle x, b(0)\rangle - \Gamma|x|\psi(|x|) \leq |b(0)||x| - \Gamma|x|\psi(|x|), \qquad x \in \mathbb{R}^d.$$

In particular, for $\mathscr{V}(x) = |x|^2$ we have that

$$\mathscr{L}\mathscr{V}(x) = 2\langle x, b(x)\rangle + \mathrm{Tr}\,\sigma\sigma^T \leq \mathrm{Tr}\,\sigma\sigma^T + 2|b(0)||x| - 2\Gamma|x|\psi(|x|), \qquad x \in \mathbb{R}^d.$$

Now, since every super-additive convex function is necessarily non-decreasing and unbounded, we conclude that there is $r_0 > 0$ large enough such that

$$\mathrm{Tr}\,\sigma\sigma^T + 2|b(0)||x| \leq \Gamma|x|\psi(|x|), \qquad |x| \geq r_0,$$

that is,

$$\begin{aligned}\mathscr{L}\mathscr{V}(x) &\le \left(\mathrm{Tr}\,\sigma\sigma^T + 2|b(0)||x| - 2\Gamma|x|\psi(|x|)\right)\mathbb{1}_{B_{r_0}(x)} \\ &\quad + \left(\mathrm{Tr}\,\sigma\sigma^T + 2|b(0)||x| - 2\Gamma|x|\psi(|x|)\right)\mathbb{1}_{B^c_{r_0}(x)} \\ &\le \left(\mathrm{Tr}\,\sigma\sigma^T + 2|b(0)||x| - 2\Gamma|x|\psi(|x|)\right)1_{B_{r_0}(x)} - \Gamma|x|\psi(|x|)\mathbb{1}_{B^c_{r_0}(x)} \\ &\le \left(\mathrm{Tr}\,\sigma\sigma^T + 2|b(0)|r_0 + \Gamma r_0\psi(r_0)\right)\mathbb{1}_{B_{r_0}(x)} - \Gamma r_0\psi(r_0)\,, \qquad |x| \ge r_0\,.\end{aligned}$$

Clearly, the above relation holds for all $r \ge r_0$ also. Now, according to [MT93b, Theorem 1.1] we conclude that for each $x \in \mathbb{R}^d$ and $r \ge r_0$ we have

$$\liminf_{t \nearrow \infty} \frac{1}{t}\int_0^t p(s,x,\bar{B}_r(0))\,ds \ge \frac{\Gamma r\psi(r)}{\mathrm{Tr}\,\sigma\sigma^T + 2|b(0)|r + \Gamma r\psi(r)}\,.$$

The assertion now follows by choosing r large enough.

Let us now show that any invariant $\pi \in \mathscr{P}$ of $\{X_t\}_{t\ge0}$ has finite all moments. Fix $p \ge 2$ and let $\mathscr{V}_p(x) = |x|^p$. By the same reasoning as above, it is easy to see that there are $r_p > 0$, $\Gamma_{p,1} > 0$ and $\Gamma_{p,2} > 0$ such that

$$\mathscr{L}\mathscr{V}_p(x) \le \Gamma_{p,1}\mathbb{1}_{B_{r_p}(0)}(x) - \Gamma_{p,2}|x|^{p-1}\psi(|x|)\,, \qquad x \in \mathbb{R}^d\,.$$

Now, from [MT93b, Theorem 4.3] it follows that

$$\int_{\mathbb{R}^d} |x|^{p-1}\psi(|x|)\pi(dx) \le \frac{\Gamma_{p,1}}{\Gamma_{p,2}}$$

for any corresponding invariant $\pi \in \mathscr{P}$.

Finally, let us prove that $\{X_t\}_{t\ge0}$ admits a unique invariant probability measure which satisfies (2.4.9). Let $\pi, \bar{\pi} \in \mathscr{P}$ be two invariant probability measures of $\{X_t\}_{t\ge0}$. Then, for any $\kappa > 0$ and $p \ge 1$ Remark 2.4.3 implies that

$$\mathscr{W}_p(\pi,\bar{\pi}) = \mathscr{W}_p(\pi P_t, \bar{\pi}P_t) \le \left(\frac{\mathscr{W}_p(\pi,\bar{\pi})}{\kappa} + 1\right)\Psi_\kappa^{-1}(\Gamma t)\,, \qquad t \ge 0\,.$$

Now, by letting $t \to \infty$ we see that $\mathscr{W}_p(\pi,\bar{\pi}) = 0$, that is, $\{X_t\}_{t\ge0}$ admits a unique invariant $\pi \in \mathscr{P}$. Finally, for any $\kappa > 0$, $p \ge 1$ and $\mu \in \mathscr{P}_p$, by employing Remark 2.4.3 again, we have that

$$\mathscr{W}_p(\pi,\mu P_t) = \mathscr{W}_p(\pi P_t, \mu P_t) \le \left(\frac{\mathscr{W}_p(\pi,\mu)}{\kappa} + 1\right)\Psi_\kappa^{-1}(\Gamma t)\,, \qquad t \ge 0\,,$$

which concludes the proof. ■

A special case of this is result is again for $\psi(t) = t$. In this case, ergodicity results have already been obtained, and they provide geometric rate of the convergence. However, these results do require σ to be a regular matrix and they correspond to the following version of an asymptotic flatness condition: there are some $\Gamma_1 > 0$, $\Gamma_2 > 0$ and $\Delta > 0$ such that

$$\langle x-y, b(x)-b(y)\rangle \leq \begin{cases} \Gamma_1|x-y|^2, & |x-y| \leq \Delta, \\ -\Gamma_2|x-y|^2, & |x-y| \geq \Delta, \end{cases} \qquad x, y \in \mathbb{R}^d, \tag{2.4.10}$$

The connection to our result is the following:

- If $\sigma(x) \equiv \sigma$ is quadratic and non-singular matrix, and $b(x)$ satisfies (2.4.10), by using the so-called coupling by reflection method (see [Che05, Example 2.16] for details), in [Ebe11] it has been shown that there is a concave function $f(t)$ (given explicitly in terms of the constants Γ_1, Γ_2 and Δ, and coefficients σ and $b(x)$) defining a metric $\rho(x,y) = f(|x-y|)$ on $\mathbb{R}^d$ under which $\{X_t\}_{t\geq 0}$ satisfies contraction property of the type (2.4.5) with geometric rate of convergence, and geometric ergodicity property of the type (2.4.9).
  - An example from before: $b(x) = -\text{sgn}(x)|x|^p$, $p > 1$, satisfies (2.4.1) and (2.4.8), but clearly it also satisfies (2.4.10). However, if σ is non-singular, we would not be able to use this result to obtain geometric ergodicity. On the other hand, singularity does not impact the validity of Theorem 2.4.4 and we can thus apply it in this case.
- If σ is non-singular, by taking $y = 0$ in (2.4.10), one can easily see that $\{X_t\}_{t\geq 0}$ is geometrically ergodic with respect to the total variation distance (see Proposition 2.3.6).
- In [BGG12] and [vRS05], the coupling approach and the asymptotic flatness property in (2.4.6) are employed to establish geometric contractivity and ergodicity of the semigroup of a diffusion process with possibly singular diffusion coefficient, with respect to a Wasserstein distance.

As an illustration, we give a simple example satisfying (2.4.1) and (2.4.8).

Example 2.4.5. Let $p > 1$, $b(x) = -\mathrm{sgn}(x)|x|^p$, $\sigma(x) \equiv \sigma \in \mathbb{R}$, $f(t) = t$, $\gamma > 0$ and $\psi(t) = t^p$. Now, it is easy to see that $b(x)$ cannot satisfy the relation in (2.4.6). On the other hand, an elementary computation shows that there is $\Gamma > 0$ such that (2.4.1) holds true. Thus, we have (2.4.5) with $\delta = \gamma^{-1}$. Also, $\lim_{t\to\infty} \sqrt[p-1]{t}\,\Psi_\kappa^{-1}(t) = 1/\sqrt[p-1]{p-1}$, $\kappa > 0$.

Let us also remark that one can show that the same result holds in the multidimensional case with $b(x_1,\dots,x_d) = (-\mathrm{sgn}(x_1)|x_1|^p,\dots,-\mathrm{sgn}(x_d)|x_d|^p)$.

2.4.1. Ergodicity of jump-processes

Just as in the case of a total variation distance, so in the case of Wasserstein distances too we take into account the situation of a jump-process and its ergodicity.

We will consider the situation where jumps come from a Lévy process. We considered this situation in the case of the total variation distance too. More precisely, we are looking at the solutions of a **Lévy driven SDE** that takes the form:

$$dX_t \,=\, b(X_t)\,dt + dY_t\,, \qquad X_0 = x \in \mathbb{R}^d\,, \tag{2.4.11}$$

where $\{Y_t\}_{t\geq 0}$ be a d-dimensional Lévy process with Lévy triplet (β,γ,ν) (coming from the Lévy-Itô decomposition) and $b : \mathbb{R}^d \to \mathbb{R}^d$ a continuous function.

We impose the following conditions on the coefficients:

(J1) for any $r > 0$ there is $\Gamma_r > 0$ such that for all $x,y \in B_r(0)$,

$$\langle x-y, b(x)-b(y)\rangle \,\leq\, \Gamma_r|x-y|^2\,;$$

(J2) there is $\Gamma > 0$ such that for all $x \in \mathbb{R}^d$,

$$\langle x, b(x)\rangle \,\leq\, \Gamma(1+|x|^2)\,.$$

↬ **(J1)-(J2)** imply that admits a unique strong non-explosive solution $\{X_t\}_{t\geq 0}$ which is a strong Markov process and satisfies the C_b-Feller property (see [Maj16, Theorem 1.1, and Lemmas 2.4 and 2.5]).

Lemma 2.4.6. Assume that $\mathbb{E}[|Y_1|^p] < \infty$ (or, equivalently, $\int_{B_1^c(0)} |y|^p \nu(dy) < \infty$) for some $p > 0$. Then, there is a constant $\Delta > 0$ such that

$$\mathbb{E}^x\big[|X_t|^p\big] \,\leq\, (|x|^p+1)\mathrm{e}^{\Delta t}\,, \qquad t \geq 0\,,\; x \in \mathbb{R}^d\,.$$

Proof. Let $\chi\in C^2(\mathbb{R}^d)$ be such that $\chi(x)\geq 0$, $\chi(x)\leq |x|^p$ and $\chi(x)=|x|^p$ for $x\in B_1^c(0)$. Further, for $n\in\mathbb{N}$, let $\chi_n\in C_b^2(\mathbb{R}^d)$ be such that $\chi_n(x)\geq 0$, $\chi_n(x)=\chi|_{B_{n+1}(0)}(x)$ and $\chi_n(x)\to\chi(x)$ as $n\to\infty$, and $\tau_n:=\inf\{t\geq 0: X_t\in B_n^c(0)\}$. Then, according to Itô's formula (see [ABW10, Remark 2.2]), we have that

$$\begin{aligned}\mathbb{E}^x[\chi_n(X_{t\wedge\tau_n})] &\leq \chi_n(x)+\Delta_n(t\wedge\tau_n)+\Delta_n\mathbb{E}^x\left[\int_0^{t\wedge\tau_n}\chi_n(X_s)ds\right]\\ &\leq \chi_n(x)+\Delta_n t+\Delta_n\int_0^t\mathbb{E}^x\left[\chi_n(X_{s\wedge\tau_n})\right]ds\,,\qquad n\in\mathbb{N},\,t\geq 0,\,x\in\mathbb{R}^d\,,\end{aligned}$$

where the constants $\Delta_n>0$ depend on p, β, γ, $b(x)$ and the constants $\int_{B_1(0)}|y|^2\nu(dy)$, $\nu(B_1^c(0))$, $\sup_{x\in B_R(0)}|\nabla\chi_n(x)|$ and $\sup_{x\in B_R(0)}|\nabla^2\chi_n(x)|$, for $R>0$ large enough. Clearly, the functions $\chi_n(x)$ can be chosen such that $\Delta:=\sup_{n\in\mathbb{N}}\Delta_n<\infty$. Now, since the function $t\mapsto\mathbb{E}^x[\chi_n(X_{t\wedge\tau_n})]$ is bounded and càdlàg, Gronwall's lemma implies that

$$\mathbb{E}^x[\chi_n(X_{t\wedge\tau_n})]\leq(\chi_n(x)+1)e^{\Delta t}-1\,,\qquad n\in\mathbb{N},\,t\geq 0,\,x\in\mathbb{R}^d\,.$$

By letting $n\to\infty$ monotone convergence theorem and non-explosivity of $\{X_t\}_{t\geq 0}$ imply that

$$\mathbb{E}^x[\chi(X_t)]\leq(\chi(x)+1)e^{\Delta t}-1\,,\qquad t\geq 0,\,x\in\mathbb{R}^d\,.$$

Finally, we have that

$$\mathbb{E}^x\left[|X_t|^p\right]\leq\mathbb{E}^x[\chi(X_t)]+1\leq(\chi(x)+1)\mathrm{e}^{\Delta t}\leq(|x|^p+1)\mathrm{e}^{\Delta t}\,,\qquad t\geq 0,\,x\in\mathbb{R}^d\,.$$

■

Lemma 2.4.7. Assume that $\nu(\mathbb{R}^d)<\infty$. Then, the sample paths of $\{X_t\}_{t\geq 0}$ are piecewise continuous $\mathbb{P}^x$-a.s.

Proof. Define $\tau_0:=0$ and

$$\tau_n:=\inf\{t\geq\tau_{n-1}:|X_t-X_{t-}|>0\}=\inf\{t\geq\tau_{n-1}:|Y_t-Y_{t-}|>0\}\,,\qquad n\geq 1\,.$$

Clearly, $\{\tau_n\}_{n\in\mathbb{N}}$ are i.i.d. and $\mathbb{P}^x(\tau_1>t)=\mathrm{e}^{-\nu(\mathbb{R}^d)t}$ (that is, τ_1 is exponentially distributed with parameter $\nu(\mathbb{R}^d)$) for any $x\in\mathbb{R}^d$. Hence, $\{X_t\}_{t\geq 0}$ is continuous on $[\tau_n,\tau_{n+1})$, $n\geq 0$, $\mathbb{P}^x$-a.s. for all $x\in\mathbb{R}^d$. ■

In establishing ergodicity, we will need the help of a generator of our process. Under certain additional conditions, we will be able to calculate it. Namely, let $\{X_t\}_{t\geq 0}$ be a solution to (2.4.11) with $b(x)$ satisfying **(J1)** and **(J2)**, and with $\{Y_t\}_{t\geq 0}$ having finite p-th moment, $p \geq 1$, and finite Lévy measure. Then, according to Lemmas 2.4.6 and 2.4.7, if $b(x)$ satisfies (2.4.1) we conclude that $\{X_t\}_{t\geq 0}$ satisfies (2.4.2), (2.4.3), (2.4.4) and (2.4.5). Further, according to [ABW10] and [Mas07], for any $f \in C^2(\mathbb{R}^d)$ such that $x \mapsto \int_{B_1^c(0)} f(x+y)\nu(dy)$ is locally bounded,

$$f(X_t) - f(X_0) - \int_0^t \mathscr{L} f(X_s)ds, \qquad t \geq 0,$$

is a local $\mathbb{P}^x$-martingale, $x \in \mathbb{R}^d$, where

$$\begin{aligned}\mathscr{L} f(x) &= \langle b(x), \nabla f(x)\rangle + \langle \beta, \nabla f(x)\rangle + \frac{1}{2}\mathrm{Tr}\,\gamma\nabla^2 f(x)\\ &\quad + \int_{\mathbb{R}^d} \Big(f(y+y) - f(x) - \langle y, \nabla f(x)\rangle \mathbb{1}_{B_1(0)}(y)\Big)\,\nu(dy)\,.\end{aligned}$$

Proposition 2.4.8. Let $p \geq 1$. Assume that $b(x)$ satisfies **(J1)**, **(J2)** and (2.4.8), and that $\{Y_t\}_{t\geq 0}$ has finite p-th moment and finite Lévy measure. Then, $\{X_t\}_{t\geq 0}$ admits a unique invariant $\pi \in \mathscr{P}_p$ such that for any $\kappa > 0$, $1 \leq q \leq p$ and $\mu \in \mathscr{P}_q$ it holds that

$$\mathscr{W}_q(\pi, \mu P_t) \leq \left(\frac{\mathscr{W}_q(\pi,\mu)}{\kappa} + 1\right)\Psi_\kappa^{-1}(\Gamma t)\,, \qquad t \geq 0\,. \tag{2.4.12}$$

Proof. First, observe that

$$\begin{aligned}\mathscr{L} f(x) &= \langle b(x), \nabla f(x)\rangle + \langle \beta + \int_{B_1^c(0)} y\,\nu(dy), \nabla f(x)\rangle + \frac{1}{2}\mathrm{Tr}\,\gamma\nabla^2 f(x)\\ &\quad + \int_{\mathbb{R}^d} \left(f(y+y) - f(x) - \langle y, \nabla f(x)\rangle\right)\nu(dy)\,.\end{aligned}$$

By taking a non-negative $\mathscr{V}_p \in C^2(\mathbb{R}^d)$ such that $\mathscr{V}_p(x) = |x|^p$ on $B_1^c(0)$ from [APS19, Lemma 5.1] we have that

$$\sup_{x\in\mathbb{R}^d}\left|\int_{\mathbb{R}^d}\left(\mathscr{V}_p(y+y) - \mathscr{V}_p(x) - \langle y, \nabla \mathscr{V}_p(x)\rangle\right)\nu(dy)\right| < \infty.$$

Now, by completely the same approach as in the proof of Theorem 2.4.4 we conclude that $\{X_t\}_{t\geq 0}$ admits a unique invariant $\pi \in \mathscr{P}$ such that $\int_{\mathbb{R}^d} |x|^{p-1}\psi(|x|)\pi(dx) < \infty$. Thus, $\pi \in \mathscr{P}_p$, and the relation in (2.4.12) follows by the same reasoning as in the proof of Theorem 2.4.4. ∎

The second type of jump processes that we discuss will be, just like in the case of a total variation distance, Markov processes with jumps obtained through Bochner's subordination method.

Proposition 2.4.9. Let $\{X_t\}_{t\geq 0}$ be a Markov process with state space $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$ and semigroup $\{P_t\}_{t\geq 0}$. Let $\{S_t\}_{t\geq 0}$ be a subordinator with characteristic exponent $\phi(u)$, independent of $\{X_t\}_{t\geq 0}$. Further, let ρ be a metric on $\mathbb{R}^d$ such that $(\mathbb{R}^d, \rho)$ is a Polish space and $\mathscr{B}(\mathbb{R}^d_\rho) \subseteq \mathscr{B}(\mathbb{R}^d)$, that is, ρ induces a coarser topology than the standard d-dimensional Euclidean metric on $\mathbb{R}^d$. Assume, that $\{X_t\}_{t\geq 0}$ admits an invariant $\pi \in \mathscr{P}$ such that $\mathscr{W}_{\rho,p}(\delta_x P_t, \pi) \leq \Gamma(x) r(t)$, $t \geq 0$, $x \in \mathbb{R}^d$, where $r : [0, +\infty\rangle \to [1, +\infty\rangle$ is Borel measurable and $\Gamma(x) \geq 0$. Then, $\mathscr{W}_{\rho,p}(\delta_x P_t^\phi, \pi) \leq \Gamma(x) r_\phi(t)$, $t \geq 0$, $x \in \mathbb{R}^d$, where $r_\phi(t) = (\mathbb{E}[r^p(S_t)])^{1/p}$.

Proof. First, recall that if π is an invariant measure for $\{X_t\}_{t\geq 0}$, then it is also invariant for $\{X_t^\phi\}_{t\geq 0}$. Next, [Vil09, Theorem 4.1] implies that for each $s \in [0,\infty)$ there is $\Pi_s \in \mathscr{C}(\delta_x P_s, \pi)$ such that $\mathscr{W}_{\rho,p}(\delta_x P_s^\phi, \pi) = \int_{\mathbb{R}^d \times \mathbb{R}^d} \rho(y,z) \Pi_s(dy,dz)$. Now, we have that

$$\begin{aligned}
\mathscr{W}^p_{\rho,p}(\delta_x P_t^\phi, \pi) &= \inf_{\Pi \in \mathscr{C}(\delta_x P_t^\phi, \pi)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \rho^p(y,z) \Pi(dy,dz) \\
&\leq \int_{\mathbb{R}^d \times \mathbb{R}^d} \rho^p(y,z) \int_{[0,\infty)} \Pi_s(dy,dz) \mu_t(ds) \\
&\leq \int_{[0,\infty)} \mathscr{W}^p_{\rho,p}(\delta_x P_s, \pi) \mu_t(ds) \\
&\leq \Gamma^p(x) \int_{[0,\infty)} r^p(s) \mu_t(ds) \\
&= \Gamma^p(x) \mathbb{E}[r^p(S_t)],
\end{aligned}$$

which completes the proof. ■

3. DIFFUSION PROCESSES WITH MARKOVIAN SWITCHING

A problem that motivated the research of the so-called diffusion processes with switching comes from physics - from observing systems which did not behave in the same way all the time, but instead, they changed their behaviour every now and then, at some random times. We call them *hybrid systems*. One can look at those systems as two-component ones: one component is continuous, and corresponds to a classical diffusion, and another is discrete, which corresponds to the notion of Markov chain. The role of the discrete component, usually independent of the continuous one, is to say when is a time for a continuous component to start behaving differently, and in what way (that is, how will this new motion look like). The continuous motion will then remain the same, until the discrete component changes its value again.

Mathematically, we say that such a system is a system with Markovian switching, or a regime-switching system, and look at it as a paired process (X,Λ), where X_t denotes a position of our process at time t, and Λ_t the state of the discrete component at time t. The value of X is constantly changing, and the value of Λ_t remains the same for a certain period of time after changing. We call it a **diffusion process with Markovian switching** or a **regime-switching diffusion process**. The model which describes the movement of such a process can be written in the form of the SDE, just, unlike before, this SDE will be a bit more complicated as all coefficients will depend on both components of the process. Actually, one can think about this process as a diffusion living in an random environment - and the environment is changing at discrete, random times.

We start the analysis by formally posing the problem, model and discussing its properties. Denote the process with regime-switching by $(X,\Lambda) = \{(X(x,i;t),\Lambda(x,i;t))\}_{t\geq 0}$.

The first (continuous-state) component is given by

$$\begin{aligned} dX(x,i;t) &= b\big(X(x,i;t),\Lambda(x,i;t)\big)dt + \sigma\big(X(x,i;t),\Lambda(x,i;t)\big)dB(t) \\ X(x,i;0) &= x \in \mathbb{R}^d \\ \Lambda(x,i;0) &= i \in \mathbb{S}, \end{aligned} \tag{3.0.1}$$

where $\{B(t)\}_{t\geq 0}$ denotes a standard n-dimensional Brownian motion (starting from the origin), and the second (regime-switching) component is a right-continuous and time-homogeneous Markov chain with finite state space $\mathbb{S}$. The processes $\{B(t)\}_{t\geq 0}$ and $\{\Lambda(x,i;t)\}_{t\geq 0}$ are both defined on a stochastic basis $(\Omega,\mathscr{F},\{\mathscr{F}_t\}_{t\geq 0},\mathbb{P})$ satisfying the usual conditions.

Then, Λ possesses the **generator matrix** (or the infinitesimal generator) $\mathscr{Q}(x) = (\mathrm{q}_{i,j}(x))_{i,j\in\mathbb{S}}$ which describes the movement of this Markov chain (elements of $\mathscr{Q}(x)$ are intensities of waiting times if this process which are exponentially distributed), where $x \in \mathbb{R}^d$ is the starting point of the system. So, once we determing the starting position, the behaviour of Λ is fixed, it stays the same till the end, and is not influenced by the continuous-state component X_t. Recall, this matrix satisfies the following properties:

- $\mathrm{q}_{i,j}(x) = \lim_{t\to 0} \frac{\mathbb{P}(\Lambda(x,i;t)=j)}{t}$ for $i \neq j$,
- $\mathrm{q}_{i,i}(x) = -\sum_{j\neq i} \mathrm{q}_{i,j}(x)$.

Some properties:

- Continuous-state component X does have a continuous trajectories, and in the case when we add jumps, they will be càdlàg.
- Maximum principle and Harnack inequalities do hold, see [CCTY19].
- The biggest structural difference between a regime-switched diffusion and a standard diffusion process is the fact the regime-switched process X is not a Markov process - this is obvious because the value of X at time t does depend on the past (through the discrete component Λ_t). Therefore, in order that the problem of ergodicity makes sense in this setting, we need to work with a joint process (X,Λ) as this will indeed be a Markov process. In this spirit, the problem of ergodicity can now be portrayed as:

$$r(t)\|\mathbb{P}^{x,i}((X_t,\Lambda_t)\in\cdot)-\pi(\cdot)\| \xrightarrow{?} 0,$$

and the same questions as before are being tackled. Similarly, all properties from before will be defined in the same way, just on a product space instead.

- Feller and C_b-Feller properties: have been considered, see [KZ20].

However, we may ask ourselves, <u>is this situation that much different</u> than a normal, standard diffusion process? The answer is: NO! There are some very big differences between these two types of processes which make this analysis interesting and important. Adding a switching mechanism does make the whole system much more complicated - it can exhibit some strange behaviour.

↬ For example, it is possible that all diffusions which form a switching process are positive recurrent, but the whole system is not. Or conversely, for all $i \in \mathbb{S}$ the corresponding diffusions are transient, but the regime-switched diffusion as a whole process is positive recurrent. See [PS92].

So, we proceed with the analysis. We assume that the coefficients $b:\mathbb{R}^d\times\mathbb{S}\to\mathbb{R}^d$ and $\sigma:\mathbb{R}^d\times\mathbb{S}\to\mathbb{R}^{d\times n}$, and the process $\{\Lambda(x,i;t)\}_{t\geq 0}$ satisfy the following:

(A1) for any $r>0$ and $i\in\mathbb{S}$,

$$\sup_{x\in\mathscr{B}_r(0)}\big(|b(x,i)|+\|\sigma(x,i)\|_{\mathrm{HS}}\big) < +\infty,$$

(A2) for each $(x,i)\in\mathbb{R}^d\times\mathbb{S}$ the regime-switching stochastic differential equation (shortly RSSDE) in eq. (3.0.1) admits a unique non-explosive strong solution $\{X(x,i;t)\}_{t\geq 0}$ which has continuous sample paths,

(A3) the process $\{(\mathrm{X}(x,i;t),\Lambda(x,i;t))\}_{t\geq 0}$ is a temporally-homogeneous strong Markov process with transition kernel $p(t,(x,i),dy\times\{j\})=\mathbb{P}((X(x,i;t),\Lambda(x,i;t))\in dy\times\{j\})$,

(A4) the corresponding semigroup of linear operators $\{\mathscr{P}_t\}_{t\geq 0}$, defined by

$$\mathscr{P}_t f(x,i) := \int_{\mathbb{R}^d\times\mathbb{S}} f(y,j)\,\mathsf{p}\big(t,(x,i),dy\times\{j\}\big)\,,\qquad f\in\mathscr{B}_b(\mathbb{R}^d\times\mathbb{S})\,,$$

satisfies the $\mathscr{C}_b$-Feller property, that is, $\mathscr{P}_t(\mathscr{C}_b(\mathbb{R}^d\times\mathbb{S}))\subseteq\mathscr{C}_b(\mathbb{R}^d\times\mathbb{S})$ for all $t\geq 0$,

(A5) for any $(x,i) \in \mathbb{R}^d \times \mathbb{S}$ and $f \in \mathscr{C}^2(\mathbb{R}^d \times \mathbb{S})$ the process

$$\left\{ f\big(X(x,i;t),\Lambda(x,i;t)\big) - f(x,i) - \int_0^t \mathscr{L} f\big(X(x,i;s),\Lambda(x,i;s)\big)ds \right\}_{t\geq 0}$$

is a $\mathbb{P}$-local martingale, where

$$\mathscr{L} f(x,i) = \mathscr{L}_i f(x,i) + \mathscr{Q}(x) f(x,i)$$

with

$$\mathscr{L}_i f(x) = \langle b(x,i), \nabla f(x) \rangle + \frac{1}{2}\mathrm{Tr}\big(\sigma(x,i)\sigma(x,i)^T \nabla^2 f(x)\big), \qquad f \in \mathscr{C}^2(\mathbb{R}^d),$$

and

$$\mathscr{Q}(x) f(i) = \sum_{j\in\mathbb{S}} f(j)\, \mathrm{q}_{i,j}(x), \qquad f \in \mathbb{S}^{\mathbb{S}}.$$

- $\mathscr{B}_r(x)$ denotes the open ball with radius $r > 0$ around $x \in \mathbb{R}^d$,
- $\langle \cdot,\cdot \rangle$ stands for the standard scalar product on $\mathbb{R}^d$,
- $|\cdot| := \langle \cdot,\cdot \rangle^{1/2}$ is the corresponding, Euclidean norm
- The symbols $\mathscr{B}(\mathbb{R}^d \times \mathbb{S})$, $\mathscr{B}_b(\mathbb{R}^d \times \mathbb{S})$, $\mathscr{C}_b(\mathbb{R}^d \times \mathbb{S})$ and $\mathscr{C}^2(\mathbb{R}^d \times \mathbb{S})$ stand for the spaces of all functions $f : \mathbb{R}^d \times \mathbb{S} \to \mathbb{R}$ such that $x \mapsto f(x,i)$ is Borel measurable, bounded and Borel measurable, bounded and continuous and of class $\mathscr{C}^2$ for all $i \in \mathbb{S}$, respectively.

↬ Conditions ensuring **(A1)** – **(A5)** can be found in [FGC19], [KZ20] and [MY06])

3.1. Literature overview

The research on regime-switching diffusions started in the 90's, after it was observed that standard models can not capture some interesting movements. Namely, people started to find examples of phenomena which change their behaviour at random time points. Such system, called the hybrid system, were then modelled by two stochastic components: one continuous, and one discrete.

As in the case of a classical diffusion, we can divide results based on the type of the convergence.

From the qualitative aspect, the properties of such models will relate to ones for standard SDE, since our process here (X, Λ) is a Markov process. Thereforem , we will observe the same properties, just on a different space (everything will be done on a product space instead of a just $\mathbb{R}^d$).

- [MY06]

  ↬ The main reference for this topic. Provides sufficient condition for various properties (for example, using the assumption of M-matrix).

- [KZ20]

  ↬ Here the authors discuss structural properties of regime-switching diffusions like Feller and C_b-property and irreducibility.

From the quantitative aspect, we again have much more results concerning geometric convergence with respect to the total variation norm.

- **Geometric rate:** as the whole theory itself, the results are quite recent for this case.

  TV: Most results correspond to the total-variation distance. See [Sha13], [LX22] or [Sha13].

  WASS: Only few results. See [Sha14] and [CH15].

- **Sub-geometric ergodicity:** this situation was not yet discussed in the literature for switching diffusions.

3.2. Irreducibility and Aperiodicity

As we have thoroughly discussed in the previous chapter, the necessary ingredient to use with a Foster - Lypaunov method is to assume open-set irreducibility and aperiodicity. As we have done a lot of research on those properties for classical diffusions, we will be able to extend those results to regime-switching diffusions as well.

We start by adapting the definitions of those properties to our new setting (the state space of the process is not $\mathbb{R}^d$ any more, but rather a product space $\mathbb{R}^d \times \mathbb{S}$).

Definition 3.2.1. The process $\{(X(x,i;t),\, \Lambda(x,i;t))\}_{t\geq 0}$ is said to be

- ϕ**-irreducible** if there exists a σ-finite measure ϕ on $\mathscr{B}(\mathbb{R}^d)$ (the Borel σ-algebra on $\mathbb{R}^d$) such that whenever $\phi(B) > 0$ we have $\int_0^\infty p(t,(x,i),B\times\{j\})\,dt > 0$ for all $(x,i) \in \mathbb{R}^d \times \mathbb{S}$ and $j \in \mathbb{S}$.

  If it is ϕ-irreduciblen then we call it

  - **transient** if there exists a countable covering of $\mathbb{R}^d$ with sets $\{B_k\}_{k\in\mathbb{N}} \subset \mathscr{B}(\mathbb{R}^d)$, and for each $k \in \mathbb{N}$ there exists a finite constant $c_k \geq 0$ such that $\int_0^\infty p(t,(x,i),B_k\times\{j\})\,dt \leq c_k$ holds for all $(x,i) \in \mathbb{R}^d \times \mathbb{S}$ and $j \in \mathbb{S}$
  - **recurrent** if $\phi(B) > 0$ implies $\int_0^\infty p(t,(x,i),B\times\{j\})\,dt = \infty$ for all $(x,i) \in \mathbb{R}^d \times \mathbb{S}$ and $j \in \mathbb{S}$.

- **open-set irreducible** if the support of its maximal irreducibility measure ψ,

$$\operatorname{supp}\psi = \{x \in \mathbb{R}^d : \psi(O) > 0 \text{ for every open neighborhood } O \text{ of } x\},$$

  has a non-empty interior,

- **aperiodic** if it admits an irreducible skeleton chain, that is, there exist $t_0 > 0$ and a σ-finite measure ϕ on $\mathscr{B}(\mathbb{R}^d)$, such that $\phi(B) > 0$ implies $\sum_{n=0}^\infty p(nt_0,(x,i),B\times\{j\}) > 0$ for all $(x,i) \in \mathbb{R}^d \times \mathbb{S}$ and $j \in \mathbb{S}$.

Recall, a right-continuous temporally-homogeneous Markov chain $\{\Lambda(i;t)\}_{t\geq 0}$ on $\mathbb{S}$ given by state-independent generator $\mathscr{Q}$ is irreducible if for any $i,j \in \mathbb{S}$, $i \neq j$, there are $m \in \mathbb{N}$ and $k_0,\dots,k_m \in \mathbb{S}$ with $k_0 = i$, $k_m = j$ and $k_l \neq k_{l+1}$ for $l = 0,\dots,m-1$, such that

$q_{k_l k_{l+1}} > 0$ for all $l = 0, \ldots, m-1$. Due to finiteness of $\mathbb{S}$, it is well known that $\{\Lambda(i;t)\}_{t\geq 0}$ is then geometrically ergodic. Let $\{\bar{\Lambda}(i;t)\}_{t\geq 0}$ be an independent copy of $\{\Lambda(i;t)\}_{t\geq 0}$. Put

$$\tau_{ij} := \begin{cases} \inf\{t > 0 \colon \Lambda(i;t) = \bar{\Lambda}(j;t)\}, & i \neq j, \\ 0, & i = j, \end{cases}$$

and $\zeta := \inf_{i,j\in\mathbb{S}} \mathbb{P}(\Lambda(i;1) = j)$. Observe that $0 < \zeta < 1$ (recall that $\mathbb{S}$ is finite and $\{\Lambda(i;t)\}_{t\geq 0}$ is irreducible). Define $\vartheta := -\log(1-\zeta)$. It holds that

$$\mathbb{P}(\tau_{ij} > t) \leq e^{-\vartheta\lfloor t\rfloor} \tag{3.2.1}$$

for all $i, j \in \mathbb{S}$ and $t \geq 0$.

Lemma 3.2.2. For $\{\Lambda(i;t)\}_{t\geq 0}$ irreducible and $\{\bar{\Lambda}(i;t)\}_{t\geq 0}$ its independent copy,

$$\tau_{ij} = \begin{cases} \inf\{t > 0 \colon \Lambda(i;t) = \bar{\Lambda}(j;t)\}, & i \neq j, \\ 0, & i = j, \end{cases}$$

$\zeta = \inf_{i,j\in\mathbb{S}} \mathbb{P}(\Lambda(i;1) = j)$ and $\vartheta = -\log(1-\zeta)$. Then,

$$\mathbb{P}(\tau_{ij} > t) \leq e^{-\vartheta\lfloor t\rfloor}.$$

for all $i, j \in \mathbb{S}$ and $t \geq 0$.

Proof. First, observe that for any initial distribution $\mu = (\mu_i)_{i\in\mathbb{S}}$ of $\{\Lambda(i;t)\}_{t\geq 0}$ and all $j \in \mathbb{S}$ it holds that

$$\sum_{i\in\mathbb{S}} \mathbb{P}(\Lambda(i;1) = j)\,\mu_i \geq \zeta. \tag{3.2.2}$$

Thus, for any initial distribution $\Pi = (\Pi_{i,j})_{i,j\in\mathbb{S}}$ of $\{(\Lambda(i;t), \bar{\Lambda}(j;t))\}_{t\geq 0}$ we have that

$$\begin{aligned}
&\sum_{i,j\in\mathbb{S}} \mathbb{P}(\Lambda(i;1) \neq \bar{\Lambda}(j;1))\,\Pi_{i,j} \\
&= \sum_{i,j\in\mathbb{S}} \sum_{k\in\mathbb{S}} \mathbb{P}(\Lambda(i,1) \neq \bar{\Lambda}(j;1), \Lambda(i;1) = k)\,\Pi_{i,j} \\
&= \sum_{i,j\in\mathbb{S}} \sum_{k\in\mathbb{S}} \big(\mathbb{P}(\Lambda(i;1) \neq \bar{\Lambda}(j;1), \Lambda(i;1) = k) + \mathbb{P}(\Lambda(i;1) = \bar{\Lambda}(j;1), \Lambda(i;1) = k) \\
&\qquad\qquad - \mathbb{P}(\Lambda(i;1) = \bar{\Lambda}(j;1), \Lambda(i;1) = k)\big)\,\Pi_{i,j} \\
&= \sum_{i,j\in\mathbb{S}} \sum_{k\in\mathbb{S}} \mathbb{P}(\Lambda(i;1) = k)\,\Pi_{i,j} - \sum_{i,j\in\mathbb{S}} \sum_{k\in\mathbb{S}} \mathbb{P}(\Lambda(i;1) = k, \Lambda(j;1) = k)\,\Pi_{i,j} \\
&= 1 - \sum_{i,j\in\mathbb{S}} \sum_{k\in\mathbb{S}} \mathbb{P}(\Lambda(i;1) = k)\mathbb{P}(\bar{\Lambda}(j;1) = k)\,\Pi_{i,j} \\
&\leq 1 - \zeta.
\end{aligned} \tag{3.2.3}$$

Let $t \geq 1$ be arbitrary. Then, $t = \lfloor t \rfloor + s$ for some $s \in [0,1)$. We have that

$$\mathbb{P}(\tau_{ij} > t) = \mathbb{P}(\tau_{ij} > t,\, \tau_{ij} > t-1) \leq \mathbb{P}(\Lambda(i;t) \neq \bar{\Lambda}(j;t),\, \tau_{ij} > t-1)\,.$$

Observe that $\{\tau_{ij} > t-1\} \in \bar{\mathscr{F}}_{t-1}$. Here, $\{\bar{\mathscr{F}}_t\}_{t\geq 0}$ stands for the natural filtration of the process $\{(\Lambda(i;t), \bar{\Lambda}(j;t))\}_{t\geq 0}$. This and eq. (3.2.3) imply

$$\begin{aligned}
&\mathbb{P}(\Lambda(i;t) \neq \bar{\Lambda}(j;t),\, \tau_{ij} > t-1) \\
&= \int_{\{\tau_{ij}>t-1\}} \mathbb{P}(\Lambda(i;t) \neq \bar{\Lambda}(j;t) \mid \bar{\mathscr{F}}_{t-1})\, d\mathbb{P} \\
&= \int_{\{\tau_{ij}>t-1\}} \mathbb{P}\left(\Lambda(\Lambda(i;t-1);1) \neq \bar{\Lambda}(\bar{\Lambda}(j;t-1);1)\right) d\mathbb{P} \\
&\leq (1-\zeta)\,\mathbb{P}(\tau_{ij} > t-1)\,.
\end{aligned}$$

Thus,

$$\mathbb{P}(\tau_{ij} > t) \leq (1-\zeta)\,\mathbb{P}(\tau_{ij} > t-1)\,.$$

Iterating this procedure we arrive at

$$\mathbb{P}(\tau_{ij} > t) \leq (1-\zeta)^{\lfloor t \rfloor} = \mathrm{e}^{-\vartheta \lfloor t \rfloor}$$

for all $t \geq 1$. However, it is clear that the relation holds for all $t \geq 0$. ■

We now move to discuss irreducibility and aperiodicity properties of (X,Λ).

Fix $i \in \mathbb{S}$ and consider the following stochastic differential equation (SDE):

$$\begin{aligned}
dX^{(i)}(x;t) &= b\big(X^{(i)}(x;t), i\big)dt + \sigma\big(X^{(i)}(x;t), i\big)dB(t) \\
X^{(i)}(x;0) &= x \in \mathbb{R}^d\,.
\end{aligned} \tag{3.2.4}$$

Assume the following:

$(\overline{\mathrm{A1}})$ for each $i \in \mathbb{S}$ and $x \in \mathbb{R}^d$ the SDE in eq. (2.3.1) admits a unique nonexplosive strong solution $\{X^{(i)}(x;t)\}_{t\geq 0}$ which has continuous sample paths and it is a temporally-homogeneous strong Markov process with transition kernel

$$p^{(i)}(t,x,dy) = \mathbb{P}(X^{(i)}(x;t)) \in dy),$$

$(\overline{\mathrm{A2}})$ there is $\Delta > 0$ such that for each $i \in \mathbb{S}$, $\inf_{x\in\mathbb{R}^d} \mathrm{q}_{ii}(x) > -\Delta$,

($\overline{\text{A3}}$) there are $\Delta, \alpha > 0$ such that for all $x, y \in \mathbb{R}^d$ and $i \in \mathbb{S}$,

$$\sum_{j \in \mathbb{S} \setminus \{i\}} |\mathrm{q}_{ij}(x) - \mathrm{q}_{ij}(y)| \le \Delta |x-y|^{\alpha}$$

,

($\overline{\text{A4}}$) there are $x_0 \in \mathbb{R}^d$ and $r_0 > 0$, such that

(i) there are $\alpha, \Delta, \delta > 0$, such that for all $x, y \in B_{r_0}(x_0)$ and $i \in \mathbb{S}$,

$$|b(x,i) - b(y,i)| + \|\sigma(x,i) - \sigma(y,i)\|_{\mathrm{HS}} \le \Delta |x-y|^{\alpha} \quad \text{and} \quad |\sigma(x,i)' y| \ge \delta |y|$$

,

(ii) for all $i \in \mathbb{S}$ and $x \in \mathbb{R}^d$,

$$\mathbb{P}\big(\tau^{(x,i)}_{B_{r_0}(x_0)} < \infty\big) > 0\,,$$

where $\tau^{(x,i)}_{B_{r_0}(x_0)} := \inf\{t \ge 0 : X^{(i)}(x;t) \in B_{r_0}(x_0)\}$.

↬ *Remark:* conditions from [FGC19] ensuring **(A1)-(A5)** imply also **($\overline{\text{A1}}$)**.

We now prove the main result of this section.

Theorem 3.2.3. Assume **(A1)-(A5)**, **($\overline{\text{A1}}$)-($\overline{\text{A4}}$)** and that for any $i, j \in \mathbb{S}$, $i \neq j$, there are $n \in \mathbb{N}$ and $k_0, \dots, k_n \in \mathbb{S}$ with $k_0 = i$, $k_n = j$ and $k_l \neq k_{l+1}$ for $l = 0, \dots, n-1$, such that the set $\{x \in B_{r_0}(x_0) \colon \mathrm{q}_{k_l k_{l+1}}(x) > 0\}$ has positive Lebesgue measure for all $l = 0, \dots, n-1$. Then the process $\{(X(x,i;t), \Lambda(x,i;t)\}_{t \ge 0}$ is open-set irreducible and aperiodic.

Proof. We show that for all $(x,i) \in \mathbb{R}^d \times \mathbb{S}$ and $j \in \mathbb{S}$,

$$\int_0^{\infty} p\big(t, (x,i), B \times \{j\}\big)\, dt > 0 \qquad \text{and} \qquad \sum_{m=0}^{\infty} p\big(m, (x,i), B \times \{j\}\big) > 0$$

whenever $\mathrm{Leb}(B \cap B_{r_0}(x_0)) > 0$. In other words, we show that $\{(X(x,i;t), \Lambda(x,i;t)\}_{t \ge 0}$ and the skeleton chain $\{(X(x,i;m), \Lambda(x,i;m)\}_{m \ge 0}$ are ϕ-irreducible with $\phi(\cdot) := \mathrm{Leb}(\cdot \cap B_{r_0}(x_0))$.

From the proof of [FGC19, Theorem 4.8] we have that

$$p\big(t,(x,i),B\times\{j\}\big) \,\ge\, \delta_{ij}\,\mathbb{P}\big(X^{(i)}(x;t)\in B\big)\,\mathrm{e}^{-\Delta t}$$
$$+\,(1-\delta_{ij})\,\mathrm{e}^{-\Delta t}\sum_{m=1}^{\infty}\int\!\!\cdots\!\!\int_{0<t_1<\ldots<t_m<t}\sum_{\substack{k_0,\ldots,k_m\in\mathbb{S}\\ k_l\neq k_{l+1}\\ k_0=i,\,k_m=j}}\int_{\mathbb{R}^d}\cdots\int_{\mathbb{R}^d}\mathbb{P}\big(X^{(k_0)}(x;t_1)\in dy_1\big)\,\mathrm{q}_{k_0k_1}(y_1)$$
$$\mathbb{P}\big(X^{(k_1)}(y_1;t_2-t_1)\in dy_2\big)\mathrm{q}_{k_1k_2}(y_2)\cdots\mathrm{q}_{k_{m-1}k_m}(y_m)\mathbb{P}\big(X^{(k_m)}(y_m;t-t_m)\in B\big)\,dt_1\ldots dt_m\,, \tag{3.2.5}$$

where δ_{ij} is the Kronecker delta. In particular,

$$p\big(t,(x,i),B\times\{j\}\big) \,\ge\, \delta_{ij}\,\mathbb{P}\big(X^{(i)}(x;t)\in B\big)\,\mathrm{e}^{-\Delta t}$$
$$+\,(1-\delta_{ij})\,\mathrm{e}^{-\Delta t}\int\!\!\cdots\!\!\int_{0<t_1<\ldots<t_m<t}\int_{B_{r_0}(x_0)}\cdots\int_{B_{r_0}(x_0)}\mathbb{P}\big(X^{(k_0)}(x;t_1)\in dy_1\big)\,\mathrm{q}_{k_0k_1}(y_1)$$
$$\mathbb{P}\big(X^{(k_1)}(y_1;t_2-t_1)\in dy_2\big)\,\mathrm{q}_{k_1k_2}(y_2)\cdots\mathrm{q}_{k_{m-1}k_m}(y_m)\mathbb{P}\big(X^{(k_m)}(y_m;t-t_m)\in B\big)\,dt_1\ldots dt_m\,,$$

where for $i\neq j$, $m\in\mathbb{N}$ and $k_0,\ldots,k_m\in\mathbb{S}$ are such that $k_0=i$, $k_m=j$, and $k_l\neq k_{l+1}$ and

$$\mathrm{Leb}\big(\{x\in B_{r_0}(x_0)\colon\,\mathrm{q}_{k_lk_{l+1}}(x)>0\}\big)\,>\,0$$

for $l=0,\ldots,m-1$. Next, assumptions of the theorem together with the proof of [LS21, Theorem 2.3] imply that for $B\in\mathscr{B}(\mathbb{R}^d)$ with $\mathrm{Leb}(B\cap B_{r_0}(x_0))>0$,

(i) $\mathbb{P}\big(X^{(i)}(x;t)\in B\big)>0$ for all $x\in B_{r_0}(x_0)$, $i\in\mathbb{S}$ and $t>0$

(ii) $\displaystyle\int_0^{\infty}\mathbb{P}\big(X^{(i)}(x;t)\in B\big)\,dt>0$ for all $i\in\mathbb{S}$ and $x\in\mathbb{R}^d$

(iii) $\displaystyle\sum_{m=0}^{\infty}\mathbb{P}\big(X^{(i)}(x;m)\in B\big)>0$ for all $i\in\mathbb{S}$ and $x\in\mathbb{R}^d$.

Thus, for all $(x,i)\in\mathbb{R}^d\times\mathbb{S}$, $j\in\mathbb{S}$ and $B\in\mathscr{B}(\mathbb{R}^d)$ with $\mathrm{Leb}(B\cap B_{r_0}(x_0))>0$,

$$\int_0^{\infty}p\big(t,(x,i),B\times\{j\}\big)\,dt\,>\,0\qquad\text{and}\qquad\sum_{m=0}^{\infty}p\big(m,(x,i),B\times\{j\}\big)\,>\,0\,,$$

which proves the assertion. ■

Let us now give several remarks.

Remark 3.2.4. (i) The conclusion of Theorem 3.2.3 also remains true in the case of infinitely countable state space $\mathbb{S}$ (say $\mathbb{S}=\mathbb{N}$) if, in addition to the assumptions of the theorem, there is $\gamma > 0$ such that for all $i, j \in \mathbb{S}$, $i \neq j$, $\sup_{x\in\mathbb{R}^d} \mathrm{q}_{ij}(x) \leq \gamma j 3^{-j}$. This additional assumption is required to conclude the relation in eq. (3.2.5) (see [FGC19, Lemma 4.7]).

(ii) The problem of open-set irreducibility and aperiodicity of regime-switching diffusion processes has already been considered in the literature (see [KZ20] and the references therein). In all these works a crucial assumption is uniform ellipticity of the matrix $\sigma(x,i)\sigma(x,i)^T$, that is,

$$\inf_{x\in\mathbb{R}^d, y\in\mathbb{R}^d\setminus\{0\}, i\in\mathbb{S}} \frac{|\sigma(x,i)^T y|}{|y|} > 0\,. \tag{3.2.6}$$

On the other hand, in Theorem 3.2.3 we require uniform ellipticity of $\sigma(x,i)\sigma(x,i)^T$ on the open ball $B_{r_0}(x_0)$ only, while on the rest of the state space it can degenerate provided $\mathbb{P}(\tau^{(x,i)}_{B_{r_0}(x_0)} < \infty) > 0$ for all $i \in \mathbb{S}$ and $x \in \mathbb{R}^d$.

(iii) According to [LS21, Proposition 2.4], $\mathbf{(\overline{A4})}$ (ii) will hold if $|\sigma(x,i)^T y| > 0$ for all $i \in \mathbb{S}$, $x \in \mathbb{R}^d$ and $y \in \mathbb{R}^d \setminus \{0\}$. Clearly, this condition is much weaker than (3.2.6). A simple example of a regime-switching diffusion process satisfying $\mathbf{(\overline{A4})}$ (as well as **(A1)-(A5)** and $\mathbf{(\overline{A1})}$-$\mathbf{(\overline{A3})}$) with degenerate $\sigma(x,i)\sigma(x,i)^T$ is given as follows. Let $\mathbb{S}_> \subseteq \mathbb{S}$, and let $b_1 \in \mathscr{C}^1(\mathbb{R}^d \times \mathbb{S}_>)$, $b_2(x,i) = (b_2^{(1)}(x,i), \dots, b_2^{(d)}(x,i))$ with $b_2^{(k)} \in \mathscr{C}^1(\mathbb{R}^d \times \mathbb{S}^c_>)$, $k = 1, \dots, d$, and $\sigma \in \mathscr{C}^1(\mathbb{R}^d \times \mathbb{S})$ be such that:

(a) $0 < \inf_{x\in\mathbb{R}^d} b_1(x,i) \leq \sup_{x\in\mathbb{R}^d} b_1(x,i) < \infty$ for all $i \in \mathbb{S}_>$;

(b) $\sup_{x\in\mathbb{R}^d\setminus\{0\}} \langle x, b_2(x,i)\rangle/|x|^2 < \infty$ for all $i \in \mathbb{S}^c_>$;

(c) for all $i \in \mathbb{S}_>$, $\sigma(x,i) > 0$ if, and only if, $x \in B_{r_i}(0)$ for some $r_i > 0$;

(d) $\sigma(x,i) > 0$ for all $x \in \mathbb{R}^d$ and $i \in \mathbb{S}^c_>$.

Define

$$b(x,i) := \begin{cases} -b_1(x,i)x\,, & i \in \mathbb{S}_>\,, \\ b_2(x,i)\,, & i \in \mathbb{S}^c_>\,, \end{cases}$$

and $\sigma(x,i) := \sigma(x,i)\,\mathbb{I}_d$, where $\mathbb{I}_d$ stands for the $d \times d$ identity matrix. It is clear that such a regime-switching diffusion process satisfies $\mathbf{(\overline{A4})}$ with $x_0 = 0$ and any $0 < r_0 < \min_{i\in\mathbb{S}_>} r_i$.

3.3. ERGODICITY IN THE TOTAL VARIATION DISTANCE

We start the discussion on ergodicity of a regime-switching diffusion by finding sufficient conditions for recurrence and transience. Such questions naturally appear when addressing the stability of the process as positive recurrence is a necessary condition for existence of an invariant probability measure.

Let us introduce the notation similar to the one in the previous chapter: Fix $x_0 \in \mathbb{R}^d$ and $r_0 \geq 0$, and put

$$\begin{aligned}
c(x,i) &:= \sigma(x,i)\sigma(x,i)^T, \qquad x \in \mathbb{R}^d, i \in \mathbb{S},\\
A(x,i) &:= (1/2)\mathrm{Tr}\, c(x,i), \qquad x \in \mathbb{R}^d, i \in \mathbb{S},\\
B_{x_0}(x,i) &:= \langle x - x_0, b(x,i)\rangle, \qquad x \in \mathbb{R}^d, i \in \mathbb{S},\\
C_{x_0}(x,i) &:= |x-x_0|^{-2}\langle x - x_0, c(x,i)(x-x_0)\rangle, \qquad x \in \mathbb{R}^d \setminus \{x_0\}, i \in \mathbb{S},\\
\underline{\gamma}_{x_0}(r) &:= \inf_{i\in\mathbb{S}} \inf_{|x-x_0|=r} C_{x_0}(x,i), \qquad r>0,\\
\overline{\gamma}_{x_0}(r) &:= \sup_{i\in\mathbb{S}} \sup_{|x-x_0|=r} C_{x_0}(x,i), \qquad r>0,\\
\overline{\iota}_{x_0}(r) &:= \sup_{i\in\mathbb{S}} \sup_{|x-x_0|=r} (2A(x,i) - C_{x_0}(x,i) + 2B_{x_0}(x,i))/C_{x_0}(x,i), \qquad r>0,\\
\overline{I}_{x_0,}(r) &:= \int_{r_0}^{r} \frac{\overline{\iota}_{x_0}(s)}{s} ds, \qquad r \geq r_0,\\
\underline{\iota}_{x_0}(r) &:= \inf_{i\in\mathbb{S}} \inf_{|x-x_0|=r} (2A(x,i) - C_{x_0}(x,i) + 2B_{x_0}(x,i))/C_{x_0}(x,i), \qquad r>0,\\
\underline{I}_{x_0,}(r) &:= \int_{r_0}^{r} \frac{\underline{\iota}_{x_0}(s)}{s} ds, \qquad r \geq r_0.
\end{aligned}$$

Now, similarly as in [Bha78], one obtains the following conditions ensuring recurrence and transience.

Theorem 3.3.1. Assume **(A1)-(A5)** and assume that the matrix $c(x,i)$ is positive definite for all $x \in \mathbb{R}^d$ and all $i \in \mathbb{S}$ so that all functions above are well defined. If there exist some $x_0 \in \mathbb{R}^d$ and $r_0 \geq 0$ such that

(i)

$$\int_{r_0}^{+\infty} e^{-\overline{I}_{x_0}(u)} du = +\infty,$$

then the diffusion is recurrent,

(ii)
$$\int_{r_0}^{+\infty} e^{-\underline{I}_{x_0}(u)} du < +\infty,$$
then the diffusion is transient.

Proof. (i) A sufficient condition for a process to be recurrent is to find a Lyapunov function $\mathscr{V}$ such that for all $x \in \mathbb{R}^d$ such that $|x - x_0| \geq r_0$ and all $i \in \mathbb{S}$

$$\mathscr{L}\mathscr{V}(x,i) \leq 0,$$

see [MT93b, Theorem 3.1.].

Define $\overline{\mathscr{V}} : [0,+\infty\rangle \longrightarrow [0,+\infty\rangle$ by

$$\overline{\mathscr{V}}(r) = \int_{r_0}^{r} e^{-\bar{I}_{x_0}(u)} du, \qquad r \geq r_0.$$

Then fix $r_1 > r_0$ and define $\mathscr{V} : \mathbb{R}^d \times \mathbb{S} \to [0,+\infty\rangle$ by $\mathscr{V}(x,i) = \overline{\mathscr{V}}(|x-x_0|) + 1$ for $x \in \mathbb{R}^d$ such that $|x - x_0| \geq r_1$, and for all other $x \in \mathbb{R}^d$ in such a way that $x \mapsto \mathscr{V}(x,i) \in C^2(\mathbb{R}^d)$, for all $i \in \mathbb{S}$. Since for any $i \in \mathbb{S}$ $\lim_{|x|\to+\infty} \mathscr{V}(x,i) = +\infty$, it holds that $\mathscr{V}$ is norm-like function, that is, a Lyapunov function. Note that the function $\mathscr{V}(x,i)$ does not depend on the value i if $|x - x_0| \geq r_1$. Since the sum of elements in a row of the matrix $\mathscr{Q}$ is equal to 0, for all $x \in \mathbb{R}^d$ such that $|x - x_0| \geq r_1$ and all $i \in \mathbb{S}$ we have:

$$\begin{aligned}
\mathscr{L}\mathscr{V}(x,i) &= \frac{1}{2} C_{x_0}(x,i) \overline{\mathscr{V}}''(|x-x_0|) + \frac{\overline{\mathscr{V}}'(|x-x_0|)}{2|x-x_0|}(2A(x,i) - C_{x_0}(x,i) + 2B_{x_0}(x,i)) \\
&\quad + \sum_{j \in S} \mathrm{q}_{i,j} \mathscr{V}(x,j) \\
&\leq \frac{C_{x_0}(x,i)}{2|x-x_0|}\Big[-\bar{\imath}_{x_0}(|x-x_0|) e^{-\bar{I}_{x_0}(|x-x_0|)} + \\
&\quad + e^{-\bar{I}_{x_0}(|x-x_0|)} \frac{2A(x,i) - C_{x_0}(x,i) + 2B_{x_0}(x,i)}{C_{x_0}(x,i)} \Big] + \mathscr{V}(x,i) \sum_{j \in S} \mathrm{q}_{i,j} \\
&\leq 0.
\end{aligned}$$

(ii) Define functions $\overline{\mathscr{V}}$ and $\mathscr{V}$ as above, but using $\underline{I}$ instead of $\bar{I}$. Then fix $r_1 > r_0$. So, for all $x \in \mathbb{R}^d$ such that $|x - x_0| \geq r_1$ and all $i \in \mathbb{S}$, it holds that $\overline{\mathscr{V}}'(|x - x_0|) > 0$, so we have that $\mathscr{L}\mathscr{V}(x,i) \geq 0$. Define

$$\tau_0 := \inf\{t > 0 : (X_t, \Lambda_t) \in B_{r_0}(x_0) \times \mathbb{S}\},$$
$$\tau_K := \inf\{t > 0 : |X_t| \geq K\}.$$

It holds that $\tau_K \to +\infty$ as $K \to +\infty$ because the process is non-explosive.

By the optional sampling,

$$\mathscr{V}(X_{t\wedge\tau_0\wedge\tau_K}, r_{t\wedge\tau_0\wedge\tau_K}) - \int_0^{t\wedge\tau_0\wedge\tau_K} \mathscr{L}\mathscr{V}(X_s, r_s)ds$$

is a $\mathbb{P}^{(x,i)}$-martingale, for $|x - x_0| < K$. Hence,

$$\mathbb{E}_{x,i}[\mathscr{V}(X_{t\wedge\tau_0\wedge\tau_K}, r_{t\wedge\tau_0\wedge\tau_K})] - \mathscr{V}(x,i) = \mathbb{E}_{x,i}\int_0^{t\wedge\tau_0\wedge\tau_K} \mathscr{L}\mathscr{V}(X_s, r_s)ds \geq 0.$$

Hence,

$$\mathbb{P}^{(x,i)}(\tau_0 > \tau_K)\int_{r_0}^{K} e^{-\underline{I}(u)}du \geq \int_{r_0}^{|x-x_0|} e^{-\underline{I}(u)}du. \tag{3.3.1}$$

Letting $K \to +\infty$ and using $\int_{r_0}^{+\infty} e^{-\underline{I}_{x_0}(u)}du < +\infty$, we obtain

$$\mathbb{P}_{x,i}(\tau_0 = +\infty) > 0,$$

so the process is transient.

■

We proceed by giving the illustrating this result on two examples.

Example 3.3.2. Consider a regime-switching diffusion process in one dimension where the Markov chain $\{\Lambda_t\}_{t\geq 0}$ takes two values, that is $S = \{1,2\}$, the diffusion coefficient $\sigma \equiv 1$ and the drift coefficient is as follows:

$$b(x,1) = -\frac{1}{2}\operatorname{sgn}(x)\cos(x),$$
$$b(x,2) = -\frac{1}{2}x.$$

Take $x_0 = 0$ and $r_0 = 1$. Then, for all $x \in \mathbb{R}$ and $i \in \mathbb{S}$, it holds: $C(x,i) = 1$, $A(x,i) = \frac{1}{2}$, $\underline{\gamma}(r) = \overline{\gamma}(r) = 1$,

$$\bar{\iota}(r) = \max_{i\in\mathbb{S}}\max_{|x|=r} 2B(x,i)$$

$$
\begin{aligned}
&= \max_{x=\pm r}\{-x\,\mathrm{sgn}(x)\cos(x), -x^2\} \\
&= \max\{-r\cos(r), -r^2\} = -r\cos(r), \qquad r > 1, \\
\bar{I}(r) &= \int_{r_0}^{r} \frac{\bar{\iota}(s)}{s} ds = -\int_1^r \cos(s) ds = \sin(1) - \sin(r).
\end{aligned}
$$

Now, we have

$$
\int_{r_0}^{+\infty} e^{-\bar{I}_{x_0}(u)} du = \int_1^{+\infty} e^{\sin(u)-\sin(1)} du = +\infty,
$$

so by Theorem 3.3.1 the diffusion process is recurrent.

Example 3.3.3. Consider now a similar diffusion process with random switching as in the previous example, but with the following drift coefficient:

$$
\begin{aligned}
b(x,1) &= -\frac{1}{2}\,\mathrm{sgn}(x)(\cos(x)-1), \\
b(x,2) &= x.
\end{aligned}
$$

Take again $x_0 = 0$ and $r_0 = 1$. Then, for all $x \in \mathbb{R}$ and $i \in \mathbb{S}$, it holds: $C(x,i) = 1$, $A(x,i) = \frac{1}{2}$, $\underline{\gamma}(r) = \overline{\gamma}(r) = 1$,

$$
\begin{aligned}
\underline{\iota}(r) &= \min_{i\in\mathbb{S}} \min_{|x|=r} 2B(x,i) \\
&= \min_{x=\pm r}\{-x\,\mathrm{sgn}(x)(\cos(x)-1), 2x^2\} \\
&= \min\{-r(\cos(r)-1), 2r^2\} = -r(\cos(r)-1), \qquad r > 1, \\
\underline{I}(r) &= \int_{r_0}^{r} \frac{\underline{\iota}(s)}{s} ds = -\int_1^r \cos(s) - 1\; ds = \sin(1) - 1 - \sin(r) + r.
\end{aligned}
$$

Hence,

$$
\int_{r_0}^{+\infty} e^{-\underline{I}_{x_0}(u)} du = \int_1^{+\infty} e^{\sin(u)-u-\sin(1)+1} du < +\infty,
$$

so by Theorem 3.3.1 the diffusion process is transient.

Since recurrence is not enough to conclude our process possesses an invariant probability measure, we need a stronger condition, namely, positive recurrence. In the spirit of the previous theorem, we continue by providing a sufficient condition for this property as well.

Theorem 3.3.4. Assume **(A1)-(A5)** and that the diffusion process is recurrent. If there exist $x_0 \in \mathbb{R}^d$ and $r_0 \geq 0$ such that

(i)

$$\int_{r_0}^{+\infty} \frac{e^{\bar{I}_{x_0}(u)}}{\underline{\gamma}_{x_0}(u)} du < +\infty, \tag{3.3.2}$$

and such that $\underline{\gamma}_{x_0}(r) > 0$ for all $r \geq r_0$, then the process in positive recurrent.

(ii)

$$\lim_{K \to +\infty} \frac{\int_{r_0}^{K} e^{-\bar{I}_{x_0}(s)} \int_{r_0}^{s} \frac{e^{\bar{I}_{x_0}(u)}}{\overline{\gamma}_{x_0}(u)} du\, ds}{\int_{r_0}^{K} e^{-\underline{I}_{x_0}(u)} du} = +\infty, \tag{3.3.3}$$

and such that $\overline{\gamma}_{x_0}(r) < +\infty$ for all $r \geq r_0$, then the process is null-recurrent.

Proof. (i) Let $\overline{\mathscr{V}} : [0, +\infty\rangle \longrightarrow [0, +\infty\rangle$ be such that

$$\overline{\mathscr{V}}(r) = \int_{r_0}^{r} e^{-\bar{I}_{x_0}(u)} \int_{u}^{+\infty} \frac{e^{\bar{I}_{x_0}(u)}}{\underline{\gamma}_{x_0}(u)} du, \qquad r \geq r_0.$$

Similarly as before, fix $r_1 > r_0$ and define $\mathscr{V} : \mathbb{R}^d \times \mathbb{S} \to [0, +\infty\rangle$ by $\mathscr{V}(x,i) = \overline{\mathscr{V}}(|x - x_0|) + 1$ for $x \in \mathbb{R}^d$ such that $|x - x_0| \geq r_1$, and for all other $x \in \mathbb{R}^d$ in such a way that the function $x \mapsto \mathscr{V}(x,i) \in C^2(\mathbb{R}^d)$, for all $i \in \mathbb{S}$. Because of (3.3.2), the function V is well defined. Since value $\mathscr{V}(x,i)$ does not depend on the value i if $|x - x_0| \geq r_1$ and the sum of elements in a row of $\mathscr{Q}$ is 0, we obtain the following for $|x - x_0| \geq r_1$ and all $i \in \mathbb{S}$:

$$\begin{aligned}
\mathscr{L}\mathscr{V}(x,i) &= \frac{1}{2} C_{x_0}(x,i) \overline{\mathscr{V}}''(|x - x_0|) + \frac{\overline{\mathscr{V}}'(|x - x_0|)}{2|x - x_0|} (2A(x,i) - C_{x_0}(x,i) + 2B_{x_0}(x,i)) \\
&\quad + \sum_{j \in S} \mathrm{q}_{i,j} \mathscr{V}(x,j) \\
&= -\frac{1}{2} C_{x_0}(x,i) \frac{\bar{\iota}_{x_0}(|x - x_0|)}{|x - x_0|} e^{-\bar{I}_{x_0}(|x - x_0|)} \int_{|x - x_0|}^{+\infty} \frac{e^{\bar{I}_{x_0}(u)}}{\underline{\gamma}_{x_0}(u)} du \\
&\quad - \frac{1}{2} C_{x_0}(x,i) e^{-\bar{I}_{x_0}(|x - x_0|)} \frac{e^{\bar{I}_{x_0}(|x - x_0|)}}{\underline{\gamma}_{x_0}(|x - x_0|)} \\
&\quad + \frac{e^{-\bar{I}_{x_0}(|x - x_0|)}}{2|x - x_0|} \cdot \\
&\quad \cdot \int_{|x - x_0|}^{+\infty} \frac{e^{\bar{I}_{x_0}(u)}}{\underline{\gamma}_{x_0}(u)} du \, \frac{2A(x,i) - C_{x_0}(x,i) + 2B_{x_0}(x,i)}{C_{x_0}(x,i)} C_{x_0}(x,i) \\
&\quad + \mathscr{V}(x,i) \sum_{j \in S} \mathrm{q}_{i,j} \\
&\leq -\frac{1}{2} C_{x_0}(x,i) \frac{\bar{\iota}_{x_0}(|x - x_0|)}{|x - x_0|} e^{-\bar{I}_{x_0}(|x - x_0|)} \int_{|x - x_0|}^{+\infty} \frac{e^{\bar{I}_{x_0}(u)}}{\underline{\gamma}_{x_0}(u)} du
\end{aligned}$$

$$
\begin{aligned}
&+\frac{e^{-\bar{I}_{x_0}(|x-x_0|)}}{2|x-x_0|}\int_{|x-x_0|}^{+\infty}\frac{e^{\bar{I}_{x_0}(u)}}{\underline{\gamma}_{x_0}(u)}du\,\bar{\iota}_{x_0}(|x-x_0|)C_{x_0}(x,i)\\
&-\frac{1}{2}\frac{C_{x_0}(x,i)}{\underline{\gamma}_{x_0}(|x-x_0|)}\\
&+\mathscr{V}(x,i)\sum_{j\in S}\mathrm{q}_{i,j}\\
\leq\ & 0-\frac{1}{2}+0=-\frac{1}{2},
\end{aligned}
$$

hence by [MT93b, Theorem 4.2.], our process is positive recurrent.

(ii) This time, define $\overline{\mathscr{V}}:[0,+\infty\rangle\longrightarrow[0,+\infty\rangle$ by

$$
\overline{\mathscr{V}}(r)=\int_{r_0}^{r}e^{-\bar{I}_{x_0}(s)}\int_{r_0}^{s}\frac{e^{\bar{I}_{x_0}(u)}}{\overline{\gamma}_{x_0}(u)}du,\qquad r\geq r_0,
$$

and, for a fixed $r_1>r_0$, define function $\mathscr{V}:\mathbb{R}^d\times\mathbb{S}\to[0,+\infty\rangle$ by $\mathscr{V}(x,i)=\overline{\mathscr{V}}(|x-x_0|)+1$ for $x\in\mathbb{R}^d$ such that $|x-x_0|\geq r_1$, and for all other $x\in\mathbb{R}^d$ so that $x\mapsto\mathscr{V}(x,i)\in C^2(\mathbb{R}^d)$, for all $i\in\mathbb{S}$. Now, for $|x-x_0|\geq r_1$ and all $i\in\mathbb{S}$:

$$
\begin{aligned}
\mathscr{L}\mathscr{V}(x,i)=&-\frac{1}{2}C_{x_0}(x,i)\frac{\bar{\iota}_{x_0}(|x-x_0|)}{|x-x_0|}e^{-\bar{I}_{x_0}(|x-x_0|)}\int_{r_0}^{|x-x_0|}\frac{e^{\bar{I}_{x_0}(u)}}{\overline{\gamma}_{x_0}(u)}du\\
&+\frac{1}{2}C_{x_0}(x,i)e^{-\bar{I}_{x_0}(|x-x_0|)}\frac{e^{\bar{I}_{x_0}(|x-x_0|)}}{\overline{\gamma}_{x_0}(|x-x_0|)}\\
&+\frac{e^{-\bar{I}_{x_0}(|x-x_0|)}}{2|x-x_0|}\cdot\\
&\cdot\int_{r_0}^{|x-x_0|}\frac{e^{\bar{I}_{x_0}(u)}}{\overline{\gamma}_{x_0}(u)}du\,\frac{2A(x,i)-C_{x_0}(x,i)+2B_{x_0}(x,i)}{C_{x_0}(x,i)}C_{x_0}(x,i)\\
&+\mathscr{V}(x,i)\sum_{j\in S}\mathrm{q}_{i,j}\\
\leq\ & 0+\frac{1}{2}\frac{C_{x_0}(x,i)}{\overline{\gamma}_{x_0}(|x-x_0|)}+0\\
\leq\ & \frac{1}{2}.
\end{aligned}
$$

Now we proceed similarly as in the proof of Theorem 3.3.1: first, we define stopping times τ_0 and τ_K and obtain, for $x\in\mathbb{R}^d$ such that $r_1\leq|x-x_0|\leq K$ and $i\in\mathbb{S}$,

$$
2\mathbb{E}_{x,i}[V(X_{t\wedge\tau_0\wedge\tau_K},r_{t\wedge\tau_0\wedge\tau_K})]-2\mathscr{V}(x,i)=2\mathbb{E}_{x,i}\int_0^{t\wedge\tau_0\wedge\tau_K}\mathscr{L}\mathscr{V}(X_s,r_s)ds
$$

$$\leq \mathbb{E}_{x,i}[t \wedge \tau_0 \wedge \tau_K],$$

so by letting $t \to +\infty$, we get:

$$\begin{aligned}
\mathbb{E}_{x,i}[\tau_0 \wedge \tau_K] &\geq 2\mathbb{E}_{x,i}[V(X_{\tau_0\wedge\tau_K}, r_{\tau_0\wedge\tau_K})(\mathbb{1}_{\{\tau_N<\tau_0\}} + \mathbb{1}_{\{\tau_N\geq\tau_0\}})] - 2\mathscr{V}(x,i) \\
&\geq 2\mathbb{E}_{x,i}[V(X_{\tau_0\wedge\tau_K}, r_{\tau_0\wedge\tau_K})\mathbb{1}_{\{\tau_N<\tau_0\}}] - 2\mathscr{V}(x,i) \\
&= 2\overline{\mathscr{V}}(K)\mathbb{P}_{x,i}(\tau_N < \tau_0) - 2\mathscr{V}(x,i) \\
&\geq 2\overline{\mathscr{V}}(K)\frac{\int_{r_0}^{|x-x_0|} e^{-\underline{I}(u)}du}{\int_{r_0}^{K} e^{-\underline{I}(u)}du} - 2\mathscr{V}(x,i),
\end{aligned}$$

where in the last step we used (3.3.1). Finally, by letting $K \to +\infty$ and using (3.3.3), we get:

$$\mathbb{E}_{x,i}[\tau_0] \geq 2\int_{r_0}^{|x-x_0|} e^{-\underline{I}(u)}du \lim_{K\to+\infty} \frac{\overline{\mathscr{V}}(K)}{\int_{r_0}^{K} e^{-\underline{I}(u)}du} - 2\mathscr{V}(x,i) = +\infty,$$

so by [ST97, Theorem 4.1.] the process is null-recurrent.

■

Example 3.3.5. Let $\sigma \equiv 1$ and the drift coefficient to be

$$\begin{aligned}
b(x,1) &= -\frac{1}{2}\operatorname{sgn}(x)\left(\cos(x) + \frac{1}{\sqrt{|x|}}\right), \\
b(x,2) &= -x.
\end{aligned}$$

Take again $x_0 = 0$ and $r_0 = 1$. Then, for all $x \in \mathbb{R}$ and $i \in \mathbb{S}$, it holds: $C(x,i) = 1$, $A(x,i) = \frac{1}{2}$, $\overline{\gamma}(r) = \underline{\gamma}(r) = 1$,

$$\begin{aligned}
\overline{\iota}(r) &= \max_{i\in\mathbb{S}} \max_{|x|=r} 2B(x,i) \\
&= \max_{x=\pm r}\left\{-x\operatorname{sgn}(x)\left(\cos(x) + \frac{1}{\sqrt{|x|}}\right), -2x^2\right\} \\
&= \max\left\{-r\left(\cos(r) + \frac{1}{\sqrt{r}}\right), -2r^2\right\} = -r\cos(r) - \sqrt{r}, \qquad r > 1, \\
\overline{I}(r) &= \int_{r_0}^{r} \frac{\overline{\iota}(s)}{s}ds = -\int_1^r \cos(s) + \frac{1}{\sqrt{s}}\,ds = \sin(1) + 2 - \sin(r) - 2\sqrt{r}.
\end{aligned}$$

Firstly, we have:

$$\int_{r_0}^{+\infty} e^{-\overline{I}_{x_0}(u)}du = \int_1^{+\infty} e^{\sin(u)+2\sqrt{u}-\sin(1)-2}du = +\infty,$$

so by Theorem 3.3.1 we conclude that the diffusion process is recurrent. Secondly,

$$\int_{r_0}^{+\infty} \frac{e^{\bar{I}_{x_0}(u)}}{\underline{\gamma}(u)} du = \int_1^{+\infty} e^{\sin(u)-2\sqrt{u}-\sin(1)+2} du < +\infty,$$

so (3.3.2) holds and by Theorem 3.3.4 we conclude that the process is positive recurrent.

Example 3.3.6. Take again $\sigma \equiv 1$ and

$$\begin{aligned} b(x,1) &= -\frac{1}{2}\operatorname{sgn}(x)\cos(x), \\ b(x,2) &= -\frac{1}{2}\operatorname{sgn}(x). \end{aligned}$$

Take $x_0 = 0$ and $r_0 = 1$. Then, for all $x \in \mathbb{R}$ and $i \in \mathbb{S}$, it holds: $C(x,i) = 1$, $A(x,i) = \frac{1}{2}$, $\overline{\gamma}(r) = \underline{\gamma}(r) = 1$,

$$\begin{aligned} \bar{\iota}(r) &= \max_{i\in\mathbb{S}} \max_{|x|=r} 2B(x,i) \\ &= \max_{x=\pm r}\{-x\operatorname{sgn}(x)\cos(x), -x\operatorname{sgn}(x)\} \\ &= \max\{-r\cos(r), -r\} = -r\cos(r), \\ \bar{I}(r) &= \int_{r_0}^r \frac{\bar{\iota}(s)}{s} ds = -\int_1^r \cos(s) ds = \sin(1) - \sin(r), \qquad r > 1, \\ \underline{\iota}(r) &= \min_{i\in\mathbb{S}} \min_{|x|=r} 2B(x,i) \\ &= \min_{x=\pm r}\{-x\operatorname{sgn}(x)\cos(x), -x\operatorname{sgn}(x)\} \\ &= \min\{-r\cos(r), -r\} = -r, \\ \underline{I}(r) &= \int_{r_0}^r \frac{\bar{\iota}(s)}{s} ds = \int_1^r -1 ds = -r+1, \qquad r > 1, \end{aligned}$$

Now, we have

$$\int_{r_0}^{+\infty} e^{-\bar{I}_{x_0}(u)} du = \int_1^{+\infty} e^{\sin(u)-\sin(1)} du = +\infty,$$

so by Theorem 3.3.1 the diffusion process is recurrent. Furthermore,

$$\lim_{K\to+\infty} \frac{\int_{r_0}^K e^{-\bar{I}_{x_0}(s)} \int_{r_0}^s \frac{e^{\bar{I}_{x_0}(u)}}{\overline{\gamma}_{x_0}(u)} du\, ds}{\int_{r_0}^K e^{-\underline{I}_{x_0}(u)} du} = \lim_{K\to+\infty} \frac{\int_1^K e^{-\sin(r)+\sin(1)} \int_1^r e^{\sin(u)-\sin(1)} du\, dr}{\int_1^K e^{r-1} dr} = +\infty,$$

so by Theorem 3.3.4 the process is null-recurrent.

We now address the main question of this chapter: sub-geometric ergodicity with respect to the total variation distance. As we have seen, the regime-switching process might have some unusual behaviour. Namely, since the process is basically made up as a combination of standard diffusions, one might think it will inherit their properties. So, if all diffusions are ergodic, the regime-switched process will be as well. However, this might not be true. We first consider this special case and find condition that ensures ergodicity of such a process.

Theorem 3.3.7. Assume **(A1)-(A5)** and that $(X_t,\Lambda_t)_{t\geq 0}$ is open-set irreducible and aperiodic. Further, let $\varphi:[1,+\infty\rangle\longrightarrow\langle 0,+\infty\rangle$ be a non-decreasing, differentiable and concave function satisfying $\lim_{t\nearrow+\infty}\varphi'(t)=0$ and

$$\Lambda := \int_{r_0}^{+\infty}\varphi\left(\int_{r_0}^{u}e^{-\bar{I}_{x_0}(v)}dv+1\right)\frac{e^{\bar{I}_{x_0}(u)}}{\underline{\gamma}_{x_0}(u)}du<+\infty, \tag{3.3.4}$$

for some $x_0\in\mathbb{R}^d$ and $r_0\geq 0$ such that $\underline{\gamma}_{x_0}(r)>0$ for all $r\geq r_0$ (hence, the above functions are well defined). Then, $(X_t,\Lambda_t)_{t\geq 0}$ admits a unique invariant $\pi\in\mathscr{P}$ satisfying

$$\lim_{t\nearrow+\infty}\varphi(\Phi_t^{-1})\|\delta_{x,k}P_t-\pi\|_{TV}=0,\qquad x\in\mathbb{R}^d,k\in S$$

where

$$\Phi_t := \int_1^t\frac{ds}{\varphi(s)},\qquad t\geq 1.$$

Proof. Let $\overline{\mathscr{V}}:[0,+\infty\rangle\longrightarrow[0,+\infty\rangle$ be such that

$$\overline{\mathscr{V}}(r)=\int_{r_0}^{r}e^{-\bar{I}_{x_0}(u)}\int_u^{+\infty}\varphi_\Lambda\left(\int_{r_0}^{v}e^{-\bar{I}_{x_0}(w)}dw+1\right)\frac{e^{\bar{I}_{x_0}(v)}}{\underline{\gamma}_{x_0}(v)}dvdu,\qquad r\geq r_0,$$

where $\varphi_\Lambda(t):=\varphi(t)/\Lambda$. So, the function $\overline{\mathscr{V}}$ does not depend on the state i of Markov chain r and the following inequality holds for all $r\geq r_0$:

$$\overline{\mathscr{V}}(r)\leq\int_{r_0}^{r}e^{-\bar{I}_{x_0}(u)}du.$$

Now, fix $r_1>r_0$ and define a function $\mathscr{V}:\mathbb{R}^d\times\mathbb{S}\to[0,+\infty\rangle$ by $\mathscr{V}(x,i)=\overline{\mathscr{V}}(|x-x_0|)+1$ for $x\in\mathbb{R}^d$ such that $|x-x_0|\geq r_1$, and for all other $x\in\mathbb{R}^d$ so that $x\mapsto\mathscr{V}(x,i)\in C^2(\mathbb{R}^d)$, for all $i\in\mathbb{S}$. Note that the function $\mathscr{V}(x,i)$ does not depend on the value i if $|x-x_0|\geq r_1$. Secondly, note that the sum of elements in each row on the matrix $\mathscr{Q}$ is equal to 0 (since it is the generator matrix). Then, for all $x\in\mathbb{R}^d$ such that $|x-x_0|\geq r_1$ and all $i\in\mathbb{S}$,

$$
\begin{aligned}
\mathscr{L}\mathscr{V}(x,i) &= \frac{1}{2}C_{x_0}(x,i)\overline{\mathscr{V}}''(|x-x_0|) + \frac{\overline{\mathscr{V}}'(|x-x_0|)}{2|x-x_0|}(2A(x,i) - C_{x_0}(x,i) + 2B_{x_0}(x,i)) \\
&\quad + \sum_{j\in S} \mathrm{q}_{i,j}\mathscr{V}(x,j) \\
&= -\frac{1}{2}C_{x_0}(x,i)\frac{\overline{\iota}_{x_0}(|x-x_0|)}{|x-x_0|}e^{-\overline{I}_{x_0}(|x-x_0|)}\int_{|x-x_0|}^{+\infty}\varphi_\Lambda\left(\int_{r_0}^{u}e^{-\overline{I}_{x_0}(v)}dv+1\right)\frac{e^{\overline{I}_{x_0}(u)}}{\underline{\gamma}_{x_0}(u)}du \\
&\quad - \frac{1}{2}C_{x_0}(x,i)\frac{\varphi_\Lambda\left(\int_{r_0}^{|x-x_0|}e^{-\overline{I}_{x_0}(u)}du+1\right)}{\underline{\gamma}_{x_0}(|x-x_0|)} \\
&\quad + \frac{e^{-\overline{I}_{x_0}(|x-x_0|)}}{2|x-x_0|}\int_{|x-x_0|}^{+\infty}\varphi_\Lambda\left(\int_{r_0}^{u}e^{-\overline{I}_{x_0}(v)}dv+1\right)\frac{e^{\overline{I}_{x_0}(u)}}{\underline{\gamma}_{x_0}(u)}du\,\cdot \\
&\qquad \cdot\frac{2A(x,i) - C_{x_0}(x,i) + 2B_{x_0}(x,i)}{C_{x_0}(x,i)}C_{x_0}(x,i) \\
&\quad + \sum_{j\in S} \mathrm{q}_{i,j}\mathscr{V}(x,j),
\end{aligned}
$$

so

$$
\begin{aligned}
\mathscr{L}\mathscr{V}(x,i) &\le -\frac{1}{2}C_{x_0}(x,i)\frac{\overline{\iota}_{x_0}(|x-x_0|)}{|x-x_0|}e^{-\overline{I}_{x_0}(|x-x_0|)}\int_{|x-x_0|}^{+\infty}\varphi_\Lambda\left(\int_{r_0}^{u}e^{-\overline{I}_{x_0}(v)}dv+1\right)\frac{e^{\overline{I}_{x_0}(u)}}{\underline{\gamma}_{x_0}(u)}du \\
&\quad + \frac{e^{-\overline{I}_{x_0}(|x-x_0|)}}{2|x-x_0|}\,\cdot \\
&\qquad \cdot\int_{|x-x_0|}^{+\infty}\varphi_\Lambda\left(\int_{r_0}^{u}e^{-\overline{I}_{x_0}(v)}dv+1\right)\frac{e^{\overline{I}_{x_0}(u)}}{\underline{\gamma}_{x_0}(u)}du\,\overline{\iota}_{x_0}(|x-x_0|)C_{x_0}(x,i) \\
&\quad - \frac{1}{2}C_{x_0}(x,i)\frac{\varphi_\Lambda\left(\int_{r_0}^{|x-x_0|}e^{-\overline{I}_{x_0}(u)}du+1\right)}{\underline{\gamma}_{x_0}(|x-x_0|)} \\
&\quad + \sum_{j\in S} \mathrm{q}_{i,j}\mathscr{V}(x,j) \\
&\le 0 - \frac{1}{2}\varphi_\Lambda\left(\int_{r_0}^{|x-x_0|}e^{-\overline{I}_{x_0}(u)}du+1\right) + \mathscr{V}(x,i)\sum_{j\in S}\mathrm{q}_{i,j} \\
&= -\frac{1}{2}\varphi_\Lambda(\mathscr{V}(x,i)) + 0 \\
&= -\frac{1}{2}\varphi_\Lambda(\mathscr{V}(x,i)),
\end{aligned}
$$

that is, we have established sub-geometric drift condition and conclude that the diffusion is sub-geometrically ergodic. ■

↬ *Remarks:* the condition (3.3.4) for sub-geometric ergodicity considers the worst case for all diffusion processes $\{X^{(i)}(x;t)\}_{t\geq 0}$ for a fixed starting position $x \in \mathbb{R}^d$. Namely, all those diffusions are ergodic.

- The ergodicity condition (3.3.4) does not include any information about the matrix $\mathcal{Q}$. Hence, the ergodicity depends on $\mathcal{Q}$ only through the condition that the process is irreducible and aperiodic.

Example 3.3.8. Take now a switching diffusion process with $\sigma \equiv 1$ and the drift coefficient $b(x,i)$ as follows:

$$b(x,1) = -\frac{1}{2}\operatorname{sgn}(x)(\cos(x)+1),$$
$$b(x,2) = -\operatorname{sgn}(x).$$

Take $x_0 = 0$ and $r_0 = 1$. Then, for all $x \in \mathbb{R}$ and $i \in \mathbb{S}$, it holds: $C(x,i) = 1$, $A(x,i) = \frac{1}{2}$, $\overline{\gamma}(r) = \underline{\gamma}(r) = 1$,

$$\begin{aligned}
\overline{\iota}(r) &= \max_{i\in\mathbb{S}} \max_{|x|=r} 2B(x,i) \\
&= \max_{x=\pm r}\{-x\operatorname{sgn}(x)(\cos(x)+1), -2x\operatorname{sgn}(x)\} \\
&= \max\{-r(\cos(r)+1), -2r\} = -r(\cos(r)+1), \qquad r > 1, \\
\overline{I}(r) &= \int_{r_0}^{r} \frac{\overline{\iota}(s)}{s} ds = -\int_{1}^{r} \cos(s)+1\, ds = \sin(1)+1-\sin(r)-r.
\end{aligned}$$

Consider a function $\varphi(t) = t^{\alpha}$, for $\alpha > 0$. Then, we have

$$\begin{aligned}
\Lambda &= \int_{r_0}^{+\infty} \frac{e^{\overline{I}(u)}}{\underline{\gamma}(u)} \left(\int_{r_0}^{u} e^{-\overline{I}(v)} dv + 1\right)^{\alpha} du \\
&= \int_{1}^{+\infty} e^{-\sin(u)-u+\sin(1)+1} \left(\int_{1}^{u} e^{\sin(v)+v-\sin(1)-1} dv + 1\right)^{\alpha} du < +\infty,
\end{aligned}$$

which is true for any $0 < \alpha < 1$. So, by Theorem 3.3.7, the diffusion process is subgeometrically ergodic and the rate of the convergence is polynomial - it equals $t^{\alpha(1-\alpha)}$.

A special case is $\varphi(t) = t$ - corresponding to geometric ergodicity. In this situation we can impose weaker conditions to obtain ergodicity.

Theorem 3.3.9. Assume **(A1)-(A5)** and assume that $(X_t, \Lambda_t)_{t\geq 0}$ is open-set irreducible and aperiodic. Furthermore, assume

(i)

$$\int_{r_0}^{+\infty} \frac{e^{\bar{I}_{x_0}(u)}}{\underline{\gamma}_{x_0}(u)} du < +\infty, \tag{3.3.5}$$

(ii)

$$\sup_{r>r_0} \int_{r_0}^{r} e^{-\bar{I}_{x_0}(u)} du \int_{r}^{+\infty} \frac{e^{\bar{I}_{x_0}(u)}}{\underline{\gamma}_{x_0}(u)} du < +\infty, \tag{3.3.6}$$

for some $x_0 \in \mathbb{R}^d$ and $r_0 \geq 0$ such that $\underline{\gamma}_{x_0}(r) > 0$ for all $r \geq r_0$. Then, $(X_t, \Lambda_t)_{t\geq 0}$ admits a unique invariant $\pi \in \mathscr{P}$ satisfying

$$\lim_{t \nearrow +\infty} e^t \|\delta_{x,k} P_t - \pi\|_{TV} = 0, \qquad x \in \mathbb{R}^d, k \in S$$

that is, $(X_t, \Lambda_t)_{t\geq 0}$ is geometrically ergodic.

Proof. Similarly as before, define $\overline{\mathscr{V}} : [0, +\infty\rangle \longrightarrow [0, +\infty\rangle$ be such that

$$\overline{\mathscr{V}}(r) = \int_{r_0}^{r} e^{-\bar{I}_{x_0}(u)} \int_{u}^{+\infty} \left(\int_{r_0}^{v} e^{-\bar{I}_{x_0}(w)} dw + 1 \right)^p \frac{e^{\bar{I}_{x_0}(v)}}{\underline{\gamma}_{x_0}(v)} dv du, \qquad r \geq r_0,$$

where $p \in \langle 0, 1\rangle$ is fixed. So, the function $\overline{\mathscr{V}}$ does not depend on the state i of Markov chain r.

Again, fix $r_1 > r_0$ and define a function $\mathscr{V} : \mathbb{R}^d \times \mathbb{S} \to [0, +\infty\rangle$ by $\mathscr{V}(x,i) = \overline{\mathscr{V}}(|x - x_0|) + 1$ for $x \in \mathbb{R}^d$ such that $|x - x_0| \geq r_1$, and for all other $x \in \mathbb{R}^d$ so that $x \mapsto \mathscr{V}(x,i) \in C^2(\mathbb{R}^d)$, for all $i \in \mathbb{S}$. So, the function $\mathscr{V}(x,i)$ does not depend on the value i if $|x - x_0| \geq r_1$ and the sum of elements in each row on the matrix $\mathscr{Q}$ is equal to 0. Then, for all $x \in \mathbb{R}^d$ such that $|x - x_0| \geq r_1$ and all $i \in \mathbb{S}$,

$$\begin{aligned}
\mathscr{L}\mathscr{V}(x,i) &= \frac{1}{2} C_{x_0}(x,i) \overline{\mathscr{V}}''(|x - x_0|) + \frac{\overline{\mathscr{V}}'(|x - x_0|)}{2|x - x_0|}(2A(x,i) - C_{x_0}(x,i) + 2B_{x_0}(x,i)) \\
&\quad + \sum_{j \in S} \mathrm{q}_{i,j} \mathscr{V}(x,j) \\
&\leq -\frac{1}{2} \frac{C_{x_0}(x,i)}{\underline{\gamma}_{x_0}(|x - x_0|)} \left(\int_{r_0}^{|x - x_0|} e^{-\bar{I}_{x_0}(u)} du + 1 \right)^p + 0 \\
&\leq -\frac{1}{2} \frac{\left(\int_{r_0}^{|x - x_0|} e^{-\bar{I}_{x_0}(u)} du + 1 \right)^p}{\mathscr{V}(x,i)} \mathscr{V}(x,i).
\end{aligned}$$

We will now show that

$$\inf_{r > r_0} \frac{\left(\int_{r_0}^{r} e^{-\bar{I}_{x_0}(u)} du + 1 \right)^p}{\overline{\mathscr{V}}(r)} > 0. \tag{3.3.7}$$

It holds:

$$\inf_{r>r_0}\frac{\left(\int_{r_0}^r e^{-\bar{I}_{x_0}(u)}du+1\right)^p}{\overline{\mathscr{V}}(r)} = \inf_{r>r_0}\frac{\left(\int_{r_0}^r e^{-\bar{I}_{x_0}(u)}du+1\right)^p}{\int_{r_0}^r e^{-\bar{I}_{x_0}(u)}\int_u^{+\infty}\left(\int_{r_0}^v e^{-\bar{I}_{x_0}(w)}dw+1\right)^p\frac{e^{\bar{I}_{x_0}(v)}}{\underline{\gamma}_{x_0}(v)}dvdu}$$

$$\geq \inf_{r>r_0}\frac{p\left(\int_{r_0}^r e^{-\bar{I}_{x_0}(u)}du+1\right)^{p-1}e^{-\bar{I}(r)}}{e^{-\bar{I}(r)}\int_r^{+\infty}\left(\int_{r_0}^u e^{-\bar{I}_{x_0}(v)}dv+1\right)^p\frac{e^{\bar{I}_{x_0}(u)}}{\underline{\gamma}_{x_0}(u)}du}$$

$$= \inf_{r>r_0}\frac{p\left(\int_{r_0}^r e^{-\bar{I}_{x_0}(u)}du+1\right)^{p-1}}{\int_r^{+\infty}\left(\int_{r_0}^u e^{-\bar{I}_{x_0}(v)}dv+1\right)^p\frac{e^{\bar{I}_{x_0}(u)}}{\underline{\gamma}_{x_0}(u)}du},$$

where we employed Cauchy's mean value theorem. Now, we use [Che00, Lemma 1.2] and obtain

$$\int_r^{+\infty}\left(\int_{r_0}^u e^{-\bar{I}_{x_0}(v)}dv+1\right)^p\frac{e^{\bar{I}_{x_0}(u)}}{\underline{\gamma}_{x_0}(u)}du \leq C\left(\int_{r_0}^r e^{-\bar{I}_{x_0}(u)}du+1\right)^{p-1}, \quad \text{for all } r>r_0,$$

for some constant $C>0$, hence (3.3.7) holds.

■

As an upgrade of the above result for sub-geometric ergodicity, in the following theorem we do not require nice behaviour from all separate diffusions $X^{(i)}$.

Theorem 3.3.10. Let $\{(X(x,i;t),\Lambda(x,i;t)\}_{t\geq 0}$ be an open-set irreducible and aperiodic regime-switching diffusion process satisfying **(A1)** – **(A5)**. Assume

(i) there are $\{c_i\}_{i\in\mathbb{S}}\subseteq\mathbb{R}$, twice continuously differentiable $\mathscr{V}:\mathbb{R}^d\to\langle 1,+\infty\rangle$ and twice continuously differentiable nondecreasing concave $\varphi:\langle 1,+\infty\rangle\to\langle 0,+\infty\rangle$, such that

$$\lim_{u\to+\infty}\varphi'(u)=0, \qquad \limsup_{|x|\to+\infty}\frac{\mathscr{L}_i\mathscr{V}(x)}{\varphi\circ\mathscr{V}(x)}<c_i,$$
$$\lim_{|x|\to+\infty}\frac{\varphi\circ\mathscr{V}(x)}{\mathscr{V}(x)}=0, \qquad \lim_{|x|\to+\infty}\sup_{i\in\mathbb{S}}\frac{\mathscr{L}_i\varphi\circ\mathscr{V}(x)}{\varphi\circ\mathscr{V}(x)}=0$$

(ii) $\mathscr{Q}(x)=\mathscr{Q}+\mathrm{o}(1)$ as $|x|\to+\infty$[1], where $\mathscr{Q}=(\mathrm{q}_{ij})_{i,j\in\mathbb{S}}$ is the infinitesimal generator of an irreducible right-continuous time-homogeneous Markov chain on $\mathbb{S}$ with invariant probability measure $\lambda=(\lambda_i)_{i\in\mathbb{S}}$

[1] We use the standard o notation: for $h:\mathbb{R}^p\to\mathbb{R}^q$ we write $h(x)=\mathrm{o}(1)$ as $|x|\to+\infty$ if, and only if, $\lim_{|x|\to+\infty}h(x)$ is the zero function.

(iii) $\sum_{i\in\mathbb{S}} c_i\lambda_i < 0$.

Then, $\{(X(x,i;t),\Lambda(x,i;t)\}_{t\geq 0}$ admits a unique invariant probability measure π and

$$\lim_{t\to+\infty} r(t)\|\delta_{(x,i)}\mathscr{P}_t - \pi\|_{\mathrm{TV}} = 0$$

for all $(x,i)\in\mathbb{R}^d\times\mathbb{S}$, where $r(t) = \varphi\circ\Phi^{-1}(t)$ with

$$\Phi(t) = \int_1^t \frac{du}{\varphi(u)}\,.$$

Proof. $*$*Idea*$*$ In this case, the Foster - Lyapunov method consists of finding an appropriate recurrent (petite) set $C\in\mathscr{B}(\mathbb{R}^d)\times\mathscr{P}(\mathbb{S})$ and constructing an appropriate function $\mathscr{V}:\mathbb{R}^d\times\mathbb{S}\to[1,+\infty\rangle$ (the so-called Lyapunov (energy) function), such that the Lyapunov equation

$$\mathscr{L}\mathscr{V}(x,i) \leq -\varphi\circ\mathscr{V}(x,i) + \kappa\,\mathbb{1}_C(x,i)$$

holds for some $\kappa\in\mathbb{R}$. Under the assumptions of the theorem (in particular, open-set irreducibility and aperiodicity of the process), we show that C is of the form $K\times\mathbb{S}$ for some compact set $K\subset\mathbb{R}^d$, and $\mathscr{V}(x,i)$ is given in terms of $\{c_i\}_{i\in\mathbb{S}}$, $\varphi(u)$ and $\mathscr{V}(x)$.

• Let $\beta := -\sum_{i\in\mathbb{S}} c_i\lambda_i > 0$. Clearly,

$$\sum_{i\in\mathbb{S}}(c_i+\beta)\lambda_i = 0\,.$$

From [YZ10, Lemma A.12] it then follows that the system

$$\sum_{j\in\mathbb{S}} \mathrm{q}_{ij}\gamma_j = -c_i-\beta\,,\qquad i\in\mathbb{S}\,,$$

admits a solution $(\gamma_i)_{i\in\mathbb{S}}$. Let $m>\max\{\beta,2\}$ and let $\mathscr{V}:\mathbb{R}^d\times\mathbb{S}\to[1,+\infty\rangle$ be twice continuously differentiable (in the first coordinate) and such that

$$\mathscr{V}(x,i) = \frac{m}{\beta}\big(\mathscr{V}(x)+\gamma_i\,\varphi\circ\mathscr{V}(x)\big)$$

for all $i\in\mathbb{S}$ and all $|x|$ large enough. Observe that the assumptions in (i) ensure existence of such a function. For all $i\in\mathbb{S}$ and all $|x|$ large enough, we now have

$$\begin{aligned}\mathscr{L}\mathscr{V}(x,i) &= \frac{m}{\beta}\mathscr{L}_i\mathscr{V}(x) + \frac{m\gamma_i}{\beta}\mathscr{L}_i\varphi\circ\mathscr{V}(x) + \frac{m}{\beta}\varphi\circ\mathscr{V}(x)\,\mathscr{Q}(x)\gamma_i\\ &\leq \frac{mc_i}{\beta}\varphi\circ\mathscr{V}(x) + \frac{m\gamma_i}{\beta}\mathscr{L}_i\varphi\circ\mathscr{V}(x) + \frac{m}{\beta}\varphi\circ\mathscr{V}(x)\sum_{j\in\mathbb{S}}\mathrm{q}_{ij}\gamma_j + \mathrm{o}(1)\end{aligned}$$

$$
\begin{aligned}
&= \left(c_i + \gamma_i \frac{\mathscr{L}_i \varphi \circ \mathscr{V}(x)}{\varphi \circ \mathscr{V}(x)} - c_i - \beta\right) \frac{m}{\beta} \varphi \circ \mathscr{V}(x) + \mathrm{o}(1) \\
&= \left(\gamma_i \frac{\mathscr{L}_i \varphi \circ \mathscr{V}(x)}{\varphi \circ \mathscr{V}(x)} - \beta\right) \frac{m}{\beta} \varphi \circ \mathscr{V}(x) + \mathrm{o}(1)\,,
\end{aligned}
$$

where in the second line we use assumption (ii). By assumption, for all $|x|$ large enough it holds that

$$
\sup_{i \in \mathbb{S}} \left| \gamma_i \frac{\mathscr{L}_i \varphi \circ \mathscr{V}(x)}{\varphi \circ \mathscr{V}(x)} \right| < \frac{\beta}{m}\,.
$$

Thus, for all $i \in \mathbb{S}$ and all $|x|$ large enough,

$$
\mathscr{L}\mathscr{V}(x,i) \leq -(m-1)\,\varphi \circ \mathscr{V}(x) \leq -(m-1)\,\varphi\big(\mathscr{V}(x,i)/(m-1)\big) \leq -\varphi \circ \mathscr{V}(x,i)\,,
$$

where in the last step we employed the subadditivity property of $\varphi(u)$. The assertion of the theorem now follows from [Twe94, Theorems 5.1 and 7.1] (which ensure that every compact set is petite set for $\{(X(x,i;t),\Lambda(x,i;t)\}_{t\geq 0})$ and [DFG09, Theorems 3.2 and 3.4]. ■

In the next proposition we show that if $\varphi(u)$ is linear, then $\{(X(x,i;t),\Lambda(x,i;u)\}_{t\geq 0}$ is geometrically ergodic.

Proposition 3.3.11. Let $\{(X(x,i;t),\Lambda(x,i;t)\}_{t\geq 0}$ be an open-set irreducible and aperiodic regime-switching diffusion process satisfying **(A1)-(A5)**. Assume

(i) there are $\{c_i\}_{i\in\mathbb{S}} \subset \mathbb{R}$ and twice continuously differentiable $\mathscr{V} : \mathbb{R}^d \to \langle 1, +\infty\rangle$, such that

$$
\limsup_{|x|\to+\infty} \frac{\mathscr{L}_i \mathscr{V}(x)}{\mathscr{V}(x)} < c_i\,,
$$

(ii) $\mathscr{Q}(x) = \mathscr{Q} + \mathrm{o}(1)$ as $|x| \to +\infty$,

and either one of the following two conditions

(iii) $\mathscr{Q} = (\mathrm{q}_{ij})_{i,j\in\mathbb{S}}$ is the infinitesimal generator of an irreducible right-continuous time-homogeneous Markov chain on $\mathbb{S}$ with invariant probability measure by $\lambda = (\lambda_i)_{i\in\mathbb{S}}$ and $\sum_{i\in\mathbb{S}} c_i\lambda_i < 0$.

(iii') the matrix $-(\mathscr{Q} + \operatorname{diag} c)$ is a nonsingular $\mathscr{M}$-matrix, where $c = (c_i)_{i\in\mathbb{S}}$.

Then, $\{(X(x,i;t),\Lambda(x,i;t)\}_{t\geq 0}$ is geometrically ergodic.

Proof. Assume first (i)-(iii). Analogously as in [Sha15, Theorem 2.1] we conclude that there are $\zeta\in(0,1)$, $\eta>0$ and $\gamma=(\gamma_i)_{i\in\mathbb{S}}$ with strictly positive components, such that

$$(\mathscr{Q}+\zeta\,\mathrm{diag}\,c)\gamma=-\eta\,\gamma.$$

Define $\mathscr{V}(x,i):=\gamma_i\mathscr{V}^{\zeta}(x)$. Since $\zeta\in(0,1)$, it is straightforward to check that

$$\mathscr{L}_i\mathscr{V}^{\zeta}(x)\leq\zeta\,\mathscr{V}^{\zeta-1}(x)\,\mathscr{L}_i\mathscr{V}(x)$$

for all $x\in\mathbb{R}^d$ and $i\in\mathbb{S}$. Thus, for all $i\in\mathbb{S}$ and $|x|$ large enough, we have

$$\begin{aligned}\mathscr{L}\mathscr{V}(x,i)&=\gamma_i\mathscr{L}_i\mathscr{V}^{\zeta}(x)+\mathscr{V}^{\zeta}(x)\,\mathscr{Q}(x)\gamma_i\\&\leq\zeta\,c_i\gamma_i\mathscr{V}^{\zeta}(x)+\mathscr{V}^{\zeta}(x)\sum_{j\in\mathbb{S}}\mathrm{q}_{ij}\gamma_j+\mathrm{o}(1)\\&=-\eta\,\gamma_i\mathscr{V}^{\zeta}(x)+\mathrm{o}(1)\\&\leq-\eta\,\mathscr{V}(x,i)+\mathrm{o}(1),\end{aligned}\tag{3.3.8}$$

which is exactly the Lyapunov equation on [MT93b, page 529] with $c=\eta$, $f(x,i)=\mathscr{V}(x,i)$, $C=\overline{B}_r(0)\times\mathbb{S}$ for $r>0$ large enough and $b=\sup_{(x,i)\in C}|\mathscr{L}\mathscr{V}(x,i)|$. According to [Twe94, Theorems 5.1 and 7.1], together with open-set irreducibility and $\mathscr{C}_b$-Feller property of $\{(X(x,i;t),\Lambda(x,i;t)\}_{t\geq 0}$, it follows that C is a petite set for $\{(X(x,i;t),\Lambda(x,i;t)\}_{t\geq 0}$. Consequently, from [MT93a, Proposition 6.1], [MT93b, Theorem 4.2] and aperiodicity it follows now that there are a petite set $\mathscr{C}\in\mathfrak{B}(\mathbb{R}^d)\times\mathscr{P}(\mathbb{S})$, $T>0$ and a non-trivial measure $\nu_{\mathscr{C}}$ on $\mathfrak{B}(\mathbb{R}^d)\times\mathscr{P}(\mathbb{S})$, such that $\nu_{\mathscr{C}}(\mathscr{C})>0$ and

$$p(t,(x,i),B)\geq\nu_{\mathscr{C}}(B)$$

for all $(x,i)\in\mathscr{C}$, $t\geq T$ and $B\in\mathfrak{B}(\mathbb{R}^d)\times\mathscr{P}(\mathbb{S})$. In particular,

$$p(t,(x,i),\mathscr{C})>0$$

for all $(x,i)\in\mathscr{C}$ and $t\geq T$, which is exactly the definition of aperiodicity used on [DMT95, p. 1675]. Finally, observe that eq. (3.3.8) is also the Lyapunov equation used on [DMT95, p. 1679] with $c=\eta$, $\tilde{\mathscr{V}}(x,i)=\mathscr{V}(x,i)$, $C=\overline{B}_r(0)\times\mathbb{S}$ for $r>0$ large enough and $b=\sup_{(x,i)\in C}|\mathscr{L}\mathscr{V}(x,i)|$. The assertion now follows from [DMT95, Theorem 5.2].

Assume now (i), (ii) and (iii'). Since $-(\mathscr{Q}+\operatorname{diag} c)$ is a nonsingular $\mathscr{M}$-matrix, there is a vector $\gamma=(\gamma_i)_{i\in\mathbb{S}}$ with strictly positive components such that the vector

$$\delta=(\delta_i)_{i\in\mathbb{S}}=-(\mathscr{Q}+\operatorname{diag} c)\gamma$$

also hast strictly positive components. Define $\mathscr{V}(x,i):=\gamma_i\mathscr{V}(x)$ and $\beta:=\inf_{i\in\mathbb{S}}\delta_i>0$. Analogously as above we see that for all $i\in\mathbb{S}$ and all $|x|$ large enough,

$$\mathscr{L}\mathscr{V}(x,i)\ \leq\ -\beta\,\mathscr{V}(x,i)+\mathrm{o}(1)\,,$$

which concludes the proof. ■

As a final step in this sequel, we discuss an example satisfying conditions from Theorem 3.3.10.

Example 3.3.12. Let $\mathbb{S}=\{0,1\}$, and let $\mathrm{q}_{01}=\mathrm{q}_{10}=1$. Hence, $\lambda=(1/2,1/2)$. Further, let

$$b(x,i)\ =\ \begin{cases} b\,, & i=0\,,\\ -\operatorname{sgn}(x)\beta(x)\,, & i=1\,,\end{cases}$$

with $b\in\mathbb{R}$ and $\beta:\mathbb{R}\to[0,+\infty\rangle$ satisfying

$$\lim_{|x|\to+\infty}|x|^{1-2p}\beta(x)\ =\ +\infty \tag{3.3.9}$$

for some $p\in[1/2,1)$, and let $\sigma(x,i)\equiv\sigma(i)>0$ (implying open-set irreducibility and aperiodicity of the process). Define $\mathscr{V}:\mathbb{R}\to\langle 1,+\infty\rangle$ by $\mathscr{V}(x):=1+x^2$, and $\varphi:\langle 1,+\infty\rangle\to\langle 0,+\infty\rangle$ by $\varphi(u):=u^p$. Clearly,

$$\lim_{u\to+\infty}\varphi'(u)\ =\ 0\qquad\text{and}\qquad\lim_{|x|\to+\infty}\frac{\varphi\circ\mathscr{V}(x)}{\mathscr{V}(x)}\ =\ 0\,.$$

Further,

$$\mathscr{L}_i\mathscr{V}(x)\ =\ \begin{cases} 2bx+\sigma(0)^2\,, & i=0\,,\\ -2\operatorname{sgn}(x)x\beta(x)+\sigma(1)^2\,, & i=1\,.\end{cases}$$

From this and eq. (3.3.9) it follows that

$$\lim_{|x|\to+\infty}\frac{\mathscr{L}_i\mathscr{V}(x)}{\varphi\circ\mathscr{V}(x)}\ =\ \lim_{|x|\to+\infty}\frac{\mathscr{L}_i\varphi\circ\mathscr{V}(x)}{\varphi\circ\mathscr{V}(x)}\ =\ \begin{cases} 0\,, & i=0\,,\\ -\infty\,, & i=1\,.\end{cases}$$

Hence, the process satisfies the conditions from Theorem 3.3.10 with arbitrary $c_0>0$ and $c_1=-2c_0$, which implies subgeometric ergodicity with rate $r(t)=t^{p/(1-p)}$.

3.4. ERGODICITY WITH RESPECT TO THE WASSERSTAIN DISTANCE

Again, for processes that exhibit some level of singularity, namely, are not open-set irreducible and aperiodic, we consider ergodicity with respect to the Wasserstein distance. In addition to **(A1)-(A5)**, throughout the section we assume

- $\{\Lambda(x,i;t)\}_{t\geq 0}$ and $\sigma(x,i)$ are state independent, that is, $\Lambda(x,i;t) = \Lambda(i;t)$ and $\sigma(x,i) = \sigma(i)$ for all $(x,i) \in \mathbb{R}^d \times \mathbb{S}$ and $t \geq 0$;
- $\{\Lambda(i;t)\}_{t\geq 0}$ is irreducible.

The definition of the Wasserstein distance needs to be adapted as well.

Let ρ be a distance on $\mathbb{R}^d \times \mathbb{S}$. Denote by $\mathfrak{B}_\rho(\mathbb{R}^d \times \mathbb{S})$ the Borel σ-algebra on $\mathbb{R}^d \times \mathbb{S}$ induced by ρ. For $p \geq 0$ let $\mathscr{P}_{\rho,p}$ be the space of all probability measures μ on $\mathfrak{B}_\rho(\mathbb{R}^d \times \mathbb{S})$ having finite p-th moment, that is, $\int_{\mathbb{R}^d\times\mathbb{S}} \rho((x,i),(y,j))^p \mu(dy \times \{j\}) < \infty$ for some (and then any) $(x,i) \in \mathbb{R}^d \times \mathbb{S}$. For $p \geq 1$ and $\mu,\nu \in \mathscr{P}_{\rho,p}$, the $\mathscr{L}^p$-Wasserstein distance between μ and ν is defined as

$$\mathscr{W}_{\rho,p}(\mu,\nu) := \inf_{\Pi\in\mathscr{C}(\mu,\nu)} \left(\int_{(\mathbb{R}^d\times\mathbb{S})\times(\mathbb{R}^d\times\mathbb{S})} \rho\big((x,i),(y,j)\big)^p \, \Pi\big(dx \times \{i\}, dy \times \{j\}\big) \right)^{1/p},$$

where $\mathscr{C}(\mu,\nu)$ is the family of couplings of μ and ν, that is, $\Pi \in \mathscr{C}(\mu,\nu)$ if, and only if, Π is a probability measure on $(\mathbb{R}^d \times \mathbb{S}) \times (\mathbb{R}^d \times \mathbb{S})$ having μ and ν as its marginals. The restriction of $\mathscr{W}_{\rho,p}$ to $\mathscr{P}_{\rho,p}$ defines a finite distance.

Of our special interest will be the situation when ρ takes the form

$$\rho\big((x,i),(y,j)\big) = \mathbb{1}_{\{i\neq j\}}(i,j) + f\big(|x-y|\big) \tag{3.4.1}$$

for some non-decreasing concave $f : [0,+\infty\rangle \to [0,+\infty\rangle$ satisfying $f(u) = 0$ if, and only if, $u = 0$.

Our methods used naturally generalise those used for a classical diffusion. Therefore, we first adapt the key ingredient - Lemma which is a version of the well-known Grönwall inequality.

Lemma 3.4.1. Let $i\in\mathbb{S}$, $\tau\geq 0$, $\{\Gamma_j\}_{j\in\mathbb{S}}\subset\langle-\infty,0]$, $F:[0,T\rangle\to[0,+\infty\rangle$ with $0<T\leq\infty$, and $\psi:[0,+\infty\rangle\to(0,\infty)$ be such that

(i) $F(t)$ is absolutely continuous on $[t_0,t_1]$ for any $0<t_0<t_1<T$;

(ii) $F'(t)\leq\Gamma_{\Lambda(i;\tau+t)}\psi(F(t))$ a.e. on $[0,T\rangle$;

(iii) $\Psi_{F(0)}(t):=\int_t^{F(0)}ds/\psi(s)<\infty$ for all $t\in\langle 0,F(0)]$.

Then

$$F(t)\,\leq\,\Psi_{F(0)}^{-1}\left(-\int_0^t\Gamma_{\Lambda(i;\tau+s)}ds\right)$$

for all $t\in[0,T\rangle$ such that $-\int_0^t\Gamma_{\Lambda(i;\tau+s)}ds<\Psi_{F(0)}(0)$. In addition, if there is $\gamma\in[F(0),+\infty]$ such that $\Psi_\gamma(t)=\int_t^\gamma ds/\psi(s)<\infty$ for all $t\in(0,\gamma]$, then

$$F(t)\,\leq\,\Psi_\gamma^{-1}\left(-\int_0^t\Gamma_{\Lambda(i;\tau+s)}ds\right)$$

for all $t\in[0,T\rangle$ such that $0\leq-\int_0^t\Gamma_{\Lambda(i;\tau+s)}ds<\Psi_\gamma(0)$. Furthermore, if $\psi(t)$ is convex and vanishes at zero, then $\Psi_{F(0)}(0)=+\infty$. In particular, the previous relations hold for all $t\in[0,T\rangle$.

Proof. We have,

$$-\Psi_{F(0)}(F(t))\,=\,\int_{F(0)}^{f(t)}\frac{ds}{\psi(s)}\,=\,\int_0^t\frac{F'(s)ds}{\psi(F(s))}\,\leq\,\int_0^t\Gamma_{\Lambda(i;\tau+s)}ds$$

for all $t\in[0,T\rangle$, which proves the first assertion. The second claim follows from the fact that $\Psi_{F(0)}(t)\leq\Psi_\gamma(t)$ for all $t\in(0,F(0)]$, while the last part follows from the convexity of $\psi(t)$:

$$\psi(t)\,=\,\psi(t+(1-t)0)\,\leq t\,\psi(1)+(1-t)\,\psi(0)\,=\,t\,\psi(1)$$

for all $t\in[0,1]$. ∎

Theorem 3.4.2. Assume **(A1)-(A5)**, and suppose that $\{\Lambda(x,i;t)\}_{t\geq 0}$ and $\sigma(x,i)$ are x-independent. Assume also that $\{\Lambda(i;t)\}_{t\geq 0}$ is irreducible and let $\lambda=(\lambda_i)_{i\in\mathbb{S}}$ be its invariant probability measure. Further, let $f,\psi:[0,+\infty\rangle\to[0,+\infty\rangle$ be such that

(i) $f(u)$ is bounded, concave, non-decreasing, absolutely continuous on $[u_0,u_1]$, for all $0<u_0<u_1<+\infty$, and $f(u)=0$ if, and only if, $u=0$

(ii) $\psi(u)$ is convex and $\psi(u) = 0$ if, and only if, $u = 0$

(iii) there are $\{\Gamma_i\}_{i\in\mathbb{S}} \subset \langle -\infty, 0]$ such that

$$f'\big(|x-y|\big)\big\langle x-y, b(x,i)-b(y,i)\big\rangle \leq \Gamma_i |x-y|\, \psi\big(f(|x-y|)\big) \tag{3.4.2}$$

a.e. on $\mathbb{R}^d$

(iv) $\sum_{i\in\mathbb{S}} \Gamma_i \lambda_i < 0.$

Then, for ρ given by eq. (3.4.1), and all $p \geq 1$ and $(x,i),(y,j) \in \mathbb{R}^d \times \mathbb{S}$ it holds that

$$\lim_{t\to\infty} \mathscr{W}_{f,p}\big(\delta_{(x,i)}\mathscr{P}_t, \delta_{(y,j)}\mathscr{P}_t\big) = 0\,. \tag{3.4.3}$$

Additionally, if $\psi(u) = u^q$ for some $q > 1$, then

$$\lim_{t\to\infty} t^{1/(q-1)} \mathscr{W}_{f,p}\big(\delta_{(x,i)}\mathscr{P}_t, \delta_{(y,j)}\mathscr{P}_t\big) \leq \left(\frac{1-q}{2}\sum_{i\in\mathbb{S}} \Gamma_i\lambda_i\right)^{1/(1-q)}.$$

If $\psi(u) = \kappa u$ for some $\kappa > 0$, then

$$\lim_{t\to\infty} \mathrm{e}^{\alpha t/2} \mathscr{W}_{f,p}\big(\delta_{(x,i)}\mathscr{P}_t, \delta_{(y,j)}\mathscr{P}_t\big) = 0$$

for all $0 < \alpha < \min\{\vartheta/p, -\kappa \sum_{i\in\mathbb{S}} \Gamma_i\lambda_i\}$, where ϑ is given in eq. (3.2.1).

Proof. Let $\{\bar\Lambda(i;t)\}_{t\geq 0}$ be an independent copy of $\{\Lambda(i;t)\}_{t\geq 0}$ (which is also independent of $\{B(t)\}_{t\geq 0}$). Define

$$\tilde\Lambda(j;t) := \begin{cases} \bar\Lambda(j;t)\,, & t < \tau_{ij}\,, \\ \Lambda(i;t)\,, & t \geq \tau_{ij}\,, \end{cases}$$

for $t \geq 0$. By employing Markov property, it is easy to see that $\{\tilde\Lambda(j;t)\}_{t\geq 0}$ is a Markov chain with the same law as $\{\Lambda(j;t)\}_{t\geq 0}$.

Fix now $(x,i),(y,j) \in \mathbb{R}^d \times \mathbb{S}$, and let $\{(X(x,i;t),\Lambda(i;t))\}_{t\geq 0}$ and $\{(X(y,j;t),\tilde\Lambda(j;t))\}_{t\geq 0}$ be the corresponding solutions to eq. (3.0.1). Define

$$\tau := \inf\{t > 0\colon (X(x,i;t),\Lambda(i;t)) = (X(y,j;t),\tilde\Lambda(j;t))\}\,.$$

Clearly, $\tau \geq \tau_{ij}$. Put

$$Y(y,j;t) := \begin{cases} X(y,j;t)\,, & t < \tau\,, \\ X(x,i;t)\,, & t \geq \tau\,, \end{cases}$$

for $t \geq 0$. The marginals of the process $\{(Y(y,j;t),\tilde{\Lambda}(j;t))\}_{t\geq 0}$ have the same law as the marginals of $\{(X(y,j;t),\tilde{\Lambda}(j;t))\}_{t\geq 0}$. Namely, we have that

$$\begin{aligned}
&\mathbb{P}\big((Y(y,j;t),\tilde{\Lambda}(j;t)) \in B\times\{k\}\big)\\
&= \mathbb{P}\big((Y(y,j;t),\tilde{\Lambda}(j;t)) \in B\times\{k\},\, t<\tau\big) + \mathbb{P}\big((Y(y,j;t),\tilde{\Lambda}(j;t)) \in B\times\{k\},\, t\geq\tau\big)\\
&= \mathbb{P}\big((X(y,j;t),\tilde{\Lambda}(j;t)) \in B\times\{k\}, t<\tau\big) + \mathbb{P}\big((X(x,i;t),\Lambda(i;t)) \in B\times\{k\},\, t\geq\tau\big).
\end{aligned}$$

Further, by the strong Markov property we have that

$$\begin{aligned}
&\mathbb{P}\big((X(x,i;t),\Lambda(i;t)) \in B\times\{k\},\, t\geq\tau\big)\\
&= \mathbb{E}\left[\mathbb{E}\left[\mathbb{1}_{\{t-\tau+\tau\geq\tau\}}\,\mathbb{1}_{B\times\{k\}}((X(x,i;t-\tau+\tau),\Lambda(i;t-\tau+\tau))\mid\mathscr{F}_\tau\right]\right]\\
&= \mathbb{E}\left[\mathbb{1}_{\{t\geq\tau\}}\mathbb{P}\big((X(X(x,i;\tau),\Lambda(i;\tau);t-\tau),\Lambda(\Lambda(i;\tau);t-\tau)) \in B\times\{k\},\, t-\tau\geq 0\big)\right]\\
&= \mathbb{E}\left[\mathbb{1}_{\{t\geq\tau\}}\mathbb{P}\big((X(X(y,i;\tau),\tilde{\Lambda}(i;\tau);t-\tau),\tilde{\Lambda}(\tilde{\Lambda}(i;\tau);t-\tau)) \in B\times\{k\},\, t-\tau\geq 0\big)\right]\\
&= \mathbb{P}\big((X(y,i;t),\tilde{\Lambda}(i;t)) \in B\times\{k\},\, t\geq\tau\big),
\end{aligned}$$

which proves the assertion. Consequently,

$$\mathscr{W}^p_{f,p}\big(\delta_{(x,i)}\mathscr{P}_t,\delta_{(y,j)}\mathscr{P}_t\big) \leq \mathbb{E}\left[\rho\big((X(x,i;t),\Lambda(i;t)),(Y(y,j;t),\tilde{\Lambda}(j;t))\big)^p\right].$$

Now, we have

$$\begin{aligned}
\mathsf{W}^p_{f,p}\big(\delta_{(x,i)}\mathscr{P}_t,\delta_{(y,j)}\mathscr{P}_t\big) &\leq \mathbb{E}\left[\rho\big((X(x,i;t),\Lambda(i;t)),(Y(y,j;t)\tilde{\Lambda}(j;t))\big)^p\mathbb{1}_{\{\tau_{ij}>t/2\}}\right]\\
&\quad+\mathbb{E}\left[\rho\big((X(x,i;t),\Lambda(i;t)),(Y(y,j;t)\tilde{\Lambda}(j;t))\big)^p\mathbb{1}_{\{\tau_{ij}\leq t/2\}}\right]\\
&\leq (1+\gamma)^p\mathrm{e}^{-\vartheta\lfloor t/2\rfloor}+\mathbb{E}\left[f\big(|X(x,i;t)-Y(y,j;t)|\big)^p\,\mathbb{1}_{\{\tau_{ij}\leq t/2\}}\right],
\end{aligned}$$

where in the last step $\gamma := \sup_{t>0} f(t)$ and we employed Lemma 3.2.2. After time τ_{ij} the processes $\{\Lambda(i;t)\}_{t\geq 0}$ and $\{\tilde{\Lambda}(j;t)\}_{t\geq 0}$ move together. Hence, by eq. (3.4.2) it holds that (here we also use the fact that $\sigma(x,i)=\sigma(i)$)

$$\frac{d}{dt}f\big(|X(x,i;t)-Y(y,j;t)|\big) \leq 0$$

a.e. on $\left[\tau_{ij},+\infty\right\rangle$. Thus, $t\mapsto f(|X(x,i;t)-Y(y,j;t)|)$ is non-increasing on $\left[\tau_{ij},+\infty\right\rangle$ and

$$\begin{aligned}
&\mathscr{W}^p_{f,p}\big(\delta_{(x,i)}\mathscr{P}_t,\delta_{(y,j)}\mathscr{P}_t\big)\\
&\leq (1+\gamma)^p\mathrm{e}^{-\vartheta\lfloor t/2\rfloor}+\mathbb{E}\left[f\big(|X(x,i;\tau_{ij}+t/2)-Y(y,j;\tau_{ij}+t/2)|\big)^p\,\mathbb{1}_{\{\tau_{ij}\leq t/2\}}\right].
\end{aligned}$$

Define

$$F(t) := f\big(|X(x,i;\tau_{ij}+t) - Y(y,j;\tau_{ij}+t)|\big)$$

for $t \geq 0$. By employing eq. (3.4.2) again, it follows that

$$\frac{d}{dt}F(t) \leq \Gamma_{\Lambda(i;\tau_{ij}+t)}\psi\big(F(t)\big)$$

a.e. on $\big[0, \tau - \tau_{ij}\big\rangle$. By Lemma 3.4.1 we have that

$$F(t) \leq \Psi_\gamma^{-1}\left(-\int_0^t \Gamma_{\Lambda(i;\tau_{ij}+s)}ds\right) = \Psi_\gamma^{-1}\left(-\int_{\tau_{ij}}^{t+\tau_{ij}} \Gamma_{\Lambda(i;s)}ds\right)$$

on $[0, +\infty\rangle$. For $t \geq \tau$ the term on the left-hand side vanishes, and the term on the right-hand side is well defined and strictly positive ($\psi(u)$ is convex and $\psi(u) = 0$ if, and only if, $u = 0$). Thus,

$$\mathscr{W}_{f,p}^p\big(\delta_{(x,i)}\mathscr{P}_t, \delta_{(y,j)}\mathscr{P}_t\big) \leq (1+\gamma)^p \mathrm{e}^{-\vartheta\lfloor t/2\rfloor} + \mathbb{E}\left[\left(\Psi_\gamma^{-1}\left(-\int_{\tau_{ij}}^{t/2+\tau_{ij}} \Gamma_{\Lambda(i;s)}ds\right)\right)^p\right].$$

Birkhoff ergodic theorem implies that

$$\lim_{t\to\infty} \frac{2}{t}\int_{\tau_{ij}}^{t/2+\tau_{ij}} \Gamma_{\Lambda(i;s)}ds = \sum_{i\in\mathbb{S}} \Gamma_i\lambda_i < 0, \qquad \mathbb{P}\text{-a.s.}$$

Hence, since $\psi(u)$ is convex and $\psi(u) = 0$ if, and only if, $u = 0$,

$$\lim_{t\to+\infty} \Psi_\gamma^{-1}\left(-\int_{\tau_{ij}}^{t/2+\tau_{ij}} \Gamma_{\Lambda(i;s)}ds\right) = 0$$

$\mathbb{P}$-a.s. This, together with dominated convergence theorem, shows the first assertion.

Assume now that $\psi(u) = u^q$ for $q > 1$. Then,

$$\Psi_\gamma(t) = \frac{1}{q-1}\big(t^{1-q} - \gamma^{1-q}\big) \qquad \text{and} \qquad \Psi_\gamma^{-1}(t) = \big(\gamma^{1-q} + (q-1)t\big)^{1/(1-q)}.$$

By employing Birkhoff ergodic theorem and Fatou's lemma, we have

$$\begin{aligned}
&\liminf_{t\to+\infty} \frac{\mathbb{E}\left[\left(\Psi_\gamma^{-1}\left(-\int_{\tau_{ij}}^{t/2+\tau_{ij}}\Gamma_{\Lambda(i;s)}ds\right)\right)^p\right]}{\left(\frac{(1-q)t}{2}\sum_{i\in\mathbb{S}}\Gamma_i\lambda_i\right)^{p/(1-q)}} \\
&= \liminf_{t\to+\infty} \mathbb{E}\left[\frac{\left(\gamma^{1-q} + (1-q)\int_{\tau_{ij}}^{t/2+\tau_{ij}}\Gamma_{\Lambda(i;s)}ds\right)^{p/(1-q)}}{\left(\frac{(1-q)t}{2}\sum_{i\in\mathbb{S}}\Gamma_i\lambda_i\right)^{p/(1-q)}}\right] \\
&\geq 1.
\end{aligned}$$

On the other side, Birkhoff ergodic theorem, Fatou's lemma and Jensen's inequality imply

$$\begin{aligned}
&\liminf_{t\to+\infty} \frac{\left(\frac{(1-q)t}{2}\sum_{i\in\mathbb{S}}\Gamma_i\lambda_i\right)^{p/(1-q)}}{\mathbb{E}\left[\left(\Psi_\gamma^{-1}\left(-\int_{\tau_{ij}}^{t/2+\tau_{ij}}\Gamma_{\Lambda(i;s)}ds\right)\right)^p\right]} \\
&= \liminf_{t\to+\infty}\mathbb{E}\left[\frac{\left(\gamma^{1-q}+(1-q)\int_{\tau_{ij}}^{t/2+\tau_{ij}}\Gamma_{\Lambda(i;s)}ds\right)^{p/(1-q)}}{\left(\frac{(1-q)t}{2}\sum_{i\in\mathbb{S}}\Gamma_i\lambda_i\right)^{p/(1-q)}}\right]^{-1} \\
&\geq \liminf_{t\to\infty}\mathbb{E}\left[\frac{\left(\gamma^{q-1}+(1-q)\int_{\tau_{ij}}^{t/2+\tau_{ij}}\Gamma_{\Lambda(i;s)}ds\right)^{p/(1-q)}}{\left(\frac{(1-q)t}{2}\sum_{i\in\mathbb{S}}\Gamma_i\lambda_i\right)^{p/(q-1)}}\right] \\
&= 1\,.
\end{aligned}$$

Thus,

$$\lim_{t\to+\infty}\frac{\mathbb{E}\left[\left(\Psi_\gamma^{-1}\left(-\int_{\tau_{ij}}^{t/2+\tau_{ij}}\Gamma_{\Lambda(i;s)}ds\right)\right)^p\right]}{\left(\frac{(1-q)t}{2}\sum_{i\in\mathbb{S}}\Gamma_i\lambda_i\right)^{p/(1-q)}} = 1\,,$$

which proves the second relation.

If $\psi(u)=\kappa u$ for $\kappa>0$, then

$$\Psi_\gamma(t) = \frac{1}{\kappa}\ln\left(\frac{\gamma}{t}\right) \qquad \text{and} \qquad \Psi_\gamma^{-1}(t) = \gamma e^{-\kappa t}\,.$$

Arguing as above, it holds that for any $0<\beta<-\kappa\sum_{i\in\mathbb{S}}\Gamma_i\lambda_i$,

$$\lim_{t\to+\infty}\frac{\mathbb{E}\left[\left(\Psi_\gamma^{-1}\left(-\int_{\tau_{ij}}^{t/2+\tau_{ij}}\Gamma_{\Lambda(i;s)}ds\right)\right)^p\right]}{e^{-p\beta t/2}} = 0\,.$$

Hence, by taking $0<\alpha<\min\{\vartheta/p,-\kappa\sum_{i\in\mathbb{S}}\Gamma_i\lambda_i\}$ the assertion follows. ■

Crucial assumption in Theorem 3.4.2 is that the function $f(u)$, that is, distance ρ, is bounded. In the following theorem we discuss the situation when this is not necessarily the case.

Theorem 3.4.3. Assume **(A1)-(A5)**, and suppose $\{\Lambda(x,i;t)\}_{t\geq 0}$ and $\sigma(x,i)$ are x-independent. Assume also that $b(x,i)$ is locally Lipschitz continuous for every $i\in\mathbb{S}$, and that $\{\Lambda(i;t)\}_{t\geq 0}$ is irreducible and let $\lambda=(\lambda_i)_{i\in\mathbb{S}}$ be its invariant probability measure. Further, assume that there is $0<K<\vartheta$ (recall that ϑ is given in eq. (3.2.1)) such that

$$2\langle x, \mathrm{b}(x,i)\rangle + \mathrm{Tr}\big(\sigma(i)\sigma(i)^T\big) \leq K(1+|x|^2) \tag{3.4.4}$$

for all $x\in\mathbb{R}^d$ and $i\in\mathbb{S}$. Let $f,\psi:[0,+\infty\rangle\to[0,+\infty\rangle$ be such that

(i) $f(u)$ is concave, non-decreasing, absolutely continuous on $[u_0,u_1]$, for all $0<u_0<u_1<\infty$, and $f(u)=0$ if, and only if, $u=0$

(ii) $\psi(u)$ is convex and $\psi(u)=0$ if, and only if, $u=0$

(iii) there are $\{\Gamma_i\}_{i\in\mathbb{S}}\subset\langle-\infty,0]$ and $\eta>\inf\{f(u)\mid u>0\}$, such that

$$f'(|x-y|)\big\langle x-y,b(x,i)-b(y,i)\big\rangle \le \begin{cases}\Gamma_i|x-y|\psi\big(f(|x-y|)\big)\,, & f(|x-y|)\le\eta\,,\\ 0\,, & \text{otherwise}\end{cases} \tag{3.4.5}$$

a.e. on $\mathbb{R}^d$

(iv) $\sum_{i\in\mathbb{S}}\Gamma_i\lambda_i<0$.

Then, for ρ given by eq. (3.4.1), and all $(x,i),(y,j)\in\mathbb{R}^d\times\mathbb{S}$ it holds that

$$\lim_{t\to\infty}\mathscr{W}_{f,1}\big(\delta_{(x,i)}\mathscr{P}_t,\delta_{(y,j)}\mathscr{P}_t\big) \;=\; 0\,. \tag{3.4.6}$$

Additionally, if $\psi(u)=u^q$ for some $q>1$, then $\mathbb{E}\left[|X(x,i;\tau_{ij})-Y(y,j;\tau_{ij})|^2\right]<\infty$ and

$$\begin{aligned}&\lim_{t\to\infty}t^{1/(q-1)}\mathscr{W}_{f,1}\big(\delta_{(x,i)}\mathscr{P}_t,\delta_{(y,j)}\mathscr{P}_t\big)\\ &\le \mathbb{E}\left[\lceil\delta|X(x,i;\tau_{ij})-Y(y,j;\tau_{ij})|\rceil^2\right]^{1/2}\left(\frac{1-q}{2}\sum_{i\in\mathbb{S}}\Gamma_i\lambda_i\right)^{1/(1-q)},\end{aligned}$$

where $\delta:=\inf\{t\ge0\colon f(1/t)\le\eta\}$. If $\psi(u)=\kappa u$ for some $\kappa>0$, then

$$\lim_{t\to\infty}\mathrm{e}^{\alpha t/2}\mathscr{W}_{f,1}\big(\delta_{(x,i)}\mathscr{P}_t,\delta_{(y,j)}\mathscr{P}_t\big) \;=\; 0$$

for all $0<\alpha<\min\{\vartheta,-\kappa\sum_{i\in\mathbb{S}}\Gamma_i\lambda_i\}$.

In order to prove this result, we discuss conditions under which the process $\{X(x,i;t)\}_{t\ge0}$ has second moment.

Lemma 3.4.4. Assume eq. (3.4.4). Then,

$$\mathbb{E}\big[|X(x,i;t)|^2\big] \;\le\; \big(1+|x|^2\big)\mathrm{e}^{Kt}\,.$$

Furthermore, for any $\{\mathscr{F}_t\}_{t\ge0}$-stopping time τ such that $\mathbb{E}\big[\mathrm{e}^{2K\tau}\big]<\infty$ it follows that

$$\mathbb{E}\left[|X(x,i;\tau)|^2\right] \le |x|^2+\big(1+|x|^2\big)\mathbb{E}\left[\mathrm{e}^{K\tau}\right]+4\sup_{j\in\mathbb{S}}\mathrm{Tr}\big(\sigma(j)\sigma(j)^T\big)^{\frac12}\big(1+|x|^2\big)^{\frac12}\mathbb{E}\left[\mathrm{e}^{2K\tau}\right]^{\frac12}.$$

Proof. Recall that $\{X(x,i;t)\}_{t\geq 0}$ satisfies

$$X(x,i;t) = x + \int_0^t b\big(X(x,i;s),\Lambda(i;s)\big)ds + \int_0^t \sigma\big(\Lambda(i;s)\big)dB(s)\,.$$

For $n\in\mathbb{N}$, define

$$\tau_n := \inf\{t\geq 0 : |X(x,i;t)|\geq n\}\,.$$

By employing Itô's formula we conclude that

$$\begin{aligned}
&|X(x,i;t\wedge\tau_n)|^2\\
&= |x|^2 + 2\int_0^{t\wedge\tau_n}\big\langle X(x,i;s), b\big(X(x,i;s),\Lambda(i;s)\big)\big\rangle ds + \int_0^{t\wedge\tau_n}\mathrm{Tr}\big(\sigma\big(\Lambda(i;s)\big)\sigma\big(\Lambda(i;s)\big)^T\big)ds\\
&\quad + 2\int_0^{t\wedge\tau_n} X(x,i;s)^T\sigma\big(\Lambda(x,i;s)\big)dB(s)\\
&\leq |x|^2 + K\int_0^t \big(1+|X(x,i;s)|^2\big)\mathbb{1}_{[0,\tau_n]}(s)\,ds + 2\int_0^{t\wedge\tau_n} X(x,i;s)^T\sigma\big(\Lambda(x,i;s)\big)dB(s)\\
&\leq |x|^2 + K\int_0^t\big(1+|X(x,i;s\wedge\tau_n)|^2\big)ds + 2\int_0^{t\wedge\tau_n} X(x,i;s)^T\sigma\big(\Lambda(x,i;s)\big)dB(s)\,.
\end{aligned}$$

Since the last term on the right side is a martingale, we conclude that

$$1+\mathbb{E}\Big[|X(x,i;t\wedge\tau_n)|^2\Big] \leq 1+|x|^2 + K\int_0^t\Big(1+\mathbb{E}\Big[|X(x,i;s\wedge\tau_n)|^2\Big]\Big)ds\,.$$

The first assertion now follows by employing Grönwall's inequality and Fatou's lemma (observe that since $\{X(x,i;t)\}_{t\geq 0}$ is nonexplosive, $\mathbb{P}(\lim_{n\to\infty}\tau_n=\infty)=1$).

Assume now that τ is a stopping time such that $\mathbb{E}[\mathrm{e}^{2K\tau}]<\infty$. By employing Itô's lemma again we have that

$$|X(x,i;t)|^2 \leq |x|^2 + Kt + K\int_0^t |X(x,i;s)|^2 ds + 2\int_0^t X(x,i;s)^T\sigma\big(\Lambda(i;s)\big)dB(s)\,.$$

Grönwall's inequality then gives

$$\begin{aligned}
|X(x,i;t)|^2 &\leq |x|^2 + Kt + 2\int_0^t X(x,i;s)^T\sigma\big(\Lambda(i;s)\big)dB(s)\\
&\quad + \int_0^t K\left(|x|^2 + Ks + 2\int_0^s X(x,i;u)^T\sigma\big(\Lambda(i;u)\big)dB(u)\right)\mathrm{e}^{K(t-s)}ds\,.
\end{aligned}$$

Consequently,

$$\begin{aligned}
&|X(x,i;t\wedge\tau)|^2\\
&\leq |x|^2 + K(t\wedge\tau) + 2\int_0^{t\wedge\tau} X(x,i;s)^T\sigma\big(\Lambda(i;s)\big)dB(s)
\end{aligned}$$

$$
\begin{aligned}
&+\int_0^{t\wedge\tau} K\left(|x|^2+Ks+2\int_0^s X(x,i;u)^T\sigma\big(\Lambda(i;u)\big)dB(u)\right)\mathrm{e}^{K(t\wedge\tau-s)}ds\\
&=|x|^2+K(t\wedge\tau)+2\int_0^{t\wedge\tau}X(x,i;s)^T\sigma\big(\Lambda(i;s)\big)dB(s)\\
&\quad+|x|^2\big(\mathrm{e}^{Kt\wedge\tau}-1\big)+\mathrm{e}^{Kt\wedge\tau}-K(t\wedge\tau)\\
&\quad+2K\int_0^t\left(\mathbb{1}_{[0,t\wedge\tau]}(s)\,\mathrm{e}^{K(t\wedge\tau-s)}\int_0^s X(x,i;u)^T\sigma\big(\Lambda(i;u)\big)dB(u)\right)ds\\
&\le |x|^2+2\int_0^{t\wedge\tau}X(x,i;s)^T\sigma\big(\Lambda(i;s)\big)dB(s)+\big(1+|x|^2\big)\mathrm{e}^{K\tau}\\
&\quad+2K\int_0^t\left(\mathbb{1}_{[0,t\wedge\tau]}(s)\,\mathrm{e}^{K(t\wedge\tau-s)}\int_0^s X(x,i;u)^T\sigma\big(\Lambda(i;u)\big)dB(u)\right)ds\,.
\end{aligned}
$$

Taking expectation (and the previous result) we have that

$$
\begin{aligned}
&\mathbb{E}\big[|X(x,i;t\wedge\tau)|^2\big]\\
&\le |x|^2+\big(1+|x|^2\big)\mathbb{E}\left[\mathrm{e}^{K\tau}\right]\\
&\quad+2K\int_0^t\mathbb{E}\left[\mathbb{1}_{[0,t\wedge\tau]}(s)\,\mathrm{e}^{K(t\wedge\tau-s)}\int_0^s X(x,i;u)^T\sigma\big(\Lambda(i;u)\big)dB(u)\right]ds\\
&\le |x|^2+\big(1+|x|^2\big)\mathbb{E}\left[\mathrm{e}^{K\tau}\right]\\
&\quad+2K\int_0^t\mathbb{E}\left[\mathbb{1}_{[0,t\wedge\tau]}(s)\,\mathrm{e}^{2K(t\wedge\tau-s)}\right]^{1/2}\mathbb{E}\left[\left(\int_0^s\mathbb{1}_{[0,t\wedge\tau]}X(x,i;u)^T\sigma(\Lambda(i;u))dB(u)\right)^2\right]^{1/2}ds\\
&\le |x|^2+\big(1+|x|^2\big)\mathbb{E}\left[\mathrm{e}^{K\tau}\right]+2K\sup_{j\in\mathbb{S}}\mathrm{Tr}\big(\sigma(j)\sigma(j)^T\big)^{1/2}\mathbb{E}\left[\mathrm{e}^{2K\tau}\right]^{1/2}\\
&\quad\int_0^t\mathrm{e}^{-Ks}\mathbb{E}\left[\int_0^s\mathbb{1}_{[0,t\wedge\tau]}(s)\,|X(x,i;u)|^2du\right]^{1/2}ds\\
&\le |x|^2+\big(1+|x|^2\big)\mathbb{E}\left[\mathrm{e}^{K\tau}\right]+2K\sup_{j\in\mathbb{S}}\mathrm{Tr}\big(\sigma(j)\sigma(j)^T\big)^{1/2}\mathbb{E}\left[\mathrm{e}^{2K\tau}\right]^{1/2}\\
&\quad\int_0^\infty\mathrm{e}^{-Ks}\mathbb{E}\left[\int_0^s|X(x,i;u)|^2du\right]^{1/2}ds\\
&\le |x|^2+\big(1+|x|^2\big)\mathbb{E}\left[\mathrm{e}^{K\tau}\right]+2K\sup_{j\in\mathbb{S}}\mathrm{Tr}\big(\sigma(j)\sigma(j)^T\big)^{1/2}\big(1+|x|^2\big)^{1/2}\mathbb{E}\left[\mathrm{e}^{2K\tau}\right]^{1/2}\int_0^\infty\mathrm{e}^{-Ks/2}ds\\
&= |x|^2+\big(1+|x|^2\big)\mathbb{E}\left[\mathrm{e}^{K\tau}\right]+4\sup_{j\in\mathbb{S}}\mathrm{Tr}\big(\sigma(j)\sigma(j)^T\big)^{1/2}\big(1+|x|^2\big)^{1/2}\mathbb{E}\left[\mathrm{e}^{2K\tau}\right]^{1/2},
\end{aligned}
$$

where in the third step we used Itô's isometry and in the fifth step we used the first assertion of the lemma. ■

We are now ready to prove Theorem 3.4.3.

Proof of Theorem 3.4.3. By using the same reasoning (and notation) as in the proof of Theorem 3.4.2, we have that

$$\mathscr{W}_{f,1}\left(\delta_{(x,i)}\mathscr{P}_t,\delta_{(y,j)}\mathscr{P}_t\right) \leq \mathbb{E}\left[\rho\big((X(x,i;t),\Lambda(i;t)),(Y(y,j;t),\tilde{\Lambda}(j;t))\big)\right].$$

Further, for $\varepsilon>0$ such that $K+K\varepsilon<\vartheta$ (such ε exists since by assumption $K<\vartheta$) it follows that

$$\begin{aligned}
&\mathbb{E}\left[\rho\big((X(x,i;t),\Lambda(i;t)),(Y(y,j;t),\tilde{\Lambda}(j;t))\big)\right]\\
&= \mathbb{E}\left[\left(\mathbb{1}_{\{\Lambda(i;t)\neq\tilde{\Lambda}(j;t)\}}+f(|X(x,i;t)-Y(y,j;t)|)\right)\mathbb{1}_{\{\tau_{ij}>t/(1+\varepsilon)\}}\right]\\
&\quad+\mathbb{E}\left[f(|X(x,i;t)-Y(y,j;t)|)\mathbb{1}_{\{\tau_{ij}\leq t/(1+\varepsilon)\}}\right]\\
&\leq \mathbb{E}\left[\left(1+f(|X(x,i;t)-Y(y,j;t)|)\right)^2\right]^{1/2}\mathbb{P}(\tau_{ij}>t/(1+\varepsilon))^{1/2}\\
&\quad+\mathbb{E}\left[f(|X(x,i;t)-Y(y,j;t)|)\mathbb{1}_{\{\tau_{ij}\leq t/(1+\varepsilon)\}}\right]\\
&\leq \mathbb{E}\left[\left(1+f(|X(x,i;t)|)+f(|Y(y,j;t)|)\right)^2\right]^{1/2}\mathrm{e}^{-(\vartheta/2)\lfloor t/(1+\varepsilon)\rfloor}\\
&\quad+\mathbb{E}\left[f(|X(x,i;t)-Y(y,j;t)|)\mathbb{1}_{\{\tau_{ij}\leq t/(1+\varepsilon)\}}\right]\\
&\leq 3^{1/2}\left(1+\mathbb{E}\left[f(|X(x,i;t)|)^2\right]+\mathbb{E}\left[f(|Y(y,j;t)|)^2\right]\right)^{1/2}\mathrm{e}^{-(\vartheta/2)\lfloor t/(1+\varepsilon)\rfloor}\\
&\quad+\mathbb{E}\left[f(|X(x,i;t)-Y(y,j;t)|)\mathbb{1}_{\{\tau_{ij}\leq t/(1+\varepsilon)\}}\right]\\
&\leq C_1\left(1+\mathbb{E}\left[|X(x,i;t)|^2\right]+\mathbb{E}\left[|Y(y,j;t)|^2\right]\right)^{1/2}\mathrm{e}^{-(\vartheta/2)\lfloor t/(1+\varepsilon)\rfloor}\\
&\quad+\mathbb{E}\left[f(|X(x,i;t)-Y(y,j;t)|)\mathbb{1}_{\{\tau_{ij}\leq t/(1+\varepsilon)\}}\right]\\
&\leq C_2\left(1+|x|^2+|y|^2\right)^{1/2}\mathrm{e}^{Kt/2-(\vartheta/2)\lfloor t/(1+\varepsilon)\rfloor}+\mathbb{E}\left[f(|X(x,i;t)-Y(y,j;t)|)\mathbb{1}_{\{\tau_{ij}\leq t/(1+\varepsilon)\}}\right],
\end{aligned}$$

for some $C_1,C_2>0$. Here, in the third step we used subadditivity property of concave functions and Lemma 3.2.2, in the fifth step we used the fact that $f(u)\leq Au+B$ for some $A,B>0$ ($f(u)$ is concave), and in the last step we used Lemma 3.4.4. Clearly, the first term on the right-hand side will converge to zero (as t goes to infinity) due to the choice of $\varepsilon>0$.

We next discuss the second term. Analogously as in the proof of Theorem 3.4.2, by employing eq. (3.4.5), it holds that

$$\frac{d}{dt}f\big(|X(x,i;t)-Y(y,j;t)|\big) \leq 0$$

a.e. on $\left[\tau_{ij},+\infty\right\rangle$. Thus, $t\mapsto f(|X(x,i;t)-Y(y,j;t)|)$ is non-increasing on $[\tau_{ij},\infty)$ and

$$\mathscr{W}_{f,1}\left(\delta_{(x,i)}\mathscr{P}_t,\delta_{(y,j)}\mathscr{P}_t\right) \leq C_2\left(1+|x|^2+|y|^2\right)^{1/2}\mathrm{e}^{Kt/2-(\vartheta/2)\lfloor t/(1+\varepsilon)\rfloor}$$

$$+\mathbb{E}\left[F(\varepsilon t/(1+\varepsilon))\mathbb{1}_{\{\tau_{ij}\le t/(1+\varepsilon)\}}\right],$$

where $F(t)$ is given (as in the proof of Theorem 3.4.2) by

$$F(t) = f\big(|X(x,i;\tau_{ij}+t)-Y(y,j;\tau_{ij}+t)|\big)$$

for $t\ge 0$. We now have that

$$\begin{aligned}&\mathscr{W}_{f,1}\big(\delta_{(x,i)}\mathscr{P}_t,\delta_{(y,j)}\mathscr{P}_t\big)\\&\le C_2\left(1+|x|^2+|y|^2\right)^{1/2}\mathrm{e}^{Kt/2-(\vartheta/2)\lfloor t/(1+\varepsilon)\rfloor}\\&\quad+\mathbb{E}\left[F(\varepsilon t/(1+\varepsilon))\mathbb{1}_{\{\tau_{ij}\le t/(1+\varepsilon)\}}\mathbb{1}_{\{f(|X(x,i;\tau_{ij})-Y(y,j;\tau_{ij})|)\le\eta\}}\right]\\&\quad+\mathbb{E}\left[F(\varepsilon t/(1+\varepsilon))\mathbb{1}_{\{\tau_{ij}\le t/(1+\varepsilon)\}}\mathbb{1}_{\{f(|X(x,i;\tau_{ij})-Y(y,j;\tau_{ij})|)>\eta\}}\right].\end{aligned}\tag{3.4.7}$$

On the event $\{f(|X(x,i;\tau_{ij})-Y(y,j;\tau_{ij})|)\le\eta\}$ we have the following. By employing eq. (3.4.5) again, it follows that

$$\frac{d}{dt}F(t)\le\Gamma_{\Lambda(i;\tau_{ij}+t)}\psi\big(F(t)\big)$$

a.e. on $\left[0,\tau-\tau_{ij}\right\rangle$. Lemma 3.4.1 now implies that

$$F(t)\le\Psi^{-1}_{f(|X(x,i;\tau_{ij})-Y(y,j;\tau_{ij})|)}\left(-\int_0^t\Gamma_{\Lambda(i;\tau_{ij}+s)}ds\right)\le\Psi_\eta^{-1}\left(-\int_{\tau_{ij}}^{t+\tau_{ij}}\Gamma_{\Lambda(i;s)}ds\right)$$

on $[0,+\infty\rangle$. For $t\ge\tau$ the term on the left-hand side vanishes, and the term on the right-hand side is well defined and strictly positive ($\psi(u)$ is convex and $\psi(u)=0$ if, and only if, $u=0$). Thus,

$$\begin{aligned}&\mathbb{E}\left[F(\varepsilon t/(1+\varepsilon))\mathbb{1}_{\{\tau_{ij}\le t/(1+\varepsilon)\}}\mathbb{1}_{\{f(|X(x,i;\tau_{ij})-Y(y,j;\tau_{ij})|)\le\eta\}}\right]\\&\le\mathbb{E}\left[\Psi_\eta^{-1}\left(-\int_{\tau_{ij}}^{\varepsilon t/(1+\varepsilon)+\tau_{ij}}\Gamma_{\Lambda(i;s)}ds\right)\mathbb{1}_{\{f(|X(x,i;\tau_{ij})-Y(y,j;\tau_{ij})|)\le\eta\}}\right].\end{aligned}$$

Birkhoff ergodic theorem implies that

$$\lim_{t\to\infty}\frac{1+\varepsilon}{\varepsilon t}\int_{\tau_{ij}}^{\varepsilon t/(1+\varepsilon)+\tau_{ij}}\Gamma_{\Lambda(i;s)}ds=\sum_{i\in\mathbb{S}}\Gamma_i\lambda_i<0$$

$\mathbb{P}$-a.s. on $\{f(|X(x,i;\tau_{ij})-Y(y,j;\tau_{ij})|)\le\eta\}$. Hence, since $\psi(u)$ is convex and $\psi(u)=0$ if, and only if, $u=0$,

$$\lim_{t\to\infty}\Psi_\eta^{-1}\left(-\int_{\tau_{ij}}^{\varepsilon t/(1+\varepsilon)+\tau_{ij}}\Gamma_{\Lambda(i;s)}ds\right)=0$$

$\mathbb{P}$-a.s. This, together with dominated convergence theorem, shows that

$$\lim_{t\to\infty}\mathbb{E}\Big[F(\varepsilon t/(1+\varepsilon))\mathbb{1}_{\{\tau_{ij}\le t/(1+\varepsilon)\}}\mathbb{1}_{\{f(|X(x,i;\tau_{ij})-Y(y,j;\tau_{ij})|)\le\eta\}}\Big] = 0\,. \tag{3.4.8}$$

On the event $\{f(|X(x,i;\tau_{ij})-Y(y,j;\tau_{ij})|)>\eta\}$ we proceed as follows. Recall that $\delta=\inf\{t\ge 0\colon f(1/t)\le\eta\}$. It clearly must hold that $\delta>0$. Thus, since for $x,y\in\mathbb{R}^d$, $\lceil\delta|x-y|\rceil\ge\delta|x-y|$, we have that

$$f\left(\frac{|x-y|}{\lceil\delta|x-y|\rceil}\right)\le f(1/\delta)\le\eta\,.$$

Let $z_0,\dots,z_{\lceil\delta|X(x,i;\tau_{ij})-Y(y,j;\tau_{ij})|\rceil}\in\mathbb{R}^d$ be such that $z_0:=X(x,i;\tau_{ij})$ and

$$z_{k+1}:=z_k+\frac{X(x,i;\tau_{ij})-Y(y,j;\tau_{ij})}{\lceil\delta|X(x,i;\tau_{ij})-Y(y,j;\tau_{ij})|\rceil}\,,\qquad k=0,\dots,\lceil\delta|X(x,i;\tau_{ij})-Y(y,j;\tau_{ij})|\rceil-1\,.$$

By definition, $z_{\lceil\delta|X(x,i;\tau_{ij})-Y(y,j;\tau_{ij})|\rceil}=Y(y,j;\tau_{ij})$, $z_0,\dots,z_{\lceil\delta|X(x,i;\tau_{ij})-Y(y,j;\tau_{ij})|\rceil}$ are $\mathscr{F}_{\tau_{ij}}$-measurable and

$$|z_{k+1}-z_k|\le\frac{|X(x,i;\tau_{ij})-Y(y,j;\tau_{ij})|}{\lceil\delta|X(x,i;\tau_{ij})-Y(y,j;\tau_{ij})|\rceil}\,,$$

so $f(|z_{k+1}-z_k|)\le\eta$ for $k=0,\dots,\lceil\delta|X(x,i;\tau_{ij})-Y(y,j;\tau_{ij})|\rceil-1$. For $t\ge0$, let $\tilde B(t):=B(\tau_{ij}+t)-B(t)$. Clearly, $\{\tilde B(t)\}_{t\ge0}$ is a Brownian motion. Further, let $\{\tilde X^{(\lceil\delta|X(x,i;\tau_{ij})-Y(y,j;\tau_{ij})|\rceil)}(t)\}_{t\ge0}=\{Y(y,j;\tau_{ij}+t)\}_{t\ge0}$, and for $k=0,\dots,\lceil\delta|X(x,i;\tau_{ij})-Y(y,j;\tau_{ij})|\rceil-1$ let $\{\tilde X^{(k)}(t)\}_{t\ge0}$ be solution to

$$\begin{aligned} d\tilde X^{(k)}(t) &= b\big(\tilde X^{(k)}(t),\Lambda(i;t)\big)dt+\sigma\big(\Lambda(i;t)\big)d\tilde B(t)\\ \tilde X^{(k)}(0) &= z_k\\ \Lambda(i;0) &= i\in\mathbb{S}. \end{aligned}$$

Observe that $\{\tilde X^{(0)}(t)\}_{t\ge0}=\{X(x,i;\tau_{ij}+t)\}_{t\ge0}$. We now have that

$$F(t)\le f\big(|\tilde X^{(0)}(t)-\tilde X^{(1)}(t)|\big)+\dots+f\big(|\tilde X^{(\lceil\delta|X(x,i;\tau_{ij})-Y(y,j;\tau_{ij})|\rceil-1)}(t)-\tilde X^{(\lceil\delta|X(x,i;\tau_{ij})-Y(y,j;\tau_{ij})|\rceil)}(t)|\big)\,,$$

and from the first part of the proof it follows that

$$F(t)\le\lceil\delta|X(x,i;\tau_{ij})-Y(y,j;\tau_{ij})|\rceil\Psi_\eta^{-1}\left(-\int_{\tau_{ij}}^{t+\tau_{ij}}\Gamma_{\Lambda(i;s)}ds\right)\,.$$

By taking expectation we get

$$\mathbb{E}\Big[F(\varepsilon t/(1+\varepsilon))\mathbb{1}_{\{\tau_{ij}\le t/(1+\varepsilon)\}}\mathbb{1}_{\{f(|X(x,i;\tau_{ij})-Y(y,j;\tau_{ij})|)>\eta\}}\Big]$$

$$\le \mathbb{E}\Bigg[\lceil \delta|X(x,i;\tau_{ij})-Y(y,j;\tau_{ij})|\rceil \Psi_\eta^{-1}\left(-\int_{\tau_{ij}}^{\varepsilon t/(1+\varepsilon)+\tau_{ij}}\Gamma_{\Lambda(i;s)}ds\right)$$

$$\mathbb{1}_{\{\tau_{ij}\le t/(1+\varepsilon)\}}\mathbb{1}_{\{f(|X(x,i;\tau_{ij})-Y(y,j;\tau_{ij})|)>\eta\}}\Bigg]$$

$$\le \mathbb{E}\left[\lceil \delta|X(x,i;\tau_{ij})-Y(y,j;\tau_{ij})|\rceil^2\right]^{1/2}$$

$$\mathbb{E}\left[\Psi_\eta^{-1}\left(-\int_{\tau_{ij}}^{\varepsilon t/(1+\varepsilon)+\tau_{ij}}\Gamma_{\Lambda(i;s)}ds\right)^2 \mathbb{1}_{\{\tau_{ij}\le t/(1+\varepsilon)\}}\mathbb{1}_{\{f(|X(x,i;\tau_{ij})-Y(y,j;\tau_{ij})|)>\eta\}}\right]^{1/2}.$$

From Lemma 3.4.4 we now that $\mathbb{E}\left[\lceil \delta|X(x,i;\tau_{ij})-Y(y,j;\tau_{ij})|\rceil^2\right]<\infty$. Thus, analogously as in eq. (3.4.8) we have that

$$\lim_{t\to\infty}\mathbb{E}\left[F(\varepsilon t/(1+\varepsilon))\mathbb{1}_{\{\tau_{ij}\le t/(1+\varepsilon)\}}\mathbb{1}_{\{f(|X(x,i;\tau_{ij})-Y(y,j;\tau_{ij})|)>\eta\}}\right]=0. \qquad (3.4.9)$$

Now, by combining eqs. (3.4.8) and (3.4.9) the first assertion follows.

The cases when $\psi(u)=u^q$ and $\psi(u)=\kappa u$ are treated in a completely the same way as in Theorem 3.4.2. ■

As a consequence of Theorems 3.4.2 and 3.4.3 we conclude the following ergodic behavior of $\{(X(x,i;t),\Lambda(x,i;t)\}_{t\ge 0}$.

Theorem 3.4.5. In addition to the assumptions of Theorem 3.4.2 or Theorem 3.4.3, suppose that there are non-negative $\mathscr{V}\in\mathscr{C}^2(\mathbb{R}^d\times\mathbb{S})$ and locally bounded $g:\mathbb{R}^d\to\mathbb{R}$, such that

$$\lim_{|x|\to\infty} g(x)=\infty \qquad \text{and} \qquad \mathscr{L}\mathscr{V}(x,i)\le -g(x) \qquad (3.4.10)$$

for all $(x,i)\in\mathbb{R}^d\times\mathbb{S}$. Then, $\{(X(x,i;t),\Lambda(x,i;t)\}_{t\ge 0}$ admits a unique invariant probability measure π and

$$\lim_{t\to\infty}\mathscr{W}_{f,p}\left(\delta_{(x,i)}\mathscr{P}_t,\pi\right)=0$$

for all $(x,i)\in\mathbb{R}^d\times\mathbb{S}$. Additionally, if $\psi(u)=u^q$ for some $q>1$, then

$$\lim_{t\to\infty} t^{1/(q-1)}\mathscr{W}_{f,p}\left(\delta_{(x,i)}\mathscr{P}_t,\pi\right)\le\left(\frac{1-q}{2}\sum_{i\in\mathbb{S}}\Gamma_i\lambda_i\right)^{1/(1-q)},$$

and if $\psi(u)=\kappa u$ for some $\kappa>0$, then

$$\lim_{t\to\infty} e^{\alpha t/2}\mathscr{W}_{f,p}\left(\delta_{(x,i)}\mathscr{P}_t,\pi\right)=0$$

for all $0<\alpha<\min\{\vartheta/p,-\kappa\sum_{i\in\mathbb{S}}\Gamma_i\lambda_i\}$. Recall that in the case of Theorem 3.4.3 $p=1$.

Proof. Observe first that eq. (3.4.3) holds for any two initial distributions μ and ν of $\{(X(x,i;t),\Lambda(i;t))\}_{t\geq 0}$, that is,

$$\lim_{t\to\infty}\mathscr{W}_{f,p}\big(\mu\mathscr{P}_t,\nu\mathscr{P}_t\big) = 0\,.$$

From this we conclude that if $\{(X(x,i;t),\Lambda(i;t))\}_{t\geq 0}$ admits an invariant probability measure, then it must be unique. Namely, if π and $\overline{\pi}$ were two invariant probability measures of
$\{(X(x,i;t),\Lambda(i;t))\}_{t\geq 0}$, then

$$\mathscr{W}_{f,p}\big(\pi,\overline{\pi}\big) = \lim_{t\to\infty}\mathscr{W}_{f,p}\big(\pi\mathscr{P}_t,\overline{\pi}\mathscr{P}_t\big) = 0$$

which implies $\pi=\overline{\pi}$. Thus, if $\{(X(x,i;t),\Lambda(i;t))\}_{t\geq 0}$ admits an invariant probability measure π, then

$$\lim_{t\to+\infty}\mathscr{W}_{f,p}\big(\delta_{(x,i)}\mathscr{P}_t,\pi\big) = \lim_{t\to\infty}\mathscr{W}_{f,p}\big(\delta_{(x,i)}\mathscr{P}_t,\pi\mathscr{P}_t\big) = 0\,.$$

In the sequel we show that eq. (3.4.3) guarantees existence of an invariant probability measure of $\{(X(x,i;t),\Lambda(i;t))\}_{t\geq 0}$. According to [MT93a, Theorem 3.1] this will follow if we show that for each $(x,i)\in\mathbb{R}^d\times\mathbb{S}$ and $0<\varepsilon<1$ there is a compact set $C\subset\mathbb{R}^d$ (possibly depending on (x,i) and ε) such that

$$\liminf_{t\to\infty}\frac{1}{t}\int_0^t p(s,(x,i),C\times\mathbb{S})\,ds \geq 1-\varepsilon\,.$$

Let $r>0$ be large enough so that

$$\inf_{x\in B_r^c(0)} g(x) \geq -\inf_{x\in B_r(0)} g(x)\,.$$

Such r exists since $\lim_{|x|\to\infty}g(x)=\infty$. Observe that if the previous relation holds for some r_0, then it also holds for all $r\geq r_0$. We have that

$$\begin{aligned}
\mathscr{L}\mathscr{V}(x,i) &\leq -g(x)\mathbb{1}_{B_r(0)}(x) - g(x)\mathbb{1}_{B_r^c(0)}(x)\\
&\leq \Big(\big(\inf_{x\in B_r^c(0)} g(x)\big)^{1/2} + \inf_{x\in B_r^c(0)} g(x)\Big)\,\mathbb{1}_{B_r(0)}(x) - \frac{1}{2}\inf_{x\in B_r^c(0)} g(x)\mathbb{1}_{B_r^c(0)}(x)\\
&= \Big(\big(\inf_{x\in B_r^c(0)} g(x)\big)^{1/2} + \frac{1}{2}\inf_{x\in B_r^c(0)} g(x)\Big)\,\mathbb{1}_{B_r(0)\times\mathbb{S}}(x,i) - \frac{1}{2}\inf_{x\in B_r^c(0)} g(x)\,.
\end{aligned}$$

Now, according to [MT93b, Theorem 1.1] we conclude that for each (x,i) and r large enough,

$$\liminf_{t\to\infty}\frac{1}{t}\int_0^t p\big(s,(x,i),\overline{B}_r(0)\times\mathbb{S}\big)\,ds \geq \frac{\frac{1}{2}\inf_{x\in B_r^c(0)}g(x)}{\big(\inf_{x\in B_r^c(0)}g(x)\big)^{1/2}+\frac{1}{2}\inf_{x\in B_r^c(0)}g(x)}.$$

The assertion now follows by choosing r large enough. ■

Finally, we discuss sufficient conditions ensuring eq. (3.4.10). First, recall that an $m\times m$ matrix M is called an $\mathscr{M}$-matrix if it can be expressed as $M=\gamma\mathbb{I}_m-N$ for some $\gamma>0$ and some nonnegative $m\times m$ matrix N with the property that $\rho(N)\leq\gamma$, where $\mathbb{I}_m$ and $\rho(N)$ denote the $m\times m$ identity matrix and spectral radius of N. According to the Perron-Frobenius theorem, M is nonsingular if, and only if, $\rho(N)<\gamma$.

Theorem 3.4.6. Assume **(A1)-(A5)** and $\mathscr{Q}(x)=\mathscr{Q}+\mathrm{o}(1)$. Furthermore, assume that there are $\{c_i\}_{i\in\mathbb{S}}\subset\mathbb{R}$ such that either one of the following conditions holds:

(i) $\mathscr{Q}$ is the infinitesimal generator of an irreducible right-continuous temporally- homogeneous Markov chain on $\mathbb{S}$ with invariant probability measure $\lambda=(\lambda_i)_{i\in\mathbb{S}}$, $\sum_{i\in\mathbb{S}}c_i\lambda_i<0$ and there are twice continuously differentiable $\mathscr{V}:\mathbb{R}^d\to(0,\infty)$ and twice continuously differentiable concave $\theta:(0,\infty)\to(0,\infty)$, such that

$$\lim_{|x|\to\infty}\theta\circ\mathscr{V}(x)=\infty,\qquad \limsup_{|x|\to\infty}\frac{\mathscr{L}_i\mathscr{V}(x)}{\theta\circ\mathscr{V}(x)}<c_i,$$
$$\lim_{|x|\to\infty}\frac{\theta\circ\mathscr{V}(x)}{\mathscr{V}(x)}=0,\qquad \lim_{|x|\to\infty}\sup_{i\in\mathbb{S}}\frac{\mathscr{L}_i\theta\circ\mathscr{V}(x)}{\theta\circ\mathscr{V}(x)}=0.$$

(ii) $\mathscr{Q}$ is the infinitesimal generator of an irreducible right-continuous temporally- homogeneous Markov chain on $\mathbb{S}$ with invariant probability measure $\lambda=(\lambda_i)_{i\in\mathbb{S}}$, $\sum_{i\in\mathbb{S}}c_i\lambda_i<0$ and there is a twice continuously differentiable $\mathscr{V}:\mathbb{R}^d\to(0,\infty)$ such that

$$\lim_{|x|\to\infty}\mathscr{V}(x)=\infty\qquad\text{and}\qquad\limsup_{|x|\to\infty}\frac{\mathscr{L}_i\mathscr{V}(x)}{\mathscr{V}(x)}<c_i.\tag{3.4.11}$$

(iii) $(-\mathscr{Q}+\operatorname{diag}c)$ is a non-singular $\mathscr{M}$-matrix, where $c=(c_i)_{i\in\mathbb{S}}$, and there is a twice continuously differentiable $\mathscr{V}:\mathbb{R}^d\to(0,\infty)$ satisfying eq. (3.4.11).

Then there are non-negative $\mathscr{V}\in\mathscr{C}^2(\mathbb{R}^d\times\mathbb{S})$ and locally bounded $g:\mathbb{R}^d\to\mathbb{R}$, such that eq. (3.4.10) holds.

Proof. In case (i), analogously as in the proof of Theorem 3.3.10 we conclude that there is a non-negative $\mathscr{V}\in\mathscr{C}^2(\mathbb{R}^d\times\mathbb{S})$ such that

$$\lim_{|x|\to\infty}\inf_{i\in\mathbb{S}}\theta\circ\mathscr{V}(x,i)=\infty\qquad\text{and}\qquad\mathscr{L}\mathscr{V}(x,i)\le-\theta\circ\mathscr{V}(x,i)$$

for all $i\in\mathbb{S}$ and $|x|$ large enough. In cases (ii) and (iii), by the same reasoning as in the proof of Proposition 3.3.11 we see that there are $\eta>0$ and non-negative $\mathscr{V}\in\mathscr{C}^2(\mathbb{R}^d\times\mathbb{S})$, such that

$$\lim_{|x|\to\infty}\inf_{i\in\mathbb{S}}\mathscr{V}(x,i)=\infty\qquad\text{and}\qquad\mathscr{L}\mathscr{V}(x,i)\le-\eta\,\mathscr{V}(x,i)$$

for all $i\in\mathbb{S}$ and $|x|$ large enough. The desired result now follows by setting $f(x):=\inf_{i\in\mathbb{S}}\theta\circ\mathscr{V}(x,i)$ in the first case, and $f(x):=\eta\inf_{i\in\mathbb{S}}\mathscr{V}(x,i)$ in the second and third case. ■

Typical examples satisfying conditions of Theorems 3.4.2, 3.4.3, 3.4.5 and 3.4.6 are given as follows.

Example 3.4.7. (i) Let $\mathbb{S}=\{0,1\}$, let

$$b(x,i)=\begin{cases}b\,, & i=0\,,\\ -\mathrm{sgn}(x)|x|^q\,, & i=1\,,\end{cases}$$

with $b\in\mathbb{R}$ and $q>1$, and let $\sigma(x,i)\equiv0$. The processes

$$\begin{aligned}dX^{(0)}(x;t)&=b\big(X^{(0)}(x;t),0\big)dt+\sigma\big(X^{(0)}(x;t),0\big)dB(t)=b\,dt\\ X^{(0)}(x;0)&=x\in\mathbb{R}\,,\end{aligned}$$

and

$$\begin{aligned}dX^{(1)}(x;t)&=b\big(X^{(1)}(x;t),1\big)dt+\sigma\big(X^{(1)}(x;t),1\big)dB(t)\\ &=-\mathrm{sgn}\big(X^{(1)}(x;t)\big)|X^{(1)}(x;t)|^q\,dt\\ X^{(1)}(x;0)&=x\in\mathbb{R}\,,\end{aligned}$$

are given by $X^{(0)}(x;t)=x+bt$ and

$$X^{(1)}(x;t)=\begin{cases}\mathrm{sgn}(x)\big(|x|^{1-q}+(q-1)t\big)^{1/(1-q)}\,, & x\neq0\,,\\ 0\,, & x=0\,.\end{cases}$$

Clearly, both $\{X^{(0)}(x;t)\}_{t\geq 0}$ and $\{X^{(1)}(x;t)\}_{t\geq 0}$ are not irreducible and aperiodic. Hence, we cannot apply Theorem 3.3.10 to these processes. In the case when $b\neq 0$ the process $\{X^{(0)}(x;t)\}_{t\geq 0}$ does not admit an invariant probability measure, while in the case when $b=0$ it admits uncountably many invariant probability measures: $\{\delta_{\{x\}}\}_{x\in\mathbb{R}}$. On the other hand, $\delta_{\{0\}}$ is a unique invariant probability measure for $\{X^{(1)}(x;t)\}_{t\geq 0}$. However, convergence of the corresponding semigroup to $\delta_{\{0\}}$ (with respect to some distance function) cannot have exponential rate and this convergence cannot hold in the total variation norm.

Let now $\mathrm{q}_{01}=\mathrm{q}_{10}=1$. Hence, $\lambda=(1/2,1/2)$. The process $\{(X(x,i;t),\Lambda(i;t))\}_{t\geq 0}$ is also not irreducible and aperiodic (hence, we cannot apply Theorem 3.3.10), and since $\|\delta_{(x,i)}\mathscr{P}_t-\delta_{(y,i)}\mathscr{P}_t\|_{\mathrm{TV}}=1$ for all $i\in\mathbb{S}$, $x\neq y$ and $t\geq 0$, the semigroup cannot converge to the corresponding invariant probability measure (if it exists) in the total variation norm. The previous discussion suggest that this convergence (with respect to some distance function) cannot have exponential rate. Observe that in the case when $b=0$ the unique invariant probability measure for $\{(X(x,i;t),\Lambda(i;t))\}_{t\geq 0}$ is $\delta_{\{0\}}\times\lambda$.

Let $f(u)=u$ for all u small enough and $f(u)=1-1/(1+u)$ for all u large enough, and let $\psi(u)=u^q$ (with $q>1$). Obviously, $b(x,0)$ satisfies eq. (3.4.2) with $\Gamma_0=0$, and an elementary computation shows that $b(x,1)$ satisfies eq. (3.4.2) with some $\Gamma_1<0$. Hence, we can apply Theorem 3.4.2 to $\{(X(x,i;t),\Lambda(i;t))\}_{t\geq 0}$. Further, take $\mathscr{V}(x)=x^2$ and observe that

$$\mathscr{L}_0\mathscr{V}(x) = 2bx \qquad \text{and} \qquad \mathscr{L}_1\mathscr{V}(x) = -2|x|^{q+1}.$$

Thus, for arbitrary small $c_0>0$ and arbitrary large $-c_1>0$ (recall that $q>1$) it holds that

$$\mathscr{L}_0\mathscr{V}(x) \leq c_0\mathscr{V}(x) \qquad \text{and} \qquad \mathscr{L}_1\mathscr{V}(x) \leq c_1\mathscr{V}(x)$$

for all $|x|$ large enough. Hence, according to Theorems 3.4.5 and 3.4.6 the process $\{(X(x,i;t),\Lambda(i;t))\}_{t\geq 0}$ admits a unique invariant probability measure π and the corresponding semigroup converges to π with respect to $\mathscr{W}_{f,p}$ with subgeometric rate $t^{1/(q-1)}$.

(ii) Let $b(x,i)$, $\{\Lambda(i;t)\}_{t\geq 0}$ and $\psi(u)$ be as in (i). Further, let $\sigma(x,i)\equiv\sigma(i)$, $\eta\in(0,1)$ and $f(u)=u$. Observe that

$$\begin{aligned} dX^{(0)}(x;t) &= b\,dt+\sigma(0)dB(t)\\ X^{(0)}(x;0) &= x\in\mathbb{R} \end{aligned}$$

is transient if $b\neq 0$ (as a deterministic drift process or Brownian motion with drift) and nullrecurrent if $b=0$ (as a trivial process or Brownian motion). In [LS21, Example 3.3] it has been shown that

$$\begin{aligned} dX^{(1)}(x;t) &= -\mathrm{sgn}\big(X^{(1)}(x;t)\big)|X^{(1)}(x;t)|^q\,dt+\sigma(1)dB(t)\\ X^{(1)}(x;0) &= x\in\mathbb{R} \end{aligned}$$

is subgeometrically ergodic with respect to $\mathscr{W}_{f,1}$ with rate $t^{1/(q-1)}$. Further, obviously $b(x,0)$ satisfies eq. (3.4.5) with $\Gamma_0=0$, and an elementary computation shows that $b(x,1)$ satisfies eq. (3.4.5) with some $\Gamma_1<0$ for all $x,y\in\mathbb{R}$ satisfying $f(|x-y|)=|x-y|\leq\eta$. Hence, $\{(X(x,i;t),\Lambda(i;t))\}_{t\geq 0}$ satisfies assumptions of Theorem 3.4.3. Finally, by completely the same reasoning as in (i) we again conclude that $\{(X(x,i;t),\Lambda(i;t))\}_{t\geq 0}$ is subgeometrically ergodic with respect to $\mathscr{W}_{f,1}$ with rate $t^{1/(q-1)}$.

3.4.1. Regime-switching Markov processes with jumps

In this section, we briefly discuss ergodicity properties of a class of regime-switching Markov processes with jumps.

Firstly, we consider the case of jump-process obtained through the Bochner's subordination method.

Let now $\{\mathsf{S}(t)\}_{t\geq 0}$ be a subordinator with characteristic exponent $\phi(u)$, independent of $\{(X(x,i;t),\Lambda(x,i;t))\}_{t\geq 0}$. Recall, the process

$$(X^{\phi}(x,i;t),\Lambda^{\phi}(x,i;t)):=(X(x,i;\mathsf{S}(t)),\Lambda(x,i;\mathsf{S}(t))),\quad t\geq 0$$

is obtained from $\{(X(x,i;t),\Lambda(x,i;t))\}_{t\geq 0}$ by a random time change through $\{\mathsf{S}(t)\}_{t\geq 0}$. Also, as marked before, it is known that many fine properties of Markov processes (and the corresponding semigroups) are preserved under subordination. It is easy to see that $\{(X^{\phi}(x,i;t),\Lambda^{\phi}(x,i;t))\}_{t\geq 0}$ is again a Markov process with transition kernel

$$p^{\phi}\big(t,(x,i),dy\times\{j\}\big) = \int_{[0,\infty)} p\big(s,(x,i),dy\times\{j\}\big)\,\mu_t(ds)\,,$$

where $\mu_t(\cdot)=\mathbb{P}(\mathsf{S}(t)\in\cdot)$ is the transition probability of $\mathsf{S}(t)$, $t\geq 0$. Also, it is elementary to check that if π is an invariant probability measure for $\{(X(x,i;t),\Lambda(x,i;t))\}_{t\geq 0}$, then it is also invariant for the subordinate process $\{(X^{\phi}(x,i;t),\Lambda^{\phi}(x,i;t))\}_{t\geq 0}$. In [DSS17] and [LS21, Proposition 3.7] it has been shown that if $\{(X(x,i;t),\Lambda(x,i;t))\}_{t\geq 0}$ is subgeometrically ergodic with Borel measurable rate $r(t)$ (with respect to the total variation distance or an $\mathscr{L}^p$-Wasserstein distance), then $\{(X^{\phi}(x,i;t),\Lambda^{\phi}(x,i;t))\}_{t\geq 0}$ is subgeometrically ergodic with rate $r_{\phi}(t)=\mathbb{E}[r(\mathsf{S}(t))]$ (in the total variation distance case) and $r_{\phi}(t)=(\mathbb{E}[r^p(\mathsf{S}(t))])^{1/p}$ (in the $\mathscr{L}^p$-Wasserstein distance case). Therefore, as an direct application of Theorems 2.3.1 and 3.4.5 we obtain subgeometric ergodicity results for a class of subordinate regime-switching diffusion processes.

Second approach, also discussed before for diffusion processes, is to replace the Brownian motion $\{B(t)\}_{t\geq 0}$ in eq. (3.0.1) by a general Lévy process and adapt Theorems 3.4.2, 3.4.3, 3.4.5 and 3.4.6.

Let $\{\mathsf{L}(t)\}_{t\geq 0}$ be an n-dimensional Lévy process (starting from the origin) with Lévy triplet (β,γ,ν). Consider the regime-switching jump-diffusion process $\{(X(x,i;t),\Lambda(i;t))\}_{t\geq 0}$ with the first component given by

$$\begin{aligned} dX(x,i;t) &= b\big(X(x,i;t),\Lambda(i;t)\big)dt+\sigma\big(\Lambda(i;t-)\big)d\mathsf{L}(t)\\ X(x,i;0) &= x\in\mathbb{R}^d \\ \Lambda(i;0) &= i\in\mathbb{S}\,, \end{aligned} \tag{3.4.12}$$

and the second component, as before, is a right-continuous temporally-homogeneous Markov chain with finite state space $\mathbb{S}$. The processes $\{\mathsf{L}(t)\}_{t\geq 0}$ and $\{\Lambda(i;t)\}_{t\geq 0}$ are independent and defined on a stochastic basis $(\Omega,\mathscr{F},\{\mathscr{F}_t\}_{t\geq 0},\mathbb{P})$ (satisfying the usual conditions). Assume that the coefficients $b:\mathbb{R}^d\times\mathbb{S}\to\mathbb{R}^d$ and $\sigma:\mathbb{S}\to\mathbb{R}^{d\times n}$, and the process $\{\Lambda(i;t)\}_{t\geq 0}$ satisfy the following:

$(\widetilde{\mathbf{A1}})$ for any $r>0$ and $i\in\mathbb{S}$,

$$\sup_{x\in B_r(0)}|b(x,i)|<\infty$$

($\widetilde{\mathbf{A2}}$) for each $(x,i) \in \mathbb{R}^d \times \mathbb{S}$ the RSSDE in eq. (3.4.12) admits a unique nonexplosive strong solution $\{X(x,i;t)\}_{t\geq 0}$ which has càdlàg sample paths

($\widetilde{\mathbf{A3}}$) the process $\{(X(x,i;t),\Lambda(i;t))\}_{t\geq 0}$ is a temporally-homogeneous strong Markov process with transition kernel $p(t,(x,i),dy\times\{j\}) = \mathbb{P}((X(x,i;t),\Lambda(i;t)) \in dy \times \{j\})$

($\widetilde{\mathbf{A4}}$) the corresponding semigroup of linear operators $\{\mathscr{P}_t\}_{t\geq 0}$ satisfies the $\mathscr{C}_b$-Feller property

($\widetilde{\mathbf{A5}}$) for any $(x,i) \in \mathbb{R}^d \times \mathbb{S}$ and $f \in \mathscr{C}^2(\mathbb{R}^d \times \mathbb{S})$ such that $(x,i) \mapsto \int_{\mathbb{R}^d} f(x+y,i)\nu_i(dy)$ is locally bounded, the process

$$\left\{ f\big(X(x,i;t),\Lambda(x,i;t)\big) - f(x,i) - \int_0^t \mathscr{L} f\big(X(x,i;s),\Lambda(x,i;s)\big)ds \right\}_{t\geq 0}$$

is a $\mathbb{P}$-local martingale, where $\nu_i(B) = \nu(\{x \in \mathbb{R}^n\colon \sigma(i)x \in B\})$ for $B \in \mathfrak{B}(\mathbb{R}^d)$ and

$$\mathscr{L} f(x,i) = \mathscr{L}_i f(x,i) + \mathscr{Q} f(x,i)$$

with

$$\begin{aligned}\mathscr{L}_i f(x) = &\left\langle b(x,i) + \sigma(i)\beta + \int_{\mathbb{R}^n} \sigma(i)y\big(\mathbb{1}_{B_1(0)}(\sigma(i)y) - \mathbb{1}_{B_1(0)}(y)\big)\nu(dy), \nabla f(x) \right\rangle \\ &+ \frac{1}{2}\mathrm{Tr}\left[\sigma(i)\gamma\sigma(i)^T \nabla^2 f(x)\right] \\ &+ \int_{\mathbb{R}^d} \big(f(x+y) - f(x) - \langle y, \nabla f(x)\rangle \mathbb{1}_{B_1(0)}(y)\big)\nu_i(dy)\end{aligned}$$

and $\mathscr{Q} = (\mathrm{q}_{ij})_{i,j\in\mathbb{S}}$ being the infinitesimal generator of the process $\{\Lambda(i;t)\}_{t\geq 0}$.

⤞ Conditions ensuring **($\widetilde{\mathbf{A1}}$)-($\widetilde{\mathbf{A5}}$)** can be found in [FGC19] (see also [KZ20]).

It is straightforward to check that Theorem 3.4.2 (and Theorems 3.4.5 and 3.4.6) holds also in this situation (under the additional assumption that the functions $\mathscr{V}(x,i)$, $\mathscr{V}(x)$ and $\theta(u)$ appearing in Theorems 3.4.5 and 3.4.6 are such that $(x,i) \mapsto \int_{\mathbb{R}^d} \mathscr{V}(x+y,i)\nu_i(dy)$, $(x,i) \mapsto \int_{\mathbb{R}^d} \mathscr{V}(x+y)\nu_i(dy)$ and $(x,i) \mapsto \int_{\mathbb{R}^d} \theta \circ \mathscr{V}(x+y)\nu_i(dy)$ are locally bounded). On the other hand, in order to conclude the results from Theorem 3.4.3 we need to extend the results from Lemma 3.4.4 to the jump case.

Lemma 3.4.8. Assume that $\int_{\mathbb{R}^n}(|y|^2\vee|y|^4)\nu(dy)<\infty$ (or, equivalently, $\mathbb{E}[|\mathsf{L}_t|^4]<\infty$ for all $t\geq 0$) and

$$\begin{aligned}&2\big\langle x,\mathrm{b}(x,i)+\sigma(i)\beta+\int_{\mathbb{R}^n}\sigma(i)y\big(\mathbb{1}_{\mathbb{R}^n}\big(\sigma(i)y\big)-\mathbb{1}_{B_1(0)}(y)\big)\nu(dy)\big\rangle\\&+\mathrm{Tr}\big(\sigma(i)\gamma\sigma(i)^T\big)+\int_{\mathbb{R}^d}|y|^2\nu_i(dy)\\&\leq K(1+|x|^2).\end{aligned}\tag{3.4.13}$$

Then,

$$\mathbb{E}\big[|X(x,i;t)|^2\big]\ \leq\ \big(1+|x|^2\big)\mathrm{e}^{Kt}.$$

Furthermore, for any $\{\mathscr{F}_t\}_{t\geq0}$-stopping time τ such that $\mathbb{E}\big[\mathrm{e}^{2K\tau}\big]<\infty$ it follows that

$$\begin{aligned}&\mathbb{E}\left[|X(x,i;\tau)|^2\right]\\&\leq|x|^2+\big(1+|x|^2\big)\mathbb{E}\left[\mathrm{e}^{K\tau}\right]\\&\quad+\big(1+|x|^2\big)\mathbb{E}\left[\mathrm{e}^{2K\tau}\right]^{1/2}\Bigg(4\sup_{j\in\mathbb{S}}\mathrm{Tr}\big(\sigma(j)\sigma(j)^T\big)+\sup_{j\in\mathbb{S}}\mathrm{Tr}\big(\sigma(j)\sigma(j)^T\big)^2\int_{\mathbb{R}^n}|y|^4\nu(dy)/K\\&\quad+4\sup_{j\in\mathbb{S}}\mathrm{Tr}\big(\sigma(j)\sigma(j)^T\big)\int_{\mathbb{R}^n}|y|^2\nu(dy)+4\sup_{j\in\mathbb{S}}\mathrm{Tr}\big(\sigma(j)\sigma(j)^T\big)^{3/2}\int_{\mathbb{R}^n}|y|^3\nu(dy)\Bigg)^{1/2}.\end{aligned}$$

Proof. For $n\in\mathbb{N}$, let $f_n:\mathbb{R}^d\to[0,+\infty\rangle$ be such that $f_n\in\mathscr{C}_b^2(\mathbb{R}^d)$ (the space of bounded and twice continuously differentiable functions with bounded first and second order derivatives), $f_n(x)=|x|^2$ on $B_{n+1}(0)$ and $f_n(x)\leq f_{n+1}(x)$ for all $x\in\mathbb{R}^d$, and

$$\tau_n\ :=\ \inf\{t\geq0:|X(x,i;t)|\geq n\}.$$

Further, for $t>0$ and $B\in\mathfrak{B}(\mathbb{R}^n)$ denote

$$\mathrm{N}((0,t],B)\ :=\sum_{0<s\leq t}\mathbb{1}_B(\mathsf{L}(s)-\mathsf{L}(s-))\qquad\text{and}\qquad\tilde{\mathrm{N}}(dt,dy)\ :=\ \mathrm{N}(ds,dy)-\nu(dy)ds,$$

as we defined in the first chapter the poisson random measure and its compensated version. By employing Itô's formula and the assumption that $\int_{\mathbb{R}^n}(|y|^2\vee|y|^4)\nu(dy)<\infty$ we conclude that for n large enough,

$$\begin{aligned}&f_n\big(X(x,i;t\wedge\tau_n)\big)\\&=f_n(x)+\int_0^{t\wedge\tau_n}\Bigg(\Big\langle\nabla f_n\big(X(x,i;s)\big),b\big(X(x,i;s),\Lambda(i;s)\big)+\sigma\big(\Lambda(i;s)\big)\beta\end{aligned}$$

$$
\begin{aligned}
&\qquad\qquad + \int_{\mathbb{R}^n} \sigma\big(\Lambda(i;s)\big) y \big(\mathbb{1}_{\mathbb{R}^n}\big(\sigma\big(\Lambda(i;s)\big)y\big) - \mathbb{1}_{B_1(0)}(y)\big)\nu(dy)\Big\rangle \\
&+ \frac{1}{2}\mathrm{Tr}\big(\sigma\big(\Lambda(i;s)\big)\gamma\sigma\big(\Lambda(i;s)\big)^T \nabla^2 f_n\big(X(x,i;s)\big)\big) \\
&+ \int_{\mathbb{R}^d} \big(f_n\big(X(x,i;s)+y\big) - f_n\big(X(x,i;s)\big) - \big\langle y, \nabla f_n\big(X(x,i;s)\big)\big\rangle\big)\,\nu_{\Lambda(i;s)}(dy)\Bigg)ds \\
&+ \int_0^{t\wedge\tau_n} \nabla f_n\big(X(x,i;s)\big)^T \sigma\big(\Lambda(i;s)\big)dB(s) \\
&+ \int_0^{t\wedge\tau_n}\int_{\mathbb{R}^n} \big(f_n\big(X(x,i;s-)+\sigma\big(\Lambda(i;s-)\big)y\big) - f_n\big(X(x,i;s-)\big)\big)\tilde{\mathrm{N}}(dy,ds) \\
= {}& |x|^2 + 2\int_0^{t\wedge\tau_n}\Bigg(\Big\langle X(x,i;s), b\big(X(x,i;s),\Lambda(i;s)\big) + \sigma\big(\Lambda(i;s)\big)\beta \\
&\qquad\qquad + \int_{\mathbb{R}^n} \sigma\big(\Lambda(i;s)\big) y \big(\mathbb{1}_{\mathbb{R}^n}\big(\sigma\big(\Lambda(i;s)\big)y\big) - \mathbb{1}_{B_1(0)}(y)\big)\nu(dy)\Big\rangle \\
&+ \mathrm{Tr}\big(\sigma\big(\Lambda(i;s)\big)\gamma\sigma\big(\Lambda(i;s)\big)^T\big) + \int_{\mathbb{R}^d} |y|^2 \nu_{\Lambda(i;s)}(dy)\Bigg)ds \\
&+ \int_0^{t\wedge\tau_n} \nabla f_n\big(X(x,i;s)\big)^T \sigma\big(\Lambda(i;s)\big)dB(s) \\
&+ \int_0^{t\wedge\tau_n}\int_{\mathbb{R}^n} \big(f_n\big(X(x,i;s-)+\sigma\big(\Lambda(i;s-)\big)y\big) - f_n\big(X(x,i;s-)\big)\big)\tilde{\mathrm{N}}(dy,ds) \\
\le {}& |x|^2 + K\int_0^t \big(1 + |X(x,i;s)|^2 \mathbb{1}_{[0,\tau_n]}(s)\big)\,ds \\
&+ \int_0^{t\wedge\tau_n} \nabla f_n\big(X(x,i;s)\big)^T \sigma\big(\Lambda(i;s)\big)dB(s) \\
&+ \int_0^{t\wedge\tau_n}\int_{\mathbb{R}^n} \big(f_n\big(X(x,i;s-)+\sigma\big(\Lambda(i;s-)\big)y\big) - f_n\big(X(x,i;s-)\big)\big)\tilde{\mathrm{N}}(dy,ds)\,,
\end{aligned}
$$

where in the last step we used eq. (3.4.13). By taking expectation, we have that

$$
\begin{aligned}
&1 + \mathbb{E}\left[|X(x,i;t)|^2 \mathbb{1}_{[0,\tau_n)}(t)\right] \\
&\le 1 + \mathbb{E}\left[f_n\big(X(x,i;t\wedge\tau_n)\big)\right] \\
&\le 1 + |x|^2 + K\int_0^t \big(1 + \mathbb{E}\left[|X(x,i;s)|^2\right] \mathbb{1}_{[0,\tau_n]}(s)\big)ds \\
&= 1 + |x|^2 + K\int_0^t \big(1 + \mathbb{E}\left[|X(x,i;s)|^2 \mathbb{1}_{[0,\tau_n)}(s)\right]\big)ds\,.
\end{aligned}
$$

The first assertion now follows by employing Grönwall's inequality and Fatou's lemma.

Let now τ be a stopping time such that $\mathbb{E}[\mathrm{e}^{2K\tau}] < \infty$. Itô's lemma then gives

$$
\begin{aligned}
|X(x,i;t)|^2 \le {}& |x|^2 + Kt + K\int_0^t |X(x,i;s)|^2 ds + 2\int_0^t X(x,i;s)^T \sigma\big(\Lambda(i;s)\big)dB(s) \\
&+ \int_0^t\int_{\mathbb{R}^n} \big(|\sigma\big(\Lambda(i;s-)\big)y|^2 + 2X(x,i;s-)^T\sigma\big(\Lambda(i;s-)\big)y\big)\tilde{\mathrm{N}}(dy,ds)\,.
\end{aligned}
$$

Denote

$$\begin{aligned}\alpha(t) := 2\int_0^t X(x,i;s)^T\sigma\big(\Lambda(i;s)\big)dB(s)\\ +\int_0^t\int_{\mathbb{R}^n}\big(|\sigma\big(\Lambda(i;s-)\big)y|^2+2X(x,i;s-)^T\sigma\big(\Lambda(i;s-)\big)y\big)\tilde{\mathrm{N}}(dy,ds)\,.\end{aligned}$$

Grönwall's inequality then gives

$$|X(x,i;t)|^2 \le |x|^2+Kt+\alpha(t)+\int_0^t K\left(|x|^2+Ks+\alpha(s)\right)\mathrm{e}^{K(t-s)}ds\,.$$

Consequently,

$$\begin{aligned}&|X(x,i;t\wedge\tau)|^2\\ &\le |x|^2+K(t\wedge\tau)+\alpha(t\wedge\tau)+|x|^2\big(\mathrm{e}^{Kt\wedge\tau}-1\big)+\mathrm{e}^{Kt\wedge\tau}-K(t\wedge\tau)\\ &\quad +K\int_0^t\left(\mathbb{1}_{[0,\tau]}(s)\,\mathrm{e}^{K(t\wedge\tau-s)}\alpha(s)\right)ds\\ &\le |x|^2+\alpha(t\wedge\tau)+\big(1+|x|^2\big)\mathrm{e}^{K\tau}+K\int_0^t\left(\mathbb{1}_{[0,\tau]}(s)\,\mathrm{e}^{K(t\wedge\tau-s)}\alpha(s)\right)ds\,.\end{aligned}$$

By taking expectation, we have that

$$\begin{aligned}&\mathbb{E}\big[|X(x,i;t\wedge\tau)|^2\big]\\ &\le |x|^2+\big(1+|x|^2\big)\mathbb{E}\left[\mathrm{e}^{K\tau}\right]+K\int_0^t\mathbb{E}\left[\mathbb{1}_{[0,\tau]}(s)\,\mathrm{e}^{K(t\wedge\tau-s)}\alpha(s)\right]ds\\ &\le |x|^2+\big(1+|x|^2\big)\mathbb{E}\left[\mathrm{e}^{K\tau}\right]+K\int_0^t\mathbb{E}\left[\mathbb{1}_{[0,\tau]}(s)\,\mathrm{e}^{2K(t\wedge\tau-s)}\right]^{1/2}\mathbb{E}\left[\alpha(s)^2\right]^{1/2}ds\\ &\le |x|^2+\big(1+|x|^2\big)\mathbb{E}\left[\mathrm{e}^{K\tau}\right]\\ &\quad +K\mathbb{E}\left[\mathrm{e}^{2K\tau}\right]^{1/2}\int_0^t\mathrm{e}^{-Ks}\Bigg(4\sup_{j\in\mathbb{S}}\mathrm{Tr}\big(\sigma(j)\sigma(j)^T\big)\mathbb{E}\left[\int_0^s|X(x,i;u)|^2du\right]\\ &\quad +\sup_{j\in\mathbb{S}}\mathrm{Tr}\big(\sigma(j)\sigma(j)^T\big)^2 s\int_{\mathbb{R}^n}|y|^4\nu(dy)\\ &\quad +4\sup_{j\in\mathbb{S}}\mathrm{Tr}\big(\sigma(j)\sigma(j)^T\big)\int_{\mathbb{R}^n}|y|^2\nu(dy)\,\mathbb{E}\left[\int_0^s|X(x,i;u)|^2du\right]\\ &\quad +4\sup_{j\in\mathbb{S}}\mathrm{Tr}\big(\sigma(j)\sigma(j)^T\big)^{3/2}\int_{\mathbb{R}^n}|y|^3\nu(dy)\,\mathbb{E}\left[\int_0^s|X(x,i;u)|du\right]\Bigg)^{1/2}ds\\ &\le |x|^2+\big(1+|x|^2\big)\mathbb{E}\left[\mathrm{e}^{K\tau}\right]\\ &\quad +K\mathbb{E}\left[\mathrm{e}^{2K\tau}\right]^{1/2}\int_0^\infty\mathrm{e}^{-Ks}\Bigg(4\sup_{j\in\mathbb{S}}\mathrm{Tr}\big(\sigma(j)\sigma(j)^T\big)\big(1+|x|\big)^2\mathrm{e}^{Ks}\\ &\quad +\sup_{j\in\mathbb{S}}\mathrm{Tr}\big(\sigma(j)\sigma(j)^T\big)^2 s\int_{\mathbb{R}^n}|y|^4\nu(dy)\end{aligned}$$

$$
\begin{aligned}
&+4\sup_{j\in\mathbb{S}}\mathrm{Tr}\big(\sigma(j)\sigma(j)^T\big)\int_{\mathbb{R}^n}|y|^2\nu(dy)\big(1+|x|\big)^2\mathrm{e}^{Ks}\\
&+4\sup_{j\in\mathbb{S}}\mathrm{Tr}\big(\sigma(j)\sigma(j)^T\big)^{3/2}\int_{\mathbb{R}^n}|y|^3\nu(dy)\big(1+|x|\big)\mathrm{e}^{Ks/2}\Bigg)^{1/2}ds\\
\leq\ & |x|^2+\big(1+|x|^2\big)\mathbb{E}\left[\mathrm{e}^{K\tau}\right]\\
&+\mathbb{E}\left[\mathrm{e}^{2K\tau}\right]^{1/2}\Bigg(4\sup_{j\in\mathbb{S}}\mathrm{Tr}\big(\sigma(j)\sigma(j)^T\big)\big(1+|x|\big)^2+\sup_{j\in\mathbb{S}}\mathrm{Tr}\big(\sigma(j)\sigma(j)^T\big)^2\int_{\mathbb{R}^n}|y|^4\nu(dy)/K\\
&+4\sup_{j\in\mathbb{S}}\mathrm{Tr}\big(\sigma(j)\sigma(j)^T\big)\int_{\mathbb{R}^n}|y|^2\nu(dy)\big(1+|x|\big)^2\\
&+4\sup_{j\in\mathbb{S}}\mathrm{Tr}\big(\sigma(j)\sigma(j)^T\big)^{3/2}\int_{\mathbb{R}^n}|y|^3\nu(dy)\big(1+|x|\big)\Bigg)^{1/2},
\end{aligned}
$$

where in the third step we used Itô's isometry and in the fourth step we used the first assertion of the lemma. ■

Theorem 3.4.3 now follows by replacing eq. (3.4.4) by eq. (3.4.13) and $\int_{\mathbb{R}^n}(|y|^2\vee|y|^4)\nu(dy)<\infty$.

CONCLUSION

The aim of my thesis was to explore the ergodicity of a wide range of diffusion processes with respect to two distance functions: the total variation distance and the class of Wasserstein distances. The focus was placed on identifying conditions on the drift and diffusion coefficients which result in sub-geometric ergodicity of the corresponding semigroup. In order to consider convergence with respect to the total variation distance, the process needed to possesses certain regularity properties (open-set irreducibility and aperiodicty). In cases when that was not true, convergence was considered in some weaker sense, namely, with respect to the Wasserstein distances.

The sub-geometric ergodicity was discussed for two types of processes: diffusion processes and diffusion processes with switching. In each case, I obtained sharp conditions on the coefficients of the process that implied sub-geometric ergodicity with respect to both distance functions. All results were followed by examples that illustrated application to some specific choice of coefficients. Since structural properties of the process are crucial in obtaining the convergence, I also provided several sufficient conditions that ensure those properties for each type of the process. Finally, the results were also extended to a situation when the trajectory of the process is not continuous, that is, when the processes jumps at some random time. In each case, I considered two types of jump processes.

BIBLIOGRAPHY

[ABG12] A. Arapostathis, V. S. Borkar, and M. K. Ghosh. *Ergodic control of diffusion processes*. Cambridge University Press, Cambridge, 2012. ↑ 83.

[ABW10] S. Albeverio, Z. Brzeźniak, and J. Wu. Existence of global solutions and invariant measures for stochastic differential equations driven by Poisson type noise with non-Lipschitz coefficients. *J. Math. Anal. Appl.*, 371(1):309–322, 2010. ↑ 89, 90.

[APS19] A. Arapostathis, G. Pang, and N. Sandrić. Ergodicity of a Lévy–driven SDE arising from multiclass many-server queues. *Ann. Appl. Probab.*, 29(2):1070–1126, 2019. ↑ 73, 90.

[AV10] N. Abourashchi and A. Yu. Veretennikov. On stochastic averaging and mixing. *Theory Stoch. Process.*, 16(1):111–129, 2010. ↑ 53.

[BC86] R. F. Bass and M. Cranston. The Malliavin calculus for pure jump processes and applications to local time. *Ann. Probab.*, 14(2):490–532, 1986. ↑ 73.

[BG68] R. M. Blumenthal and R. K. Getoor. *Markov processes and potential theory*. Academic Press, New York-London, 1968. ↑ 45.

[BGG12] F. Bolley, I. Gentil, and A. Guillin. Convergence to equilibrium in Wasserstein distance for Fokker-Planck equations. *J. Funct. Anal.*, 263(8):2430–2457, 2012. ↑ 87.

[Bha78] R. N. Bhattacharya. Criteria for recurrence and existence of invariant measures for multidimensional diffusions. *Ann. Probab.*, 6(4):541–553, 1978. ↑ 65, 66, 103.

[BSW13] B. Böttcher, R. L. Schilling, and J. Wang. *Lévy matters. III*. Springer, Cham, 2013. ↑ 72, 75.

[But14] O. Butkovsky. Subgeometric rates of convergence of Markov processes in the Wasserstein metric. *Ann. Appl. Probab.*, 24(2):526–552, 2014. ↑ 56.

[CCTY19] X. Chen, Z. C. Chen, K. Tran, and G. Yin. Properties of switching jump diffusions: Maximum principles and harnack inequalities, 2019. ↑ 93.

[CH15] B. Cloez and M. Hairer. Exponential ergodicity for markov processes with random switching. *Bernoulli*, 21(1), Feb 2015. ↑ 96.

[Che00] M. Chen. Explicit bounds of the first eigenvalue. *Sci. China Ser. A*, 43(10):1051–1059, 2000. ↑ 68, 115.

[Che05] M. Chen. *Eigenvalues, inequalities, and ergodic theory*. Springer-Verlag London, Ltd., London, 2005. ↑ 79, 81, 87.

[CT04] R. Cont and P. Tankov. *Financial modelling with jump processes*. Chapman and Hall/CRC, 2004. ↑ 21.

[DFG09] R. Douc, G. Fort, and A. Guillin. Subgeometric rates of convergence of f-ergodic strong Markov processes. *Stochastic Process. Appl.*, 119(3):897–923, 2009. ↑ 49, 50, 51, 56, 64, 66, 75, 117.

[DFM16] A. Durmus, G. Fort, and É. Moulines. Subgeometric rates of convergence in Wasserstein distance for Markov chains. *Ann. Inst. Henri Poincaré Probab. Stat.*, 52(4):1799–1822, 2016. ↑ 56.

[DMT95] D. Down, S. P. Meyn, and R. L. Tweedie. Exponential and uniform ergodicity of Markov processes. *Ann. Probab.*, 23(4):1671–1691, 1995. ↑ 56, 71, 118.

[DSS17] C. Deng, R. L. Schilling, and Y. Song. Subgeometric rates of convergence for Markov processes under subordination. *Adv. in Appl. Probab.*, 49(1):162–181, 2017. ↑ 78, 138.

[Dur96] R. Durrett. *Stochastic calculus*. CRC Press, Boca Raton, FL, 1996. ↑ 5, 35, 38, 57, 58, 65.

[Ebe11] A. Eberle. Reflection coupling and Wasserstein contractivity without convexity. *C. R. Math. Acad. Sci. Paris*, 349(19-20):1101–1104, 2011. ↑ 56, 87.

[Ebe15] A. Eberle. Reflection couplings and contraction rates for diffusions. *To appear in Probab. Theory Relat. Fields. Available at http://dx.doi.org/10.1007/s00440-015-0673-1)*, 2015. ↑ 56.

[EK86] S. N. Ethier and T. G. Kurtz. *Markov processes*. John Wiley & Sons Inc., New York, 1986. ↑ 72.

[FGC19] Xi F., Yin G., and Zhu C. Regime-switching jump diffusions with non-lipschitz coefficients and countably many switching states: Existence and uniqueness, feller, and strong feller properties, 2019. ↑ 95, 100, 101, 102, 139.

[FR05] G. Fort and G. O. Roberts. Subgeometric ergodicity of strong Markov processes. *Ann. Appl. Probab.*, 15(2):1565–1589, 2005. ↑ 45, 51, 55, 56, 64, 66.

[GAM97] M. K. Gosh, A. Arapostathis, and S. I. Marcus. Ergodic control of switching diffusions. *SIAM J.*, 35(6):1952–1988, 1997. ↑ 18.

[GS79] I. I. Gihman and A. V. Skorohod. *Stochastic Differential Equations*, pages 113–219. Springer, Berlin, 1979. ↑ 5, 36.

[GS82] I. I. Gikhman and A. V. Skorokhod. *Stochastic differential equations and their applications*. Naukova Dumka, Kiev, in Russian, 1982. ↑ 18.

[Hai16] M. Hairer. *Convergence of Markov processes*. Lecture notes, University of Warwick. Available at http://www.hairer.org/notes/Convergence.pdf, 2016. ↑ 49, 52, 53.

[HMS11] M. Hairer, J. C. Mattingly, and M. Scheutzow. Asymptotic coupling and a general form of Harris' theorem with applications to stochastic delay equations. *Probab. Theory Related Fields*, 149(1-2):223–259, 2011. ↑ 56.

[Ish01] Y. Ishikawa. Density estimate in small time for jump processes with singular Lévy measures. *Tohoku Math. J. (2)*, 53(2):183–202, 2001. ↑ 73.

[IW89] N. Ikeda and S. Watanabe. *Stochastic Differential Equations and Diffusion Processes, second edition*. North-Holland Mathematical Library 24. North Holland, 2nd edition, 1989. ↑ 5, 32.

[KC99] Y. Kwon and Lee C. Strong Feller property and irreducibility of diffusions with jumps. *Stochastics*, 67(1-2):147–157, 1999. ↑ 73, 76.

[Kha12] R. Khasminskii. *Stochastic stability of differential equations*. Springer, Heidelberg, second edition, 2012. ↑ 52.

[KK18] V. Knopova and A. M. Kulik. Parametrix construction of the transition probability density of the solution to an SDE driven by α-stable noise. *Ann. Inst. Henri Poincaré Probab. Stat.*, 54(1):100–140, 2018. ↑ 73.

[Kol00] V. N. Kolokoltsov. Symmetric stable laws and stable-like jump-diffusions. *Proc. Lond. Math. Soc. (3)*, 80(3):725–768, 2000. ↑ 73.

[Kol11] V. N. Kolokoltsov. *Markov processes, semigroups and generators*, volume 38. Walter de Gruyter & Co., Berlin, 2011. ↑ 73.

[KS12] V. P. Knopova and R. L. Schilling. Transition density estimates for a class of Lévy and Lévy-type processes. *J. Theoret. Probab.*, 25(1):144–170, 2012. ↑ 73.

[KS13] V. P. Knopova and R. L. Schilling. A note on the existence of transition probability densities of Lévy processes. *Forum Math.*, 25(1):125–149, 2013. ↑ 73.

[Kul09] A. M. Kulik. Exponential ergodicity of the solutions to SDE's with a jump noise. *Stochastic Process. Appl.*, 119(2):602–632, 2009. ↑ 53.

[Kul15] A. M. Kulik. *Introduction to ergodic rates for Markov chains and processes*. Potsdam University Press, Potsdam, 2015. ↑ 52, 53, 56, 66.

[Kul18] A. M. Kulik. *Ergodic behavior of Markov processes*. De Gruyter, Berlin, 2018. ↑ 52, 53, 56, 66.

[KZ20] K. Kunwai and C. Zhu. On Feller and Strong Feller Properties and Irreducibility of Regime-Switching Jump Diffusion Processes with Countable Regimes. *arXiv e-prints*, page arXiv:2002.01422, February 2020. ↑ 94, 95, 96, 102, 139.

[LP17] G. Last and M. Penrose. *Lectures on the Poisson Process*. Institute of Mathematical Stat. Cambridge University Press, 2017. ↑ 5.

[LS21] P. Lazić and N. Sandrić. On sub-geometric ergodicity of diffusion processes. *Bernoulli*, 27(1):348–380, 2021. ↑ 101, 102, 137, 138.

[LW16] D. Luo and J. Wang. Exponential convergence in L^p-Wasserstein distance for diffusion processes without uniformly dissipative drift. *Math. Nachr.*, 289(14-15):1909–1926, 2016. ↑ 56.

[LX22] J. Li and F. Xi. Exponential ergodicity for regime-switching diffusion processes in total variation norm. *Discrete and Continuous Dynamical Systems - B*, 0:–, 2022. ↑ 96.

[Maj16] M. B. Majka. A note on existence of global solutions and invariant measures for jump SDEs with locally one-sided lipschitz drift. *Preprint. Available at https://arxiv.org/pdf/1612.03824.pdf*, 2016. ↑ 35, 88.

[Maj17] M. B. Majka. Coupling and exponential ergodicity for stochastic differential equations driven by Lévy processes. *Stochastic Process. Appl.*, 127(12):4083–4125, 2017. ↑ 56.

[Man68] P. Mandl. *Analytical treatment of one-dimensional Markov processes*. Academia Publishing House of the Czechoslovak Academy of Sciences, Prague; Springer-Verlag New York Inc., New York, 1968. ↑ 66.

[Mas07] H. Masuda. Ergodicity and exponential β-mixing bounds for multidimensional diffusions with jumps. *Stochastic Process. Appl.*, 117(1):35–56, 2007. ↑ 73, 76, 90.

[Mas09] H. Masuda. Erratum to: "Ergodicity and exponential β-mixing bound for multidimensional diffusions with jumps" [Stochastic Process. Appl. 117 (2007) 35–56]. *Stochastic Process. Appl.*, 119(2):676–678, 2009. ↑ 73.

[MT93a] S. P. Meyn and R. L. Tweedie. Stability of Markovian processes. II. Continuous-time processes and sampled chains. *Adv. in Appl. Probab.*, 25(3):487–517, 1993. ↑ 5, 45, 48, 51, 53, 55, 70, 85, 118, 133.

[MT93b] S. P. Meyn and R. L. Tweedie. Stability of Markovian processes. III. Foster-Lyapunov criteria for continuous-time processes. *Adv. in Appl. Probab.*, 25(3):518–548, 1993. ↑ 5, 39, 55, 65, 70, 86, 104, 108, 118, 134.

[MT09] S. Meyn and R. L. Tweedie. *Markov Chains and Stochastic Stability*. Cambridge University Press, Cambridge, second edition, 2009. ↑ 51.

[MY06] X. Mao and C. Yuan. Stochastic differential equation with Markovian switching, aug 2006. ↑ 17, 95, 96.

[Pic96] J. Picard. On the existence of smooth densities for jump processes. *Probab. Theory Related Fields*, 105(4):481–511, 1996. ↑ 73.

[Pic10] J. Picard. Erratum to: On the existence of smooth densities for jump processes. *Probab. Theory Related Fields*, 147(3-4):711–713, 2010. ↑ 73.

[PS92] R. Pinsky and M. Scheutzow. Some remarks and examples concerning the transience and recurrence of random diffusions. *Annales de l'I.H.P. Probabilités et statistiques*, 28(4):519–536, 1992. ↑ 94.

[PS16] G. Pang and N. Sandrić. Ergodicity and fluctuations of a fluid particle driven by diffusions with jumps. *Commun. Math. Sci.*, 2(14):327–362, 2016. ↑ 73.

[RW00] L. C. G. Rogers and D. Williams. *Diffusions, Markov processes, and martingales. Vol. 2*. Cambridge Mathematical Library. Cambridge University Press, Cambridge, 2000. ↑ 36, 57, 65.

[San16a] N. Sandrić. Ergodicity of Lévy-type processes. *ESAIM-Prob.-Stat.*, 20:154–177, 2016. ↑ 56, 66, 77.

[San16b] N. Sandrić. Long-time behavior for a class of Feller processes. *Trans. Am. Math. Soc.*, 3(368):1871–1910, 2016. ↑ 73.

[San17] N. Sandrić. A note on the Birkhoff ergodic theorem. *Results Math.*, 72(1-2):715–730, 2017. ↑ 42.

[Sat13] K. Sato. *Lévy Processes and Infinitely Divisible Distributions*. Cambridge University Press, 2013. ↑ 22.

[Sha13] F. Shao, J.and Xi. Strong ergodicity of the regime-switching diffusion processes. *Stochastic Process. Appl.*, 123(11):3903–3918, 2013. ↑ 96.

[Sha14] J. Shao. Ergodicity of regime-switching diffusions in wasserstein distances, 2014. ↑ 96.

[Sha15] J. Shao. Criteria for transience and recurrence of regime-switching diffusion processes. *Electron. J. Probab.*, 20, 2015. ↑ 118.

[SS10] R. L. Schilling and A. Schnurr. The symbol associated with the solution of a stochastic differential equation. *Electron. J. Probab.*, 15:1369–1393, 2010. ↑ 75, 76.

[SSV12] R. L. Schilling, R. Song, and Z. Vondraček. *Bernstein functions*. Walter de Gruyter & Co., Berlin, 2012. ↑ 77.

[ST97] O. Stramer and R. L. Tweedie. Existence and stability of weak solutions to stochastic differential equations with non-smooth coefficients. *Statist. Sinica*, 7(3):577–593, 1997. ↑ 58, 109.

[Str75] D. W. Stroock. Diffusion processes associated with Lévy generators. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 32(3):209–244, 1975. ↑ 72, 73.

[SV06] D. W. Stroock and S. R. S. Varadhan. *Multidimensional diffusion processes*. Springer-Verlag, Berlin, 2006. Reprint of the 1997 edition. ↑ 5, 36.

[Twe94] R. L. Tweedie. Topological conditions enabling use of Harris methods in discrete and continuous time. *Acta Appl. Math.*, 34(1-2):175–188, 1994. ↑ 45, 47, 64, 70, 73, 75, 117, 118.

[Ver97] A. Yu. Veretennikov. On polynomial mixing bounds for stochastic differential equations. *Stochastic Process. Appl.*, 70, 1997. ↑ 52, 56, 66.

[Ver99] A. Yu. Veretennikov. On polynomial mixing and the rate of convergence for stochastic differential and difference equations. *Teor. Veroyatnost. i Primenen.*, 44(2):312–327, 1999. ↑ 52, 56, 66.

[Vil09] C. Villani. *Optimal transport*. Springer-Verlag, Berlin, 2009. ↑ 42, 43, 91.

[vRS05] M. von Renesse and K. Sturm. Transport inequalities, gradient estimates, entropy, and Ricci curvature. *Comm. Pure Appl. Math.*, 58(7):923–940, 2005. ↑ 56, 83, 84, 87.

[Wan08] J. Wang. Criteria for ergodicity of Lévy type operators in dimension one. *Stochastic Process. Appl.*, 118(10):1909–1928, 2008. ↑ 56, 66, 75.

[Wan16] J. Wang. L^p-Wasserstein distance for stochastic differential equations driven by Lévy processes. *Bernoulli*, 22(3):1598–1616, 2016. ↑ 56.

[YZ10] G. Yin and C. Zhu. *Hybrid switching diffusions*, volume 63. Springer, New York, 2010. ↑ 116.

[Zit15] G. Zitkovic. Semimartingales, 2015. ↑ 5.

Curriculum Vitae

Petra Lazić was born on October 31st, 1991 in Zagreb, were she attended the elementary school Jabukovac and the 5th Gymnasium. After graduating from gymnasium, she enrolled in the Bachelor programme in Mathematics in 2010 at the Faculty of Science in Zagreb. Her third year, via the Erasmus exchange programme for students, she spent studying at the Uppsala University in Sweden. She got her Bachelor's degree in 2013. In the same year she started the Master programme in Financial and business mathematics at the University of Zagreb where she graduated in 2015. The topic of her master's thesis was "Brownian motion and Hausdorff dimension" and it was done under the supervision of Dr. Ante Mimica. During her studies, she received several prizes and awards.

After graduation, she started the Postgraduate program in Mathematics at the same university. During her studies, she agreed with professor Dr. Nikola Sandrić to be her supervisor for the doctoral dissertation. They started to work in the area of stochastic stability and decided that the topic of her thesis will be ergodicity of diffusion process.

At the same time, she applied for the position of a teaching assistant at the Department of mathematics and started to work there in April 2016. Since then, she was holding exercise classes for numerous subjects, including Measure and integration, Statistics lab 1 and 2, Statistics, Financial lab, Elementary mathematics, and so on.

As a doctoral student, she has visited a number of summer schools and conferences in the area of probability (48$^{\text{th}}$ Probability Summer School Saint Flour, France; Bielefeld Stochastic Summer, Germany, 2018; 6$^{\text{th}}$ Workshop on Fractional Calculus, Probability and Non-local Operators, Bilbao, Spain, 2018; Korean-Croatian Summer Probability Camp, Zagreb) and has given several talks on the topic she was working on (at the Seminar for stochastics and analysis at TU Dresden, Lévy summer school in Athens, Faculty of Science PhD Student Symposium, and so on). She is an active member and secretary of the departmental seminar for the probability theory.

The work on her doctoral dissertation resulted in two papers, one of which is published so far:

(1) P. Lazić, N. Sandrić, *On Sub-geometric ergodicity of diffusion processes*, Bernoulli (1350-7265) 27, 1; 348-380, 2021.