Implications of Markov stability theory for nonparametric statistics, Markov additive fluctuations and data-driven stochastic control

Inauguraldissertation
zur Erlangung des akademischen Grades
eines Doktors der Naturwissenschaften
der Universität Mannheim

vorgelegt von

Lukas Trottner
aus Wiesbaden

Mannheim, 2021

Dekan:	Dr. Bernd Lübcke, Universität Mannheim
Referent:	Prof. Dr. Leif Döring, Universität Mannheim
Korreferentin:	Prof. Dr. Claudia Strauch, Aarhus University
Korreferent:	Prof. Dr. Mark Podolskij, Université du Luxembourg

Tag der mündlichen Prüfung: 24. September 2021

Abstract

The central topic of this thesis is the influence of stability properties of continuous time Markov processes on their nonparametric statistical analysis. In particular sup-norm adaptive invariant density estimation under assumptions on the ergodic behavior of the process is investigated and consequently applied to jump diffusions with Lévy-driven jump part. Furthermore, the findings are used to demonstrate how statistical procedures for Markov processes can be implemented for the development of efficient data-driven strategies for stochastic optimal control problems associated to both continuous diffusion processes and Lévy processes. As one of the main theoretical tools in this regard, we give a detailed analysis of stability properties of overshoots associated to Markov additive processes. This allows incorporating fluctuation theory for Markov additive processes and Lévy processes into our general statistical framework, which is essential for the data-driven Lévy control strategy. Moreover, the overshoot analysis guides us naturally into extending some well-known fluctuation results for Lévy processes to the more general case of Markov additive processes and—making use of the one-to-one correspondence between Markov additive processes and real self-similar Markov processes through the Lamperti–Kiu transform—gives us the right tool to analyze the mixing behavior of self-similar Markov processes sampled at first hitting times.

Zusammenfassung

Zentrales Thema dieser Arbeit ist der Einfluss von Stabilitätseigenschaften zeitstetiger Markovprozesse auf ihre nichtparametrische statistische Analyse. Insbesondere betrachten wir die adaptive Schätzung der invarianten Verteilung bezüglich des sup-Norm Risikos unter Annahmen an das ergodische Verhalten des Prozesses und wenden die Resultate auf Diffusionen mit Lévy-gesteuerter Sprungkomponente an. Wir nutzen unsere Ergebnisse, um datengesteuerte statistische Ansätze für Lösungsstrategien stochastischer optimaler Kontrollprobleme sowohl für stetige Diffusionen als auch für Lévyprozesse zu entwickeln. Als eine der fundamentalen theoretischen Entwicklungen in dieser Hinsicht, geben wir eine detaillierte Analyse des Stabilitätsverhaltens von Overshoots von Markov additiven Prozessen, was uns gestattet Fluktuationstheorie für Lévyprozesse und Markov additive Prozesse in unseren allgemeinen statistischen Rahmen einzubetten. Zudem motiviert uns die Overshoot-Analyse dazu, einige zentrale Fluktuationsresultate für Lévyprozesse auf den allgemeineren Fall Markov additiver Prozesse zu erweitern. Schließlich erlauben es unsere Overshoot Resultate ebenso, das Mixingverhalten reeller selbstähnlicher Markovprozesse, die an Ersteintrittszeiten ausgewertet werden, zu analysieren. Dazu nutzen wir die bijektive Beziehung zwischen dieser Prozessklasse mit Markov additiven Prozessen, die durch die Lamperti–Kiu Transformation beschrieben wird.

Danksagung

Hier möchte ich die Gelegenheit nutzen, all jenen zu danken, die fachlich oder emotional dazu beigetragen haben meine Promotionszeit zu einem Lebensabschnitt werden zu lassen, an den ich mich immer mit Freude und sicherlich auch ein wenig Wehmut zurückerinnern werde.

An erste Stelle möchte ich meinem Betreuer Leif Döring danken. Er hat mich in die Wissenschaftswelt eingeführt und mich mit seiner Passion für die Wahrscheinlichkeistheorie sowie klugen und ertragreichen Ideen immer wieder dazu motiviert, tiefer zu graben, um so mehr aus meinen Ansätzen herauszuholen. Gleichzeitig hat er mir das stets spürbare Vertrauen entgegengebracht, eigene Pfade in der Statistik zu erkunden und mir so die Möglichkeit gegeben, mich in vielfältige Themengebiete einzuarbeiten und in Zusammenhang zu bringen. Die Summe dieser Einzelteile ist das Ergebnis dieser Arbeit. Unabhängig von diesen fachlichen Punkten, können wohl die wenigsten Doktoranden von sich behaupten, schon mal einen Nachmittag im Hochsommer gemeinsam mit ihrem Doktorvater knietief im Müll verbracht zu haben, was eine besondere zwischenmenschliche Erfahrung war, die uns immer verbinden wird.

Genauso möchte ich meine tiefe Dankbarkeit gegenüber Claudia Strauch ausdrücken, die mich von der reinen Wahrscheinlichkeitstheorie in die mathematische Statistik entführt hat und dabei stets voller Hingabe und Euphorie ihr Wissen und Gespür für die größeren Zusammenhänge mit mir geteilt hat. Ohne ihren Einfluss und ihre Mitarbeit wären große Teile der Dissertation in der jetzigen Form nicht entstanden. Zudem war sie über das Fachliche hinaus stets eine große menschliche Stütze für mich, die mir über frustrierende Phasen der Promotion—gerade in Zeiten der Pandemie—hinweggeholfen hat.

In diesem Zusammenhang möchte ich auch Sören Christensen danken, mit dessen zusätzlicher Unterstützung das letzte Kapitel der Dissertation entstanden ist und von dessen breitem Wissen und Intuition für kluge Lösungswege ich sehr profitiert habe. In den letzten Monaten vor Abgabe der Dissertation konnte ich mich zudem glücklich schätzen wöchentliche Diskussionen mit Alex Watson über MAPs, Wiener–Hopf Zerlegungen und Philanthropen haben zu dürfen.

Das tägliche Büroleben haben mir meine Freunde, Mitdoktoranden und Bürokollegen Niklas, Philipp und Simon enorm verschönert. Auch wenn es mir nie zu meiner vollen Zufriedenheit gelungen ist sie von den Vorzügen der Mensa-Kaltschale an heißen Sommertagen zu überzeugen und meine Produktivität nicht immer von ihrer Anwesenheit profitiert hat, bin ich ihnen sehr dankbar für die freundschaftliche und offene Atmosphäre. Es war sehr hilfreich für mich, in Anbetracht der manchmal aufkommenden mathematischen Frustrationen, zumindest Erfolge im Bürogolf feiern zu dürfen oder die Uni zu Gunsten anderer Genüsse gemeinsam hinter sich zu lassen. Auch André möchte ich nicht übergehen, der zwar leider das Pech hatte, erst zu Corona-Zeiten dazuzustoßen, mit dem ich aber seitdem sehr gerne und selbstverständlich mit Abstand das Büro teilte. Zudem möchte ich meine (ehemaligen) Mitdoktoranden bei Leif, Sam und Philip, dankend erwähnen, mit denen ich nicht nur über Markovprozesse reden konnte, sondern auch den Leistungssport Barbillard in Bath kennen- und schätzen gelernt habe. Genauso danke ich allen Mitgliedern des Graduiertenkollegs *RTG 1953: Statistical modeling of complex systems* für das professionelle und stimulierende wissenschaftliche Umfeld und insbesondere Sanja Juric für ihr herausragendes Management aller Verwaltungsangelegenheiten, die mir in den letzten drei Jahren begegnet sind. Besonders will ich noch Niklas hervorheben. Wir sind

zwar auch mathematisch durch unsere Zusammenarbeit, die entscheidend für das dritte Kapitel der Dissertation war, verbunden, aber vielleicht noch bedeutender dadurch, dass wir ein Zimmer teilten während er gegen die Salmonellen auf Samos kämpfte.

Im gleichen Sommer der Samos-Reise zur Lévy-Konferenz im ersten Jahr meiner Promotion, hatte ich die große Freude mit meinem besten Freund Anton einen grandiosen Urlaub in Vietnam zu verbringen, der mir trotz der Nahtoderfahrungen auf den Busfahrten sehr gut tat. Ich schätze mich glücklich auf ihn seit dem ersten Semester in Mannheim als Freund zählen zu können. Gleiches gilt für Urs, Patrick, Maxi-P., Jonas, Sven, Alex Z., Alex S., Yang, Jin, Drea und viele andere, die meine Studienzeit als Freunde begleitet und bereichert haben.

Worte können hier meine Dankbarkeit gegenüber Jovi kaum zum Ausdruck bringen, die in den letzten Monaten auch meine schlechten Launen ertragen und mich zu jedem Zeitpunkt unterstützt, motiviert und umsorgt—und manchmal bestimmt auch ein klein wenig gehasst hat. Ohne dich wäre das Leben nicht das gleiche. Auch Iwona und Zenon möchte ich hier meinen Dank aussprechen—nicht nur dafür, dass sie meine handwerklichen Unzulänglichkeiten während des Umzugs im letzten Jahr mehr als ausgeglichen haben, sondern auch dafür, dass sie mir durch ihre großzügige und herzliche Art das Gefühl geben bei ihnen zu Hause zu sein.

Zu guter letzt möchte ich zumindest versuchen meine unendliche Dankbarkeit gegenüber meiner Familie zum Ausdruck zu bringen. Mein großer Bruder Fabian war mir in vielen Dingen im Leben bis heute ein Vorbild und ist mir trotz der großen räumlichen Distanz so wichtig wie kaum ein anderer Mensch in meinem Leben. Meine Eltern Ute und Frank haben mich mein gesamtes Leben zu jedem Zeitpunkt selbstlos und großzügig unterstützt und mich nicht nur gut ernährt, sondern mir Werte und Liebe mitgegeben, die mich im Leben tragen. Ich weiß, dass ich immer auf sie zählen kann. Ihnen ist diese Arbeit gewidmet.

Contents

List of Figures

Introduction

1

Along with semimartingales, Markov processes are certainly the most studied and best understood class of stochastic processes in continuous time. With their flexibility and analytical tractability this makes them an accessible and frequently used tool for the modeling of random phenomena in many disciplines whenever the central assumption of memorylessness is considered adequate. Quite naturally, this raises the interest in efficient statistical estimation of Markov processes based on different observation schemes, with continuous observations on the one hand and discrete low- and high-frequency sampling schemes on the other hand being the most prominent in the literature. Any continuos time Markov process that is sampled at equally spaced time intervals (low-frequency) is a discrete time Markov chain and in applications the availability of discrete data is more realistic than a continuous record of the process. Consequently, the majority of the statistical literature on Markov processes focusses on the discrete perspective and therefore does not differentiate between estimation of Markov processes and Markov chains.

However, from a purely mathematical point of view it is of fundamental interest to investigate whether the loss of information imposed by incomplete sampled data also results in a loss of performance of statistical procedures. In fact, it has been observed in the literature on nonparametric statistics for continuous diffusion processes that under suitable structural assumptions, the invariant density—quantifying the distributional equilibrium of the process—can be estimated at a faster rate wrt. different risk measures in any dimension compared to the classical nonparametric rate known from i.i.d. data in discrete time [59, 106, 158]. This raises the question whether such phenomena can also be observed for more general classes of stochastic processes. An essential contribution in this regard was made in [39] and [34], where general criteria on the transition densities of a Markov process were formulated that guarantee a dimension independent parametric L^2 rate of convergence of nonparametric invariant density estimators. Such independence of the dimension, however, cannot be observed for the aforementioned case of continuous diffusion processes, for which the minimax estimator performance decreases substantially with increasing dimension of the process, which is generally referred to as the *curse of dimensionality*.

One main goal of this thesis is therefore to formulate a general framework for Markovian statistics that inhibits the nonparametric estimation performance of continuous diffusion processes but is applicable to a much wider range of stochastic processes, especially allowing jump structures. As a central ingredient for this framework we work under assumptions on the ergodic and mixing behavior of the process. Heuristically, it is clear that a fast convergence of the stochastic process to equilibrium increases the performance of estimators based on data that is collected with increasing time horizon. Technically, the notion of speed of convergence to equilibrium is forged into the frame of convergence of probability measures. Let $(P_t)_{t\geq 0}$ be the semigroup associated to a given *nice* $\mathcal{X}$-valued Markov process $\boldsymbol{X}$ that describes the transitional behavior of the process. Assuming that $\boldsymbol{X}$ possesses a unique invariant distribution μ—that is, the marginal distributions of $\boldsymbol{X}$ started in μ do not change over time—we may express convergence of $\boldsymbol{X}$ to μ through

$$d(P_t(x,\cdot),\mu) \underset{t\to\infty}{\longrightarrow} 0, \quad x \in \mathcal{X},$$

where $d(\cdot,\cdot)$ is a suitable metric on the space of probability distributions. Popular choices for d are the Prokhorov metric that metrizes the topology of weak convergence, the metric induced by the total variation norm $\|\cdot\|_{\mathrm{TV}}$ of signed measures or the Wasserstein distance, which is associated to an optimal transport problem.

For Markov processes, the most prominent choice is certainly the total variation metric $d_{\mathrm{TV}}(\mu,\nu) = \|\mu-\nu\|_{\mathrm{TV}}$ (or, alternatively, some metric induced by a generalized operator norm) since a large branch of literature, commonly referred to as *stability theory of Markov processes*, deals with establishing tractable conditions on the characteristics of $\boldsymbol{X}$ that allow finding some rate function $\Xi\colon \mathbb{R}_+ \longrightarrow \mathbb{R}_+$ decreasing to 0 such that

$$\|P_t(x,\cdot) - \mu\|_{\mathrm{TV}} \leq V(x)\Xi(t), \quad t \geq 0, x \in \mathcal{X},$$

where V is some penalty function allowing non-uniformity of convergence wrt. the initial distribution of the process. These conditions are stated in terms of nice characterizing objects of $\boldsymbol{X}$ such as its semigroup, its generator or its resolvent and it is a challenging mathematical task to translate these general results into concrete assumptions for particular Markov models. Among others, such precise convergence statements can be used to describe the mixing behavior of the process, where—in general terms—the process is said to be mixing if the σ-algebras $\sigma(X_s, s \leq t)$ and $\sigma(X_s, s \geq t+h)$ encoding the temporal t-past and $t+h$-future of the process, respectively, become asymptotically independent as $h \longrightarrow \infty$. For our purposes, the concrete notion of β-mixing (or absolute regularity) will be central throughout the whole thesis thanks to its direct association to total variation convergence.

For discrete time data it is well-known that certain mixing requirements allow to reproduce nonparametric rates associated to i.i.d. observations, since coupling results essentially allow to split the risk of an estimator into an i.i.d. contribution and a part that can be controlled through the coupling error expressed in terms of the mixing coefficient. We will demonstrate how mixing and ergodic requirements in continuous time together with an on-diagonal heat kernel estimate of the transition semigroup—essentially controlling the speed at which the marginal distributions of the process approach a singular Dirac-distribution as $t \downarrow 0$—provide the right mix of long- and short time control on the transitions of the process to obtain the minimax rates observed for continuous diffusion processes. In particular, we focus on tight variance bounds as well as uniform moment bounds and deviation inequalities for additive functionals of β-mixing Markov processes as the main ingredient for sup-norm adaptive estimation of the invariant density. As a consequence, without much additional effort we are able to include jump diffusions with Lévy-driven jump part in arbtirary dimension into our general modeling approach and obtain novel results on sup-norm adaptive invariant density estimation of such processes.

Apart from the substantial increase in technical difficulty of proving such (adaptive) estimation rates wrt. the sup-norm risk compared to the more popular (but certainly not more useful) L^2 risk of estimators, this general approach equips us with the right tools to develop data-driven substitutes for theoretically known solutions of stochastic optimal control problems in presence of uncertainty on the dynamics of the underlying stochastic process. This is dealt with as another main part of the thesis. With the exception of the recent article [50], which deals with data-driven solutions to impulse control problems for diffusion processes, so far there are no comparable studies of data-driven controls for continuous time Markov processes, although the practical consequences of such results are substantial. One type of problem we deal with is a singular ergodic control problem for continuous scalar diffusion processes, whose theoretical solution is given by a two-sided reflection strategy. Another problem we consider is a stochastic optimal control problem associated to a Lévy process diverging to $+\infty$, which is formulated as an impulse type control problem, but whose optimal solution is given by a strategy approximating reflection at a certain boundary, thus being essentially of singular type as well.

For such singular strategies, starting with a continuous observation scheme is inherently reasonable, since the execution of the optimal strategy requires not only continuously tracking but also continuously controlling the process. Any practical implementation of the strategy based on discretely observed data and finitely many interventions on any time interval must therefore be treated as an approximation to the optimal strategy. Likewise, the performance of an approximating discrete data-driven strategy should be evaluated relative to the performance of our continuous data-driven controls, similarly to numerical approximations of analytic objects or stochastic simulation algorithms.

From a technical point of view, our data-driven approximation of the optimal reflection boundaries of the continuous diffusion process—given data $(X_t)_{0\leq t\leq T}$—can be elegantly included into our general nonparametric statistical framework since the optimal solution can be expressed in terms of the invariant density of the diffusion and we work under natural coefficient assumptions that ensure exponential convergence to equilibrium and a sufficient control on the small-time transitions either through a heat-kernel bound or, alternatively, through local time arguments. Thus, our estimator arises directly from nonparametric kernel density estimation techniques. For this specific control problem however, an additional difficulty arises from the more natural scenario when we are not given a data set $(X_t)_{0\leq t\leq T}$ from the beginning but must collect data simultaneously to estimation and application of the reflection boundaries to minimize the costs in the ergodic problem formulation. This way, we face an exploration/exploitation type dilemma. Controlling the process reduces the costs but at the same time we cannot collect data on the behavior of the diffusion away from the estimated boundaries, which is essential for convergence of our estimation procedure. Conversely, not controlling the process in favor of data collection to improve the approximation of the optimal solution results in potentially high costs from the lack of intervention. Consequently, these effects must be balanced. We propose a splitting into exploration and exploitation periods separated by random hitting times of the diffusion, where the time spent in exploration and exploitation periods on average is balanced such that the algorithm converges at a rate, which includes an additional but unavoidable loss compared to the optimal rate without exploration/exploitation problem due to the balancing process.

Such exploration/exploitation type problem does not occur for Lévy processes since spatial homogeneity of the process—or equivalently, translation invariance of the semigroup—allows the controller to recover the path of a process from the controlled path that has the same law as the uncontrolled Lévy process. The significant challenge in this scenario comes from the fact that the optimal reflection boundary is characterized as the maximizer of the generator functional

$$f(x) = \mathcal{A}_H \gamma(x), \quad x \in \mathbb{R},$$

where γ is the reward function underlying the control problem and $\mathcal{A}_H$ is the extended generator of the ascending ladder height process $\boldsymbol{H}$, which is one of the central objects of Lévy fluctuation theory. This process is a subordinator obtained from time changing the Lévy process $\boldsymbol{X}$ by inverse local time at the supremum L. However, in general L cannot be recovered even from observations $(X_t)_{0\leq t\leq T}$ and $\boldsymbol{H}$ as an increasing Lévy process is not ergodic in time such that estimation of $\boldsymbol{H}$ cannot be directly treated within our proposed general statistical framework. At this point the third main part of this thesis comes into play, which treats convergence analysis of overshoots of Lévy processes—and even more generally of Markov additive processes.

Our main idea is to rewrite the generator functional f in terms of an integral wrt. the stationary limiting distribution μ of the overshoots $(\mathcal{O}_x)_{x\geq 0}$ associated to $\boldsymbol{X}$ that are defined by

$$\mathcal{O}_x := X_{T_x} - x, \quad x \geq 0,$$

where $T_x = \inf\{t \geq 0 : X_t > x\}$ is the first passage time of x. This stationarity is not of temporal but spatial nature since convergence has been observed in [24] in distribution as the level x tends to ∞. Our goal is therefore to establish conditions on $\boldsymbol{X}$ that improve classical distributional convergence of overshoots to explicit rates of convergence in total variation and the (exponential) β-mixing property of the Markov process $(\mathcal{O}_x)_{x\geq 0}$, in order to have access to the uniform moment bounds from the general statistical framework that can provide sup-norm estimation results for the generator functional f. Another challenge that we must address then consists of transforming the spatial estimator into a temporal one to obtain intepretable convergence rates for given observations $(X_t)_{0\leq t\leq T}$.

As mentioned above, we consider the overshoot convergence problem more generally for Markov additive processes (MAPs). Loosely speaking, these are Lévy processes living in a Markovian random environment. Apart from fundamental reasons, our motivation to widen the view to MAPs is twofold. First, there is a one-to-one relation between self similar Markov processes—including strictly α-stable Lévy processes—and MAPs with bivariate background Markov chain that is captured by the Lamperti–Kiu transformation. Since the overshoot process is invariant under continuous time changes of the underlying MAP, this establishes a direct connection between Markov additive overshoots and self-similar Markov processes sampled at first hitting times. The β-mixing results that we establish for Markov additive overshoots therefore translate to mixing rates for self-similar Markov processes sampled at first hitting times that are of independent interest. Second, the assumptions that we shall impose on the given MAP to prove our ergodicity results with general techniques obtained from stability theory of Markov processes, are stated in terms of the ascending ladder height MAP—which in the special case of Lévy processes is identical to $\boldsymbol{H}$ introduced above. To make these assumptions transparent in terms of assumptions on the observable parent MAP $\boldsymbol{X}$, we prove fluctuation identities that generalize the by now classical *équations amicales inversés* for Lévy processes obtained in Vigon [172]. This contributes significantly to the understanding of fluctuations of MAPs and has potential applications that go beyond overshoot convergence considered in this thesis.

1.1 Outline

Chapter 2: Stability of Markov processes

We introduce the most important definitions and objects in the general theory of Markov processes and give a condensed overview of stability theory of Markov processes to establish the most important notions and results needed for the following chapters. Moreover, we give a novel criterion for invariant measures in terms of a suitable form of resolvent convergence and characterize exponential β-mixing of a Markov process by local uniform transition density convergence at exponential speed.

Chapter 3: Markovian statistics under mixing assumptions

In this chapter we introduce a general framework for nonparametric Markovian statistics based on verifiable assumptions on the short-and long time transitional behavior of a given Borel right Markov process. We demonstrate how exponential β-mixing can be considered as the key ingredient for sup-norm adaptive invariant density estimation through the use of general moment bounds for empirical processes associated to additive functionals of exponentially β-mixing Markov processes. As particular examples, we show that Lévy-driven Ornstein–Uhlenbeck processes as well as jump diffusions with Lévy driven jump part and bounded Lipschitz coefficients can be estimated optimally within our framework.

Chapter 4: Stability of overshoots of Markov additive processes

The asymptotic behavior of overshoots of MAPs is analyzed in detail with techniques from general stability theory of Markov processes. Assumptions on the ascending ladder height MAP are introduced that yield exponential and polyonmial ergodicity as well as exponential and polynomial β-mixing of overshoots. Main ingredient for the proofs is the derivation of an explicit formula of the overshoot resolvent that allows us to prove a necessary and sufficient condition for the existence of an (explicit) invariant overshoot measure and to find appropriate Lyapunov-functions for establishing the ergodicity results. By extending Vigon's équations amicales inversés, the conditions on the ascending ladder height MAP are expressed in terms of the characteristics of the parent MAP and the mixing results are translated into mixing rates for real self-similar Markov processes.

Chapter 5: Data-driven control strategies for diffusions and Lévy processes

In this chapter the data-driven solution strategies to ergodic stochastic control problems associated to continuous diffusion processes and Lévy processes are developed, bringing together our work from the previous chapters. As a byproduct of independent interest, a nonasymptotic deviation inequality for Lévy processes with bounded jumps is established.

1.2 Collaborative Work

The results from Section 2.3 and Chapter 3 were obtained in collaboration with Niklas Dexheimer and Prof. Dr. Claudia Strauch from Aarhus University and are available as a preprint [67]. The contents of Chapter 4 and partially of Sections 2.1 and 2.2 are based on joint work with Prof. Dr. Leif Döring from University of Mannheim and can be found in the preprint [71]. Chapter 5 is the result of joint work with Prof. Dr. Sören Christensen from Kiel University and Prof. Dr. Claudia Strauch and is also available as a preprint [51].

STABILITY OF MARKOV PROCESSES

2

THE theory of stability of continuous time Markov processes is the common thread running through the different parts of the thesis. In Section 2.1 we therefore introduce the general terminology and central results from the literature that we shall frequently fall back to in the following chapters. We then proceed in Section 2.2 by establishing two technical results of general character which will be important for our analysis in Chapter 4. In Section 2.3 we investigate the influence of local uniform transition density convergence on the mixing behavior of a Markov process. On the one hand, this is a good warm up for the remainder of the thesis since the discussion demonstrates the power and general techniques of the Meyn and Tweedie approach. On the other hand, the section serves as direct preparation for Chapters 3 and 5, where the transition density convergence assumption plays a central role.

2.1 FUNDAMENTALS

2.1.1 *Basic concepts*

Let us start by formally introducing the most important concepts from stability theory of continuous time Markov processes. Let $\mathcal{X}$ be a topological space and $\mathcal{X}_\vartheta$ be its Alexandrov one-point compactification by some isolated state ϑ. Denote by $\mathcal{B}(\mathcal{X}_\vartheta)$ its associated Borel σ-algebra. We follow the common and convenient convention to extend any function $f\colon \mathcal{X} \to \mathbb{R}$ to $\mathcal{X}_\vartheta$ by setting $f(\vartheta) = 0$. In the same vein, we extend measures μ on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ to $(\mathcal{X}_\vartheta, \mathcal{B}(\mathcal{X}_\vartheta))$ by setting $\mu(\{\vartheta\}) = 0$. We work with the following definition of a Markov process from the standard textbook [31].

DEFINITION 2.1. A sextuple $((X_t)_{t\geq 0}, \Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, (\mathbb{P}^x)_{x\in\mathcal{X}_\vartheta}, (\theta_t)_{t\geq 0})$ is called *Markov process* if the following conditions are satisfied.

(i) for any $x \in \mathcal{X}_\vartheta$, $(X_t)_{t\geq 0}$ is a stochastic process on the probability space $(\Omega, \mathcal{F}, \mathbb{P}^x)$ adapted to the filtration $\mathbb{F} := (\mathcal{F}_t)_{t\geq 0}$;

(ii) for any $B \in \mathcal{B}(\mathcal{X}_\vartheta)$ and $t \geq 0$, $x \mapsto \mathbb{P}^x(X_t \in B)$ is measurable;

(iii) for any $x \in \mathcal{X}_\vartheta$ it holds that $\mathbb{P}^x(X_0 = x) = 1$;

(iv) for any $t \geq 0$, $\theta_t\colon \Omega \to \Omega$ is a measurable function such that for any $s \geq 0$, $X_{t+s} = X_s \circ \theta_t$. The family $(\theta_t)_{t\geq 0}$ is called family of *shift operators* of the Markov process;

(v) for any $f \in \mathcal{B}_b(\mathcal{X}_\vartheta)$, $s, t \geq 0$ and $x \in \mathcal{X}_\vartheta$, it holds that

$$\mathbb{E}^x[f(X_{t+s})|\mathcal{F}_t] = \mathbb{E}^{X_t}[f(X_s)], \quad \mathbb{P}^x\text{-a.s.}; \qquad \text{(Markov property)}$$

(vi) ϑ is an absorbing state called *cemetery state*, i.e., $X_t = \vartheta$ implies $X_s = \vartheta$ for any $s \geq t \geq 0$. The *lifetime* of the Markov process is defined by $\zeta := \inf\{t \geq 0 : X_t = \vartheta\}$ and $\boldsymbol{X}$ is called *non-explosive* if for any $x \in \mathcal{X}$, $\mathbb{P}^x(\zeta = \infty) = 1$.

If there is no room for confusion we will usually abbreviate a Markov process by $\boldsymbol{X} := (X_t)_{t\geq 0}$ or $(\boldsymbol{X}, (\mathbb{P}^x)_{x\in\mathcal{X}_\vartheta})$. Let us also set $X_\infty := \vartheta$. With standard measure theoretic arguments it can be shown that the Markov property is equivalent to the apparently stronger condition that for any $\sigma(X_s, s \geq 0)$-measurable random variable Z and $t \geq 0$, we have

$$\mathbb{E}^x[Z \circ \theta_t | \mathcal{F}_t] = \mathbb{E}^{X_t}[Z], \quad \mathbb{P}^x\text{-a.s.},$$

and if we let $\mathbb{P}^\mu(\cdot) = \int_{\mathcal{X}_\vartheta} \mathbb{P}^x(\cdot)\, \mu(\mathrm{d}x)$ be the probability measure induced by some probability measure μ on $(\mathcal{X}_\vartheta, \mathcal{B}(\mathcal{X}_\vartheta))$ we also have

$$\mathbb{E}^\mu[Z \circ \theta_t | \mathcal{F}_t] = \mathbb{E}^{X_t}[Z], \quad \mathbb{P}^\mu\text{-a.s.},$$

and $X_0 \sim \mu$ under $\mathbb{P}^\mu$, i.e., $\boldsymbol{X}$ is started with initial distribution μ. We let $(P_t)_{t\geq 0}$ be the sub-Markov *transition semigroup* (or *transition function*) of $\boldsymbol{X}$ defined on $\mathcal{B}_b(\mathcal{X}_\vartheta) \cup \mathcal{B}_+(\mathcal{X}_\vartheta)$ via

$$P_t f(x) = \mathbb{E}^x[f(X_t); t < \zeta], \quad t \geq 0, x \in \mathcal{X}_\vartheta, f \in \mathcal{B}_b(\mathcal{X}_\vartheta) \cup \mathcal{B}_+(\mathcal{X}_\vartheta),$$

where the semigroup property $P_{t+s} = P_t \circ P_s$ is a consequence of the Markov property. Note that by our convention to set $f(\vartheta) = 0$ for $f \in \mathcal{B}_b(\mathcal{X}) \cup \mathcal{B}_+(\mathcal{X})$, for such f we have $P_t f(x) = \mathbb{E}^x[f(X_t)]$ since ϑ is absorbing. For a measure μ on $(\mathcal{X}_\vartheta, \mathcal{B}(\mathcal{X}_\vartheta))$ we write $\mu P_t = \int_{\mathcal{X}_\vartheta} P_t(x, \cdot)\, \mu(\mathrm{d}x)$, where $P_t(x, B) = P_t \mathbb{1}_B(x)$, $(x, B) \in \mathcal{X}_\vartheta \times \mathcal{B}(\mathcal{X}_\vartheta)$, is a kernel by definition of a Markov process. If for a measure η on $(\mathcal{X}_\vartheta, \mathcal{B}(\mathcal{X}_\vartheta))$ and $f \in L^1(\mathcal{X}_\vartheta, \eta) \cup \mathcal{B}_+(\mathcal{X}_\vartheta)$ we write $\eta(f) = \int_{\mathcal{X}_\vartheta} f(x)\, \eta(\mathrm{d}x)$, then clearly, for any $f \in \mathcal{B}_b(\mathcal{X}_\vartheta) \cup \mathcal{B}_+(\mathcal{X}_\vartheta)$

$$\mu P_t(f) = \mu(P_t f), \quad t \geq 0,$$

which also equals $\mathbb{E}^\mu[f(X_t)]$ provided that $f(\vartheta) = 0$ and that μ is a probability measure. Finally, let us introduce the *resolvent* $(U_\lambda)_{\lambda>0}$ of a Markov process $\boldsymbol{X}$, which is a family of operators on $\mathcal{B}_b(\mathcal{X}_\vartheta) \cup \mathcal{B}_+(\mathcal{X}_\vartheta)$ defined by

$$U_\lambda f(x) = \mathbb{E}^x\Big[\int_0^\infty \mathrm{e}^{-\lambda t} f(X_t)\, \mathrm{d}t\Big] = \int_0^\infty \mathrm{e}^{-\lambda t}\mathbb{E}^x[f(X_t)]\, \mathrm{d}t, \quad \lambda > 0, x \in \mathcal{X}_\vartheta, f \in \mathcal{B}_b(\mathcal{X}_\vartheta) \cup \mathcal{B}_+(\mathcal{X}_\vartheta),$$

which for $x \in \mathcal{X}$ and $f \in \mathcal{B}_b(\mathcal{X}) \cup \mathcal{B}_+(\mathcal{X})$ can be rewritten as

$$U_\lambda f(x) = \int_0^\infty \mathrm{e}^{-\lambda t} P_t f(x)\, \mathrm{d}t.$$

Similarly, to the transition function of a Markov process, if μ is a measure on $(\mathcal{X}_\vartheta, \mathcal{B}(\mathcal{X}_\vartheta))$, let us define the measure $\mu U_\lambda := \int_{\mathcal{X}} U_\lambda(x, \cdot)\, \mu(\mathrm{d}x)$, where $U_\lambda(x, B) = U_\lambda \mathbb{1}_B(x)$ for $(x, B) \in \mathcal{X}_\vartheta \times \mathcal{B}(\mathcal{X}_\vartheta)$. The λ-resolvent U_λ can be interpreted as the operator induced by the *potential* of the Markov process $\boldsymbol{X}$ killed at an independent exponential time with rate $1/\lambda$, where for $x \in \mathcal{X}_\vartheta$, the potential $U(x, \cdot)$ defined by

$$U(x, B) := \mathbb{E}^x\Big[\int_0^\infty \mathbb{1}_{\{X_t \in B\}}\, \mathrm{d}t\Big] = \int_0^\infty \mathbb{P}^x(X_t \in B)\, \mathrm{d}t, \quad B \in \mathcal{B}(\mathcal{X}_\vartheta),$$

is the expected sojourn time of $\boldsymbol{X}$ in B when started in x.

Building on this basic terminology let us now come to more specific classes of Markov processes with useful properties. Arguably the most important property that we shall need is the *strong* Markov property. We say that a Markov process $(\boldsymbol{X}, (\mathbb{P}^x)_{x\in\mathcal{X}_\vartheta})$ with underlying filtration $\mathbb{F}$ is strong Markov, if for any $\mathbb{F}$-stopping time T and $f \in \mathcal{B}_b(\mathcal{X}_\vartheta)$ we have

$$\forall x \in \mathcal{X}_\vartheta : \ \mathbb{E}^x[f(X_{T+t})|\mathcal{F}_T]\mathbb{1}_{\{T<\infty\}} = \mathbb{E}^{X_T}[f(X_t)]\mathbb{1}_{\{T<\infty\}}, \quad \mathbb{P}^x\text{-a.s.},$$

which is equivalent to

$$\forall x \in \mathcal{X} : \ \mathbb{E}^x[f(X_{T+t})|\mathcal{F}_T] = \mathbb{E}^{X_T}[f(X_t)], \quad \mathbb{P}^x\text{-a.s.},$$

by our conventions $X_\infty = \vartheta$ and the property that ϑ is absorbing. Furthermore, we say that $\boldsymbol{X}$ is *quasi-left-continuous on* $[0, \zeta)$, if for a sequence of $\mathbb{F}$-stopping times $(T_n)_{n\in\mathbb{N}}$ increasing almost surely[1] to an $\mathbb{F}$-stopping time T, then it holds that

$$\lim_{n\to\infty} X_{T_n}\mathbb{1}_{\{T<\zeta\}} = X_T\mathbb{1}_{\{T<\zeta\}}, \quad \mathbb{P}^*\text{-a.s.}.$$

If the above convergence holds on $\{T < \infty\}$ instead of $\{T < \zeta\}$, we say that $\boldsymbol{X}$ is quasi-left-continuous. A filtration $\mathbb{F}$ is said to be right-continuous if for any $t \geq 0$

$$\mathcal{F}_t = \mathcal{F}_{t+} := \bigcap_{\varepsilon>0} \mathcal{F}_{t+\varepsilon}.$$

A set $N \subset \Omega$ is called $\mathbb{P}^*$-negligible for the family of probability measures $(\mathbb{P}^x)_{x\in\mathcal{X}_\vartheta}$ on the measurable space $(\Omega, \mathcal{F})$ if there exists some measurable set $\Lambda \in \mathcal{F}$ with $\mathbb{P}^x(\Lambda) = 0$ for all $x \in \mathcal{X}_\vartheta$ such that $N \subset \Lambda$. The family of filtered probability spaces $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P}^x)_{x\in\mathcal{X}_\vartheta}$ underlying a Markov process—often referred to as the *stochastic base*—is said to be *complete* if any $\mathbb{P}^*$-negligible set is contained in $\mathcal{F}_0$. Assuming completeness of the stochastic base is without loss of generality for a given Markov process $\boldsymbol{X}$, since by enlarging the natural stochastic base associated to the natural σ-algebra $\mathbb{F}^0 = (\mathcal{F}^0_t)_{t\geq 0}$ by a procedure called *augmentation* in order to obtain a complete stochastic base, $\boldsymbol{X}$ remains (strong) Markov with respect to the augmented stochastic base. See [31, Section I.5] and [52, Section 2.3] for details. If the underlying filtration $\mathbb{F}$ is right continuous and the stochastic base associated to $\mathbb{F}$ is complete, we say that $\boldsymbol{X}$ satisfies the *usual hypotheses*. With this at hand, we can now define the following classes of Markov processes, which are central to [31].

Definition 2.2. A Markov process $\boldsymbol{X}$ satisfying the usual hypotheses is called *standard process* if

(i) $\mathcal{X}$ is a locally compact Hausdorff space with countable base (LCCB);

(ii) the paths $t \mapsto X_t$ are right continuous on $[0, \infty)$ and have left limits on $[0, \zeta)$ almost surely.

(iii) $\boldsymbol{X}$ is strong Markov;

(iv) $\boldsymbol{X}$ is quasi-left-continuous.

If $\boldsymbol{X}$ is even quasi-left-continuous on $[0, \infty)$, then we call $\boldsymbol{X}$ a *Hunt process*.

[1] We say that a property Q holds almost surely (or $\mathbb{P}^*$-a.s.) for $\boldsymbol{X}$ if $\{\boldsymbol{X} \text{ satisfies } Q\} \in \mathcal{F}$ and there exists a set $N \in \mathcal{F}$ such that $\mathbb{P}^x(N) = 0$ for all $x \in \mathcal{X}_\vartheta$ and $N^c \subset \{\boldsymbol{X} \text{ satisfies } Q\}$.

Note that quasi-left-continuity on $[0, \infty)$ implies existence of left limits of $t \mapsto X_t$ on $[0, \infty)$ almost surely. Thus, a Hunt process has càdlàg paths almost surely. We will refer to Markov processes with almost surely càdlàg paths as càdlàg Markov processes. From here on, we will always assume that $\mathcal{X}$ is LCCB and hence in particular Polish, i.e. separable and complete metrizable. Standard processes have all the convenient properties that we frequently need when working with Markov processes. However, given a particular Markov model it is not immediately clear how to verify the strong Markov property and quasi-left-continuity. Let us therefore introduce the following convenient class of Markov process. By $\mathcal{C}_0(\mathcal{X})$ we denote the space of $\mathbb{R}$-valued continuous functions on $\mathcal{X}$ vanishing at infinity, with the latter property meaning that for any $\varepsilon > 0$ there exists a compact set $K \subset \mathcal{X}$ such that for any $x \in K^c$ it holds that $|f(x)| < \varepsilon$. If we endow $\mathcal{C}_0(\mathcal{X})$ with the the sup-norm $\|\cdot\|_\infty$, then $(\mathcal{C}_0(\mathcal{X}), \|\cdot\|_\infty)$ is a Banach space since $\mathcal{X}$ is Polish. By $\mathcal{C}_b(\mathcal{X})$ we denote the space of bounded, continuous $\mathbb{R}$-valued functions on $\mathcal{X}$.

DEFINITION 2.3. A càdlàg Markov process $\boldsymbol{X}$ satisfying the usual hypotheses is called *Feller process* if

(i) for any $f \in \mathcal{C}_0(\mathcal{X})$, $\lim_{t\downarrow 0}\|P_t f - f\|_\infty = 0$; (strong continuity)

(ii) for any $t \geq 0$, $P_t\mathcal{C}_0(\mathcal{X}) \subset \mathcal{C}_0(\mathcal{X})$. (Feller property)

If (ii) is satisfied for the transition semigroup of a Markov process $\boldsymbol{X}$, we say that $\boldsymbol{X}$ has the Feller property and $(P_t)_{t\geq 0}$ is a Feller semigroup. If the semigroup of $\boldsymbol{X}$ fulfills

$$P_t\mathcal{B}_b(\mathcal{X}) \subset \mathcal{C}_b(\mathcal{X}), \quad \forall t \geq 0, \tag{2.1}$$

then $\boldsymbol{X}$ is said to have the *strong Feller property*.

Remark 2.4. (a) Since $(P_t)_{t\geq 0}$ is a semigroup, it is easily shown that (i) is equivalent to continuity of $t \mapsto P_t f$ for any $f \in \mathcal{C}_0(\mathcal{X})$ as a mapping from $\mathbb{R}_+$ to $\mathcal{C}_0(\mathcal{X})$, cf. [79, Corollary 1.2].

(b) In presence of the Feller property (ii), strong continuity (i) is automatically satisfied whenever we have pointwise convergence $P_t f(x) \to f(x)$ as $t \downarrow 0$ for all $x \in \mathcal{X}$, see e.g. Kallenberg [100, Theorem 19.6].

(c) The notion of the strong Feller property is a bit misleading, since it does not necessarily imply the Feller property. However, the strong Feller property implies the $\mathcal{C}_b$-Feller property, $P_t\mathcal{C}_b(\mathcal{X}) \subset \mathcal{C}_b(\mathcal{X})$ for all $t \geq 0$ and one may define a $\mathcal{C}_b$-Feller process as a càdlàg Markov process, whose semigroup is (i) strongly continuous on the space $\mathcal{C}_b(\mathcal{X})$ equipped with the topology of local uniform convergence and (ii) has the $\mathcal{C}_b$-Feller property. Some authors prefer this definition of a Feller process to the one above. For $\mathcal{X} = \mathbb{R}^n$, it was shown in [149, Theorem 3.2, Corollary 3.3] that if $x \mapsto \mathbb{P}^x(\zeta < \infty)$ is continuous, then the transition semigroup $(P_t)_{t\geq 0}$ of a Feller process is strongly continuous wrt. local uniform convergence on $\mathcal{C}_b(\mathbb{R}^n)$ and has the $\mathcal{C}_b$-Feller property. Thus, a Feller process on $\mathbb{R}^n$ with infinite lifetime is necessarily a $\mathcal{C}_b$-Feller process, but the reverse implication is not true in general.

We emphasize that the requirement of càdlàg paths is not a restriction for Feller processes, since any Markov process with Feller semigroup has a càdlàg version, see [141, Theorem III.2.7]. It is well-known, see e.g. [141, Theorem III.3.1] and [52, Theoerem 2.4] that a Feller process is strong Markov and quasi-left-continuous on $[0,\infty)$ with respect to its augmented natural filtration $\mathbb{F}$. Thus, any Feller process is a Hunt process.

The last general class of Markov processes that we shall encounter in this thesis is given in the definition below, see [152, Definition I.8.1] for a reference. For a given measurable space $(E,\mathcal{E})$, we denote by $\mathcal{E}^* = \cap_\mu \mathcal{E}^\mu$ the σ-algebra of *universally measurable sets*, where the intersection is taken over all finite measures on $(E,\mathcal{E})$ and $\mathcal{E}^\mu$ is the μ-completion of $\mathcal{E}$, i.e., the collection of sets $B \subset E$ such that there exist sets $B_1, B_2 \in \mathcal{E}$ with the property that $B_1 \subset B \subset B_2$ and $\mu(B_2 \setminus B_1) = 0$.

Definition 2.5. (i) Let $\alpha \geq 0$. A non-negative function $f \in \mathcal{B}(\mathcal{X})^*$ is called *α-excessive* for the semigroup $(P_t)_{t\geq 0}$ of a Markov process $\boldsymbol{X}$ if

(a) $\mathrm{e}^{-\alpha t} P_t f \leq f$ for any $t \geq 0$;

(b) $\mathrm{e}^{-\alpha t} P_t f \downarrow f$, pointwise as $t \downarrow 0$.

(ii) A Markov process $\boldsymbol{X}$ satisfying the usual hypotheses is called *Borel right process* if $\boldsymbol{X}$ has right-continuous paths almost surely and if for any α-excessive function f with $\alpha > 0$, $t \mapsto f(X_t)$ is almost surely right-continuous.

While the strong Markov property is satisfied for a Borel right process (cf. [152, Theorem I.7.4]), there is no reason for a Borel right process to be quasi-left-continuous since we do not even assume the existence of left limits. On the other hand, any standard process is Borel right, which follows from [31, Theorem II.2.12]. Thus, Borel right processes can be considered as generalizations of standard processes, with their definition being rich enough to have deep probabilistic and potential theoretic consequences, often referred to as *the general theory*.

In this thesis, we shall not go down this rabbit hole but will only need the basic facts that working with Borel right processes is the natural way if we do not want to restrict generality or run into any measurability issues—in particular regarding measurability of hitting times $T_A = \inf\{t \geq 0 : X_t \in A\}$ for Borel sets $A \in \mathcal{B}(\mathcal{X}_\vartheta)$—one the one hand, and still have access to the strong Markov property on the other hand. This is the reason, why the theory of stability of continuous time Markov processes, which we introduce in the next section, is cast into the framework of Borel right processes. Explicit examples that we will come across on our journey will always (turn out to) be Feller processes and can thus be seamlessly integrated into the general framework. To sum up the introduction of different classes of Markov processes, we have the following set of inclusions:

$$\big\{\text{Feller processes}\big\} \subset \big\{\text{Hunt processes}\big\} \subset \big\{\text{standard processes}\big\} \subset \big\{\text{Borel right processes}\big\}.$$

2.1.2 *Elements of stability theory for Markov processes*

The ultimate goal of stability theory of Markov processes is the quantification of large time asymptotics of the process. It is therefore natural to require the lifetime ζ of a given Markov process $\boldsymbol{X}$ to be infinite almost surely. This of course raises the question why we even bothered to introduce Markov processes with potentially finite lifetime in Section 2.1. Even though not needed for stability analysis of Markov processes and associated statistical considerations, the concept of killed Markov processes naturally appears in Chapter 4 as part of fluctuation theory of

Markov additive processes. The reader may therefore rest assured that the generality of Section 2.1 will prove to be valuable after all.

Suppose therefore for the remainder of the section that $\boldsymbol{X}$ is a non-explosive Borel right Markov process on an LCCB space $\mathcal{X}$. This is the probabilitstic setting underlying Meyn and Tweedie's stability theory for continuous time Markov processes from the 1990s [74, 127, 130, 131]. Among other important contributions, the theory extends classical recurrence concepts and limit theorems from Markov process theory to explicit rates on the convergence to equilibrium, which is central for nonparametric statistical techniques. As such, Meyn and Tweedie's approach does not only provide the continuous time extension to discrete time Markov chain stability theory on uncountable state spaces which dates back at least to W. Doeblin in the 1930s (cf. the collected works [69]) but is deeply rooted in this theory since many proofs work by tracing back the statement to its discrete time analogue via an appropriate random sampling of the process. For a comprehensive account of discrete time stability theory we refer to the classical textbook treatments [128] and [134].

Let us now collect the most important concepts and results that are used throughout the thesis. We say that a σ-finite measure χ on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ is an *invariant measure* for $\boldsymbol{X}$, if

$$\forall B \in \mathcal{B}(\mathcal{X}): \ \mathbb{P}^{\chi}(X_t \in B) := \int_{\mathcal{X}} \mathbb{P}^{x}(X_t \in B)\, \chi(\mathrm{d}x) = \chi(B).$$

Note that an invariant measure is never unique, since any scaling of the measure is again invariant. We therefore say that an invariant measure χ is *essentially* unique if it is unique up to constant multiples. If $\chi(\mathcal{X}) = 1$, we call χ an *invariant distribution* (which is unique under Harris recurrence, which we define below).

A σ-finite measure ψ is called *irreducibility measure* of $\boldsymbol{X}$, if for any Borel set B, $\psi(B) > 0$ implies $U(x,B) > 0$ for any $x \in \mathcal{X}$. Whenever such a measure exists, we say that $\boldsymbol{X}$ is *ψ-irreducible* or simply irreducible when the specific measure does not matter. If $\boldsymbol{X}$ is irreducible, there exists a maximal irreducibility measure ψ in the sense that for any irreducibility measure ν of $\boldsymbol{X}$ it holds that $\nu \ll \psi$, see [164, Theorem 2.1]. We define $\mathcal{B}^+(\mathcal{X}) := \{B \in \mathcal{B}(\mathcal{X}) : \psi(B) > 0\}$ and call sets in $\mathcal{B}^+(\mathcal{X})$ accessible. Note that maximal irreducibility measures are clearly non-unique. Moreover, if $\boldsymbol{X}$ is ψ-irreducible and admits an invariant measure χ, then χ is a maximal irreducibility measure. To see this, let $\psi(B) > 0$, then

$$t\chi(B) = \int_0^t \Big(\int_{\mathcal{X}} \mathbb{P}^x(X_s \in B)\, \chi(\mathrm{d}x) \Big)\, \mathrm{d}s = \int_{\mathcal{X}} \Big(\int_0^t \mathbb{P}^x(X_s \in B)\, \mathrm{d}s \Big)\, \chi(\mathrm{d}x)$$

and by monotone convergence the right hand side converges to $\chi U(B) := \int_{\mathcal{X}} U(x,B)\,\chi(\mathrm{d}x) > 0$ since $U(x,B) > 0$ for all $x \in \mathcal{X}$ by our choice of B. Hence, $\chi(B) > 0$ and $\psi \ll \chi$ follows. The next important concept, *Harris recurrence*, is an even stronger property than irreducibility. We say that $\boldsymbol{X}$ is *μ-Harris recurrent* if there exists a σ-finite measure μ on the state space s.t.

$$\forall B \in \mathcal{B}(\mathcal{X}): \ \mu(B) > 0 \implies \mathbb{P}^x\Big(\int_0^\infty \mathbb{1}_B(X_t)\, \mathrm{d}t = \infty \Big) = 1, \quad \forall x \in \mathcal{X}, \tag{2.2}$$

i.e., if $\mu(B) > 0$, the process almost surely spends infinitely much time in the set B. A powerful implication of Harris recurrence is that an invariant measure of a Markov process having this property (we call such processes *positive Harris recurrent*) is essentially unique, see [17, Théorème

2.5]. Moreover, by the remark succeding this theorem in [17], an invariant measure χ of a Harris recurrent process is a Harris measure. Thus, it is *maximal Harris* in the sense that it dominates any other Harris measure, since any Harris measure is in particular an irreducibility measure and χ is a maximal irreducibility measure, as discussed above. The defining condition for Harris recurrence is often hard to check directly, however, Kaspi and Mandelbaum [101, Theorem 1] provide us with a simpler equivalent criterion for Borel right Markov processes: suppose that there exists a σ-finite measure ν such that for any Borel set B we have the implication

$$\nu(B) > 0 \implies \mathbb{P}^x(T_B < \infty) = 1, \quad \forall x \in \mathcal{X}, \tag{2.3}$$

where $T_B := \inf\{t \geq 0 : X_t \in B\}$ is the first hitting time of B. Then, $\boldsymbol{X}$ is Harris recurrent and a Harris recurrence measure μ is given by

$$\mu(B) = \mathbb{E}^\nu\Big[\int_0^\infty \mathrm{e}^{-t}\mathbb{1}_B(X_t)\,\mathrm{d}t\Big] = \nu U_1 \mathbb{1}_B, \quad B \in \mathcal{B}(\mathcal{X}). \tag{2.4}$$

Let us now recall the notion of *petite* and *small* sets, with the former concept being a generalization of the latter. We say that a non-empty set $C \in \mathcal{B}(\mathcal{X})$ is petite, if there exists a sampling distribution a on $((0,\infty), \mathcal{B}(0,\infty))$ and a non-trivial measure ν_a on the state space such that for the sampled kernel

$$K_a(x, \mathrm{d}y) := \int_{0+}^\infty P_t(x, \mathrm{d}y)\, a(\mathrm{d}t), \quad x, y \in \mathcal{X},$$

it holds that

$$K_a(x, \cdot) \geq \nu_a(\cdot), \quad x \in C.$$

The sampled kernel corresponds to the transition kernel of the discrete-time Markov chain obtained from $\boldsymbol{X}$ by sampling at renewal times of an independent renewal process with increment distribution a. An important special case is the λ-resolvent kernel

$$R_\lambda(x, \mathrm{d}y) := \int_{0+}^\infty \lambda \mathrm{e}^{-\lambda t} P_t(x, \mathrm{d}y)\,\mathrm{d}t = \lambda U_\lambda(x, \mathrm{d}y), \quad x, y \in \mathcal{X},$$

obtained for the sampling distribution $a = \mathrm{Exp}(\lambda)$, $\lambda > 0$. If $a = \delta_\Delta$ for some $\Delta > 0$, then C is called a *small* set and we refer to the sampled chain $\boldsymbol{X}^\Delta := (X_{n\Delta})_{n\in\mathbb{N}_0}$ as the Δ-skeleton of $\boldsymbol{X}$. The importance of petite sets comes from the fact, that petite sets are small for the sampled chain and small sets in discrete time Markov chain theory allow to construct a related Markov chain possessing an atom via the technique of Nummelin splitting, which then makes reasoning well-known for Markov chains on countable state spaces transferrable to the general state space situation, see [128, Chapter 5]. We emphasize that petite sets are by no means rare. To illustrate this point, let us introduce the class of T-*processes*.

Definition 2.6. $\boldsymbol{X}$ is said to be a T-process, if there exists a non-trivial continuous component T for the sampled kernel K_a associated to some sampling distribution a, meaning that

(i) $x \mapsto T(x, B)$ is lower semicontinuous for all $B \in \mathcal{B}(\mathcal{X})$;

(ii) $K_a(x, B) \geq T(x, B)$ for all $x \in \mathcal{X}$ and $B \in \mathcal{B}(\mathcal{X})$.

If $\boldsymbol{X}$ is an irreducible T-process, then every compact subset of $\mathcal{X}$ is petite, see Theorem 5.1 in [164]. We will discuss some examples of T-processes in Section 2.3.

Another central concept that we shall need is *aperiodicity*. We say that $\boldsymbol{X}$ is aperiodic, if there exists a petite set $C \in \mathcal{B}^+(\mathcal{X})$ (i.e., C must be accessible) and some $T \geq 0$ s.t.

$$\forall t \geq T, x \in C\colon \; P_t(x, C) > 0.$$

Alternatively, $\boldsymbol{X}$ is called aperiodic in [128] if there exists some $\Delta > 0$ such that the Δ-skeleton $\boldsymbol{X}^\Delta$ is irreducible, i.e. there exists a σ-finite measure μ on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ such that

$$\mu(B) > 0 \implies \forall x \in \mathcal{X}\colon \; \sum_{n=1}^{\infty} \mathbb{P}^x(X_{n\Delta} \in B) > 0.$$

We demonstrate in Lemma 2.9 that irreducibility of a skeleton chain implies aperidocity of the process.

We are now well-suited to discuss ergodicity of a Markov process. Let $\|\cdot\|_{\mathrm{TV}}$ denote the total variation norm on the space of signed finite measures $\mathcal{M}_b^s(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$, defined by

$$\|\nu\|_{\mathrm{TV}} := \sup_{|g| \leq 1} |\nu(g)|, \quad \nu \in \mathcal{M}_b^s(\mathcal{X}, \mathcal{B}(\mathcal{X})).$$

We say that $\boldsymbol{X}$ having a stationary distribution μ is *ergodic* if

$$\forall x \in \mathcal{X}\colon \quad \lim_{t \to \infty} \|\mathbb{P}^x(X_t \in \cdot) - \mu\|_{\mathrm{TV}} = 0.$$

Clearly, ergodicity implies weak convergence of the marginal distributions of $\boldsymbol{X}$ to its invariant distribution. If $\boldsymbol{X}$ is positive Harris recurrent, Theorem 6.1 in Meyn and Tweedie [130] provides us with a necessary and sufficient criterion for ergodicity in terms of skeletons of the process:

$$\mathbf{X} \text{ is ergodic} \iff \exists \Delta > 0 \text{ s.t. } \boldsymbol{X}^\Delta \text{ is irreducible.} \tag{2.5}$$

Once we know that $\boldsymbol{X}$ is ergodic, a natural question is the rate of convergence of the marginals to the invariant distribution. To this end, [74] investigate convergence in the so called f-norm. For a strictly positive, measurable function $f \in \mathcal{B}(\mathcal{X})$ satisfying $f \geq 1$, the f-norm on $\mathcal{M}_b^s(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ is given by

$$\|\nu\|_f := \sup_{|g| \leq f} |\nu(g)|, \quad \nu \in \mathcal{M}_b^s(\mathcal{X}, \mathcal{B}(\mathcal{X})),$$

where the supremum is taken over all measurable functions g bounded in absolute value by f. Note that for $f \equiv 1$, the f-norm reduces to the total variation norm. We say that the Markov process $\boldsymbol{X}$ with stationary distribution μ is *V-uniformly ergodic* for some measurable function $V \geq 1$ if there exist constants $D, \kappa > 0$ such that

$$\|P_t(x, \cdot) - \mu\|_V \leq DV(x)\mathrm{e}^{-\kappa t}, \quad x \in \mathcal{X}, t \geq 0, \tag{2.6}$$

which in particular implies that the marginal distributions of $\boldsymbol{X}$ converge to the stationary distribution at an exponential rate in total variation. For the latter, we also refer to the process as being *exponentially* or *geometrically ergodic*. Note that the notion V-uniform convergence is motivated by the fact that (2.6) implies

$$|||P_t - \mathbb{1} \otimes \mu|||_V \leq D\mathrm{e}^{-\kappa t}, \quad t \geq 0,$$

where for two Markov kernels P, Q their V-*norm distance* is defined by

$$|||P - Q|||_V := \sup_{x \in \mathcal{X}} \frac{\|P(x,\cdot) - Q(x,\cdot)\|_V}{V(x)},$$

and $\mathbb{1} \otimes \mu$ is the Markov kernel defined by

$$\mathbb{1} \otimes \mu(x, B) = \mu(B), \quad (x, B) \in \mathcal{X} \times \mathcal{B}(\mathcal{X}),$$

see also [128, Chapter 16].

[74] give conditions in terms of drift criteria for the generator, semigroup and resolvent kernel for V-uniform ergodicity. For our treatment of overshoots in Chapter 4, we will choose the resolvent drift criterion for determining the convergence speed of overshoots. More precisely, if $\boldsymbol{X}$ is irreducible and aperiodic and for some $\lambda > 0$ there exist constants $b \in \mathbb{R}_+, \beta \in (0,1)$, a petite set C and a measurable function $V_\lambda \geq 1$ such that

$$R_\lambda V_\lambda \leq \beta V_\lambda(x) + b\mathbb{1}_C, \tag{2.7}$$

Theorem 5.2 in [74] tells us that $\boldsymbol{X}$ is $R_\lambda V_\lambda$-uniformly ergodic. If V_λ is *unbounded off petite sets*, that is $\{x \in \mathcal{X} : V_\lambda(x) \leq z\}$ is petite for any $z > 0$, (2.7) is equivalent to demanding that there exists $\beta_0 \in (0,1)$ such that

$$R_\lambda V_\lambda \leq \beta_0 V_\lambda(x) + b. \tag{2.8}$$

To see this, for $\alpha > 1$ define the petite set $C(\alpha) := \{x \in \mathcal{X} : V_\lambda(x) \leq \alpha b/(1-\beta_0)\}$, then

$$\begin{aligned} \beta_0 V_\lambda + b &\leq \beta_0 V_\lambda + b\mathbb{1}_C + \frac{1}{\alpha}(1-\beta_0)V_\lambda \mathbb{1}_{C^c} \\ &\leq \frac{1}{\alpha}(1 + (\alpha-1)\beta_0)V_\lambda + b\mathbb{1}_C, \end{aligned} \tag{2.9}$$

showing that for any choice of $\alpha > 1$, (2.8) implies (2.7) with $C = C(\alpha)$ and $\beta = (1+(\alpha-1)\beta_0)/\alpha \in (0,1)$. The converse relation is obvious.

General drift criteria for the speed of convergence to the invariant distribution were extended in [72] to the case of *subgeometric* rates. The combined conclusions of Theorem 3.2 and Theorem 4.9 in [72] read that if $\boldsymbol{X}$ is ergodic and for some $\lambda > 0$ there exists

» a closed, petite set C and a constant $b < \infty$,

» a function $\widetilde{V}_\lambda \colon \mathcal{X} \to [1,\infty)$,

» an increasing, differentiable and concave function $\phi\colon [1,\infty) \to (0,\infty)$,

such that

$$R_\lambda \widetilde{V}_\lambda \leq \widetilde{V}_\lambda - \phi \circ \widetilde{V}_\lambda + b\mathbb{1}_C, \tag{2.10}$$

then, provided $R_\lambda \widetilde{V}_\lambda$ is continuous, there exists some constant $c > 0$ such that

$$\|P_t(x,\cdot) - \mu\|_{\mathrm{TV}} \leq c R_\lambda \widetilde{V}_\lambda(x)\Xi(t), \quad t \geq 0, x \in \mathcal{X}, \tag{2.11}$$

where $\Xi(t) = 1/(\phi \circ H_\phi^{-1})(t)$ for $H_\phi(t) = \int_1^t (1/\phi(s))\,\mathrm{d}s$. Note that (2.7) can be recovered for linear ϕ, in which case $\Xi(t) = \mathrm{e}^{-\kappa t}$ for some $\kappa > 0$, and hence exponential ergodicity can be regarded as a special case of this general result.

Alternatively to the resolvent criterion, it is much more common in the literature to construct a Lyapunov type function for the *extended generator* $\mathcal{A}$ of $\boldsymbol{X}$ [61]. We say that a measurable function f belongs to the domain $\mathcal{D}(\mathcal{A})$ of the extended generator of $\boldsymbol{X}$ if there exists a measurable function g such that the process

$$f(X_t) - f(X_0) - \int_0^t g(X_s)\,\mathrm{d}s, \quad t \geq 0,$$

is a $\mathbb{P}^x$-local martingale for any $x \in \mathcal{X}$, in which case we write $\mathcal{A}f := g$. The notion of the extended generator comes from the fact that it is a natural generalization of the concept of the *infinitesimal generator* of a Markov process. For simplicity, suppose for the moment that $\boldsymbol{X}$ is a Feller process. The inifinitesimal generator $\widetilde{\mathcal{A}}\colon \mathcal{C}_0(\mathcal{X}) \to \mathcal{C}_0(\mathcal{X})$ is the operator with domain

$$\mathcal{D}(\widetilde{\mathcal{A}}) := \big\{f \in \mathcal{C}_0(\mathcal{X}) : \lim_{t\downarrow 0}(P_t f - f)/t \text{ exists in } (\mathcal{C}_0(\mathcal{X}), \|\cdot\|_\infty)\big\},$$

defined as

$$\widetilde{\mathcal{A}}f = \lim_{t\downarrow 0}\frac{P_t f - f}{t}, \quad f \in \mathcal{D}(\widetilde{\mathcal{A}}),$$

where the limit is taken in the sense of the topology on $\mathcal{C}_0(\mathcal{X})$ induced by the $\|\cdot\|_\infty$-norm. Proposition 1.4 in Chapter 4 of [79] demonstrates that for any $f \in \mathcal{D}(\widetilde{\mathcal{A}})$ the process

$$f(X_t) - f(X_0) - \int_0^t \widetilde{\mathcal{A}}f(X_s)\,\mathrm{d}s, \quad t \geq 0,$$

is a (true) $\mathbb{P}^x$-martingale and thus $\mathcal{D}(\widetilde{\mathcal{A}}) \subset \mathcal{D}(\mathcal{A})$ and $\mathcal{A}f = \widetilde{\mathcal{A}}f$ for any $f \in \mathcal{D}(\widetilde{\mathcal{A}})$.

For the purposes of stability theory, the infinitesimal generator is too restrictive since it allows only bounded functions as test functions. For the extended generator, we have the following equivalent condition to (2.7) for V-uniform ergodicity of ergodic processes $\boldsymbol{X}$: if there exists some function $V \in \mathcal{D}(\mathcal{A})$ such that $V \geq 1$, a petite set C and constants $c, b > 0$ such that

$$\mathcal{A}V \leq -cV + b\mathbb{1}_C, \tag{2.12}$$

and $\sup_{x\in C} V(x) < \infty$, then $\boldsymbol{X}$ is V-uniformly ergodic [74, Theorem 5.2]. As for the resolvent drift criterion, if V is unbounded off petite sets, (2.12) is equivalent to requiring that $\mathcal{A}V \leq -\widetilde{c}V + b$ for some constant $\widetilde{c} > 0$. This criterion will be very convenient in Chapters 3 and 5 where we deal with exponential ergodicity of (Lévy driven) SDEs, for which Itō's formula for semimartingales provides nice expressions for the extended generator. For an extension of the generator drift criterion to subexponential ergodicity, see [72, Theorem 3.4].

Studying exponential and subgeometric convergence is not only interesting in its own right, but does have direct implications on the mixing behavior of the Markov process. For two σ-algebras $\mathcal{G}$ and $\mathcal{H}$ and a given probability measure $\mathbf{P}$, introduce the β-mixing coefficient

$$\beta_{\mathbf{P}}(\mathcal{G},\mathcal{H}) := \sup_{C\in\mathcal{G}\otimes\mathcal{H}}\big|\mathbf{P}|_{\mathcal{G}\otimes\mathcal{H}}(C) - \mathbf{P}|_{\mathcal{G}}\otimes\mathbf{P}|_{\mathcal{H}}(C)\big|, \tag{2.13}$$

where $\mathbf{P}|_{\mathcal{G}\otimes\mathcal{H}}$ is the restriction to $(\Omega\times\Omega, \mathcal{G}\otimes\mathcal{H})$ of the image measure of $\mathbf{P}$ under the canonical injection $\iota(\omega) = (\omega,\omega)$. Noting that for $A \times B \in \mathcal{G}\otimes\mathcal{H}$, it holds that $\mathbf{P}|_{\mathcal{G}\otimes\mathcal{H}}(A\times B) = \mathbf{P}(A\cap B)$, it is clear that the β-mixing coefficient should be interpreted as a measure of independence of the σ-algebras. For the Markov process $\boldsymbol{X}$ with natural filtration $\mathbb{F}^0 = (\mathcal{F}_t^0)_{t\geq 0}$ and a given initial distribution η let us now define

$$\beta(\eta,t) = \sup_{s\geq 0} \beta_{\mathbb{P}^\eta}(\mathcal{F}_s^0, \overline{\mathcal{F}}_{s+t}^0), \quad t>0, \tag{2.14}$$

where we denoted by $\overline{\mathcal{F}}_t^0 = \sigma(X_s, s\geq t)$ the σ-algebra of the future after time t. We then say that $\boldsymbol{X}$ is β-mixing when started in η, if $\lim_{t\to\infty}\beta(\eta,t) = 0$ Hence, if $\boldsymbol{X}$ is β-mixing we can roughly state that there is an asymptotic independence between the past and the future of the Markov process. If there even exist constants $C,\kappa>0$ such that $\beta(\eta,t)\leq C\mathrm{e}^{-\kappa t}$, we call $\boldsymbol{X}$ *exponentially* β-mixing.

[173, Lemma 1.4] gives

$$\beta_{\mathbb{P}^\eta}(\mathcal{F}_s^0, \overline{\mathcal{F}}_{t+s}^0) = \mathbb{E}^\eta\Big[\sup_{B\in\overline{\mathcal{F}}_{t+s}^0} |\mathbb{P}^\eta(B|\mathcal{F}_s^0) - \mathbb{P}^\eta(B)|\Big].$$

Proposition 1 in [62] therefore demonstrates that

$$\beta(\eta,t) = \sup_{s\geq 0}\int_{\mathcal{X}} \|\mathbb{P}^x(X_t\in\cdot) - \mathbb{P}^\eta(X_{t+s}\in\cdot)\|_{\mathrm{TV}}\,\mathbb{P}^\eta(X_s\in\mathrm{d}x), \quad t>0,$$

which in case $\eta=\mu$ is the stationary distribution, reduces to

$$\beta(\mu,t) = \int_{\mathcal{X}} \|P_t(x) - \mu\|_{\mathrm{TV}}\,\mu(\mathrm{d}x), \quad t>0.$$

Masuda [123, Lemma 3.9] uses this characterization to establish that if we have (sub)geometric decay as in (2.11) for $\boldsymbol{X}$ and moreover

$$\varrho(\eta) := \sup_{t\geq 0} c\mathbb{E}^\eta[R_\lambda\widetilde{V}_\lambda(X_t)] < \infty, \tag{2.15}$$

then $\boldsymbol{X}$ started in η is β-mixing at rate $\Xi(t)$ with

$$\beta(\eta,t)\leq 2\varrho(\eta)\Xi(t), \quad t>0.$$

2.2 Some technical results

We start with two technical contributions to stability theory of Markov processes, which will be useful for our subsequent developments.

Identifying an invariant measure of a given Markov process is one of the fundamental challenges from a theoretical perspective and rich in consequences for applications. Although the definition of the invariant measure is given in terms of the transition function of the Markov process, many explicit Markov models call for alternative characterizations of invariance in terms of different charateristics of the process since the transition function may not be known or takes an inconvenient analytic form. Prominent examples from the existing literature are conditions on the infinitesimal generator $\widetilde{\mathcal{A}}$ or the resolvent $(U_\lambda)_{\lambda>0}$ of the process.

Suppose that $\boldsymbol{X}$ is an unkilled Feller process with LCCB state space $\mathcal{X}$ such that an invariant distribution μ exists. Then for any $f\in\mathcal{D}(\widetilde{\mathcal{A}})$ we have

$$\int_{\mathcal{X}} \widetilde{\mathcal{A}}f(x)\,\mu(\mathrm{d}x) = \lim_{t\downarrow 0}\frac{1}{t}\int_{\mathcal{X}}(P_tf(x)-f(x))\,\mu(\mathrm{d}x) = 0,$$

where we used dominated convergence and the definition of the generator for the first equality and invariance for the second equality. Thus, an invariant distribution μ necessarily satisfies

$$\mu(\widetilde{\mathcal{A}}f) = 0, \quad \forall f \in \mathcal{D}(\widetilde{\mathcal{A}}). \tag{2.16}$$

Conversely, by Kolmogorov's backward equation $\frac{\mathrm{d}}{\mathrm{d}t}P_t f = \widetilde{\mathcal{A}}P_t f$ for $t > 0$ and $f \in \mathcal{D}(\widetilde{\mathcal{A}})$, it follows that

$$P_t f(x) = f(x) + \int_0^t \widetilde{\mathcal{A}}P_s f(x)\,\mathrm{d}s, \quad t \geq 0, x \in \mathcal{X}, f \in \mathcal{D}(\widetilde{\mathcal{A}}).$$

Consequently, an application of Fubini's theorem yields that for some probability measure μ

$$\mu P_t(f) = \mu(f) + \int_0^t \int_{\mathcal{X}} \widetilde{\mathcal{A}}P_s f(x)\,\mu(\mathrm{d}x)\,\mathrm{d}s = \mu(f) + \int_0^t \mu(\widetilde{\mathcal{A}}P_s f)\,\mathrm{d}s.$$

Hence, if μ satisfies (2.16) then $\mu P_t(f) = \mu(f)$ for all $f \in \mathcal{D}(\widetilde{\mathcal{A}})$ and $t \geq 0$. Since $\boldsymbol{X}$ is Feller the Hille–Yosida theorem tells us that $\widetilde{\mathcal{A}}$ is closed and $\mathcal{D}(\widetilde{\mathcal{A}})$ is dense in $\mathcal{C}_0(\mathcal{X})$. This together with $\sigma(\mathcal{C}_0(\mathcal{X})) = \mathcal{B}(\mathcal{X})$ since $\mathcal{X}$ is Polish, shows that

$$\mu P_t(f) = \mu(f), \quad \forall f \in \mathcal{B}_b(\mathcal{X}), t \geq 0 \iff \mu(\widetilde{\mathcal{A}}f) = 0, \quad \forall f \in \mathcal{D}(\widetilde{\mathcal{A}}). \tag{2.17}$$

This is still not a convenient characterization since apart from very few special cases, the domain $\mathcal{D}(\widetilde{\mathcal{A}})$ is not fully known. However, it is sufficient to require (2.16) restricted to a core $\mathcal{D}_0 \subset \mathcal{D}(\widetilde{\mathcal{A}})$—which can be determined for many explicit Markov models—to hold for invariance. That is, (2.17) can be improved to

$$\mu P_t(f) = \mu(f), \quad \forall f \in \mathcal{B}_b(\mathcal{X}), t \geq 0 \iff \mu(\widetilde{\mathcal{A}}f) = 0, \quad \forall f \in \mathcal{D}_0,$$

for a probability measure μ, see e.g. [118, Theorem 3.37]. Among many other applications, this statement is particularly convenient to derive the invariant distribution of the solution $\boldsymbol{X}$ of a scalar Itō-SDE of the form

$$\mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t.$$

Here, appropriate conditions on the drift coefficient b and the diffusion coefficient σ are needed to guarantee stationarity and the Feller property of $\boldsymbol{X}$, in which case $\mathcal{C}_0^2(\mathbb{R})$ is known to be a core of $\mathcal{D}(\widetilde{\mathcal{A}})$ and we have

$$\widetilde{\mathcal{A}}f(x) = b(x)f'(x) + \frac{1}{2}\sigma^2(x)f''(x), \quad x \in \mathbb{R}, f \in \mathcal{C}_0^2(\mathbb{R}).$$

We will revisit this case in full detail in Chapter 5, where the explicit form of the stationary distribution will play a central role for our statistical approach to a data-driven solution of a singular control problem for ergodic SDEs. For the overshoot process of a MAP considered in Chapter 4, the infinitesimal generator has an exceptionally simple form away from the boundary, however it is difficult to determine directly a convenient core of the generator since the essence of the process is captured in its boundary behavior. Instead of working with the generator in this case, we will therefore opt for the resolvent instead, which we determine explicitly.

In terms of the resolvent kernel $U_1 = R_1$ of $\boldsymbol{X}$, we have the following classical equivalent characterization of an invariant measure.

Proposition 2.7. *[17, Proposition 2.1] A measure μ on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ is invariant for $\boldsymbol{X}$ if and only if*

$$\mu U_1 = \mu.$$

This result is still not easily applicable for our specific overshoot application. We therefore prove the following related criterion in terms of the resolvent of $\boldsymbol{X}$, which will be used in Chapter 4 to determine the essentially unique invariant overshoot measure.

Proposition 2.8. *Suppose that $\mathcal{H} \subset \mathcal{B}_b(\mathcal{X}) \cap \mathcal{B}_+(\mathcal{X})$ such that $P_t\mathcal{H} \subset \mathcal{H}$ for any $t \geq 0$ and there is a non-trivial measure χ on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ and a family $(\alpha_\lambda)_{\lambda>0}$ of finite measures on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ satisfying $\lim_{\lambda\downarrow 0} \alpha_\lambda(\mathcal{X}) = 0$ such that for any $f \in \mathcal{H}$*

$$\lim_{\lambda\downarrow 0} \alpha_\lambda U_\lambda(f) = \chi(f). \tag{2.18}$$

Then, for any $t \geq 0$ and $f \in \mathcal{H}$,

$$\chi P_t(f) = \chi(f).$$

In particular, if $\mathcal{H} = \mathcal{B}_b(\mathcal{X}) \cap \mathcal{B}_+(\mathcal{X})$ (i.e. $\alpha_\lambda U_\lambda$ converges strongly to χ as $\lambda \downarrow 0$), then χ is an invariant measure of $\boldsymbol{X}$.

Proof. Let $f \in \mathcal{H}$ such that (2.18) holds and $t \geq 0$. We have for any $\lambda > 0$ by the semigroup property of $(P_t)_{t\geq 0}$

$$\begin{aligned}
\alpha_\lambda U_\lambda(P_t f) &= \int_{\mathcal{X}} \int_0^\infty \mathrm{e}^{-\lambda s} P_s P_t f(x)\, \mathrm{d}s\, \alpha_\lambda(\mathrm{d}x) \\
&= \int_{\mathcal{X}} \int_0^\infty \mathrm{e}^{-\lambda s} P_{s+t} f(x)\, \mathrm{d}s\, \alpha_\lambda(\mathrm{d}x) \\
&= \mathrm{e}^{\lambda t} \int_{\mathcal{X}} \int_t^\infty \mathrm{e}^{-\lambda s} P_s f(x)\, \mathrm{d}s\, \alpha_\lambda(\mathrm{d}x) \\
&= \mathrm{e}^{\lambda t} \left(\alpha_\lambda U_\lambda(f) - \int_{\mathcal{X}} \int_0^t \mathrm{e}^{-\lambda s} P_s f(x)\, \mathrm{d}s\, \alpha_\lambda(\mathrm{d}x) \right).
\end{aligned}$$

Since $|\int_{\mathcal{X}} \int_0^t \mathrm{e}^{-\lambda s} P_s f(x)\, \mathrm{d}s\, \alpha_\lambda(\mathrm{d}x)| \leq t\|f\|_\infty \alpha_\lambda(\mathcal{X})$ it therefore follows by our assumption that $\alpha_\lambda(\mathcal{X}) \to 0$ and

$$\alpha_\lambda U_\lambda(f) \to \chi(f)$$

as $\lambda \downarrow 0$ that

$$\lim_{\lambda\downarrow 0} \alpha_\lambda U_\lambda(P_t f) = \chi(f).$$

On the other hand, our assumptions and $P_t f \in \mathcal{H}$ yield that

$$\lim_{\lambda\downarrow 0} \alpha_\lambda U_\lambda(P_t f) = \chi(P_t f)$$

and hence

$$\chi P_t(f) = \chi(f)$$

follows. If $\mathcal{H} = \mathcal{B}_b(\mathcal{X}) \cap \mathcal{B}_+(\mathcal{X})$, then for any $B \in \mathcal{B}(\mathcal{X})$ the choice $f = \mathbb{1}_B$ shows that

$$\mathbb{P}^\chi(X_t \in B) = \chi(B), \quad \forall t \geq 0,$$

i.e. χ is an invariant measure. ∎

As a next step, let us give some clarifying remarks on the notion of aperiodicity of a continuous-time Markov process that was introduced in Section 2.1. It seems to be well-known in the literature that the existence of an irreducible skeleton chain for a Harris recurrent Markov process implies aperiodicity, but there is no concrete statement to be found. Proposition 6.1 in [130], which [72] refers to, does not quite state that irreducibility of skeletons implies aperiodicity, but indeed provides the right tool to prove it. For completeness we give the short proof and make the additional simple observation that if the petite set C in the definition of aperiodicity is a singleton set, then aperiodicity also implies the existence of an irreducible skeleton chain, which will be useful later on.

LEMMA 2.9. *Suppose that $\boldsymbol{X}$ is positive Harris recurrent, Borel right and its state space is locally compact and separable. Then, if there exists some irreducible skeleton chain, $\boldsymbol{X}$ is aperiodic. Conversely, if $\boldsymbol{X}$ is aperiodic and the defining set C is a singleton set, then any Δ-skeleton is irreducible.*

Proof. Suppose first that there exists some irreducible Δ-skeleton. Then, the assumptions on the process allow to use Proposition 6.1 from [129], which states that for any petite set C there exists some non-trivial measure μ and and a $T > 0$ such that for all $t \geq T$ we have

$$\mathbb{P}^x(X_t \in \cdot) \geq \mu(\cdot), \quad \forall t \geq T, x \in C, \tag{2.19}$$

which implies in particular that C is even a small set. By the Markov property it thus follows for $s \geq 0$ that

$$\mathbb{P}^x(X_{t+s} \in \cdot) = \int_{\mathcal{X}} \mathbb{P}^x(X_t \in \mathrm{d}y)\, \mathbb{P}^y(X_s \in \cdot) \geq \int_{\mathcal{X}} \mu(\mathrm{d}y)\, \mathbb{P}^y(X_s \in \cdot) = \mathbb{P}^\mu(X_s \in \cdot), \quad \forall t \geq T, x \in C. \tag{2.20}$$

By Proposition 3.4 of Meyn and Tweedie [127] the state space $\mathcal{X}$ can be covered by countably many petite sets (= small sets in our case), hence we may assume that $\psi(C) > 0$, i.e. $C \in \mathcal{B}^+(\mathcal{X})$. Note that $U(x, C) > 0$ for all $x \in \mathcal{X}$ and non-triviality of μ then yield that $\mu U(C) = \int_{\mathcal{X}} U(x, C)\, \mu(\mathrm{d}x) > 0$ and since with Fubini $\mu U(C) = \int_0^\infty \mathbb{P}^\mu(X_t \in C)\, \mathrm{d}t$ it follows that there exists $s > 0$ such that $\mathbb{P}^\mu(X_s \in C) > 0$. From (2.20) it thus follows that for such s and all $t \geq T + s$ and $x \in C$ it holds that

$$\mathbb{P}^x(X_t \in C) \geq \mathbb{P}^\mu(X_s \in C) > 0,$$

which proves aperiodicity of $\boldsymbol{X}$.

Suppose now that $\boldsymbol{X}$ is aperiodic with defining small singleton set $C = \{c\} \in \mathcal{B}^+(\mathcal{X})$ for some $c \in \mathcal{X}$. Then, there exists $T > 0$ such that

$$\mathbb{P}^c(X_t = c) > 0, \quad \forall t \geq T,$$

and δ_c is an irreducibility measure. Then, for given $x \in \mathcal{X}$, there exist t_x such that $\mathbb{P}^x(X_{t_x} = c) > 0$ and the Markov property yields for any $t \geq T$

$$\mathbb{P}^x(X_{t_x+t} = c) \geq \mathbb{P}^x(X_{t_x+t} = c, X_{t_x} = c) = \mathbb{P}^x(X_{t_x} = c)\mathbb{P}^c(X_t = c) > 0.$$

Hence, for given $\Delta > 0$, if we choose $n \in \mathbb{N}$ such that $n\Delta \geq t_x + T$, it follows that $\mathbb{P}^x(X_{n\Delta} = c) > 0$ and thus $\boldsymbol{X}^\Delta$ is δ_c-irreducible. ■

2.3 A transition density criterion for exponential β-mixing

This section motivates the statistical setting of Section 3.1.1 in Chapter 3, where we are considering $\mathbb{R}^d$-valued Borel right Markov processes having a unique invariant distribution μ and possessing transition densities, i.e.,

$$P_t(x, \mathrm{d}y) = p_t(x, y)\,\mathrm{d}y, \quad x, y \in \mathbb{R}^d, t > 0,$$

for some $\mathcal{B}(\mathbb{R}^d)$-measurable functions $(p_t)_{t>0}$. Then, $\mu(\mathrm{d}y) = \rho(y)\,\mathrm{d}y$ with some Lebesgue density ρ and we require that for any compact set $\mathcal{S} \subset \mathbb{R}^d$ there exists some measurable function $r_{\mathcal{S}} \colon (0, \infty) \longrightarrow \mathbb{R}_+$ such that

$$\forall t > 1: \quad \sup_{x,y \in \mathcal{S}} |p_t(x, y) - \rho(y)| \leq r_{\mathcal{S}}(t) \text{ with } \int_1^\infty r_{\mathcal{S}}(t)\,\mathrm{d}t = c_{\mathcal{S}} < \infty. \tag{2.21}$$

This condition together with a heat kernel bound on the short time behavior of the transition densities will allow us to prove tight variance bounds on integral functionals of $\boldsymbol{X}$, which appear naturally in nonparametric statistical estimation for stochastic processes. This is particularly noteworthy in dimension $d = 1$. In this case our alternative assumption of exponential β-mixing is without further assumptions not quite strong enough to obtain bounds that yield optimal rates on, say, kernel invariant density estimators. The statistical setting based on the transition density convergence (2.21) on the other hand can provide such bounds.

Our goal is now to show under which additional hypothesis on the process, (2.21) implies the exponentially β-mixing property in order to make the connection to the β-mixing framework underlying Section 3.1.2, where our main statistical tool for sup-norm estimation procedures is developed.

As the following proposition shows, we need no more than irreducibility as well as the property that compact sets are small together with exponential decay in (2.21) to infer exponential β-mixing of the stationary process.

Proposition 2.10. *Suppose that $\boldsymbol{X}$ is ψ-irreducible with stationary distribution μ and that every compact set $\mathcal{S} \subset \mathbb{R}^d$ is small. Moreover, let (2.21) be satisfied for*

$$r_{\mathcal{S}}(t) := C_{\mathcal{S}} \mathrm{e}^{-\kappa_{\mathcal{S}} t}, \quad t > 0, \tag{2.22}$$

with constants $C_{\mathcal{S}}, \kappa_{\mathcal{S}} > 0$. Then, $\boldsymbol{X}$ started in μ is exponentially β-mixing.

Proof. Let $\mathcal{S} \subset \mathbb{R}^d$ be compact such that $\lambda(\mathcal{S}) > 0$. Since $\mathbb{R}^d$ can be covered by countably many compact sets and the irreducibility measure ψ is σ-finite, we can also assume that $\psi(\mathcal{S}) > 0$ and $\mu(\mathcal{S}) > 0$. Letting $(P_t)_{t \geq 0}$ denote the semigroup associated to $\boldsymbol{X}$, we obtain from (2.21) and (2.22) that, for any $x \in \mathcal{S}$ and $t > 0$,

$$|P_t(x, \mathcal{S}) - \mu(\mathcal{S})| \leq \int_{\mathcal{S}} |p_t(x, y) - \rho(y)|\,\mathrm{d}y \leq C_{\mathcal{S}} \mathrm{e}^{-\kappa_{\mathcal{S}} t} \lambda(\mathcal{S}) = \widetilde{C}_{\mathcal{S}} \mathrm{e}^{-\kappa_{\mathcal{S}} t},$$

with $\widetilde{C}_{\mathcal{S}} = C_{\mathcal{S}} \lambda(\mathcal{S})$. Since $\mu(\mathcal{S}) > 0$, this implies in particular that there exists $T(\mathcal{S}) > 0$ such that $P_t(x, \mathcal{S}) > 0$ for all $t \geq T(\mathcal{S})$ and $x \in \mathcal{S}$. Since $\mathcal{S}$ is small by assumption, it follows that $\boldsymbol{X}$

is aperiodic. Hence, by Theorem 5.3 in [74] and the remarks thereafter, there exists (a) an extended real-valued measurable function $V \geq 1$ such that, for some $T > 0$, we have

$$P_T V(x) \leq \lambda V(x) + b \mathbf{1}_\Theta \tag{2.23}$$

for some $0 < \lambda < 1$, $b \geq 0$ and a small set $\Theta \in \mathcal{B}(\mathbb{R}^d)$ and (b) a set $S_V \subset \{V < \infty\}$, which is full and absorbing—that is, $\mu(S_V) = 1$ and $P_T(x, S_V) = 1$ for any $x \in S_V$—such that $\mathbf{X}$ restricted to S_V is exponentially ergodic in the sense

$$\|P_t(x, \cdot) - \mu\|_{\mathrm{TV}} \leq CV(x)\mathrm{e}^{-\kappa t}, \quad x \in S_V, \tag{2.24}$$

for some constants $C, \kappa > 0$. Noting that (2.23) implies

$$\Delta \widetilde{V} \leq -V + \frac{b}{1-\lambda} \mathbf{1}_\Theta$$

with $\widetilde{V} = V/(1-\lambda) \geq 0$ and $\Delta := P_T - \mathbb{I}$, it follows from Theorem 14.0.1 in [128] that $\mu(V) < \infty$. The claim on exponential β-mixing of the process now follows from (2.24) since

$$\begin{aligned} \beta(\mu, t) = \int_{\mathbb{R}^d} \|P_t(x, \cdot) - \mu\|_{\mathrm{TV}}\, \mu(\mathrm{d}x) &= \int_{S_V} \|P_t(x, \cdot) - \mu\|_{\mathrm{TV}}\, \mu(\mathrm{d}x) \\ &\leq C\mathrm{e}^{-\kappa t} \int_{S_V} V(x)\, \mu(\mathrm{d}x) \\ &= \widetilde{C}\mathrm{e}^{-\kappa t}, \end{aligned}$$

for any $t > 0$, where finiteness of $\widetilde{C} = C\mu(V)$ was discussed above and for the first equality we used that S_V is full. ∎

Compactness of small sets can be inferred for a quite general class of Markov processes, namely T-processes introduced in Section 2.1. Many processes in applied probability can be shown to be T-processes such as price processes driven by Lévy risk and return processes [137], certain piecewise deterministic Markov processes used for MCMC [29] or queuing networks [73]. Moreover, any open set irreducible weak $\mathcal{C}_b$-Feller process is a T-process (cf. [164, Theorem 7.1]), which is a convenient criterion whenever transition densities exist. Markov processes having the strong Feller property are trivially T-processes, since any operator P_t is a continuous component for itself. The strength of Markov processes with the strong Feller property (or T-processes as a generalization of these processes) comes from making possible to connect distributional properties of the Markov process induced by the semigroup and topological properties of the state space, thus allowing to use knowledge of the topology to infer strong stability results of the Markov process. Classical examples of Markov processes with the strong Feller property are Lévy processes with absolutely continuous semigroup with respect to the Lebesgue measure [91, Theorem 2.2], diffusion processes with hypoelliptic Fisk–Stratonovich-type generator [93, Lemma 5.1], diffusion processes on Hilbert spaces under appropriate assumptions on the coefficients [139, Theorem 1.2], or solutions of different classes of parabolic SPDEs [55, 56, 78, 122]. More recently, the strong Feller property was discussed for switching (jump-)diffusions [174, 176], for jump-diffusions with non-Lipschitz coefficients [175], or Markov semigroups generated by singular SPDEs such as the KPZ equation in Hairer and Mattingly [90]. For an account discussing conditions for which (weak) $\mathcal{C}_b$-Feller processes are even strong Feller, we refer to Schilling and Wang [151].

Let us now infer the exponential β-mixing property for T-processes given exponential decay in (2.21) and, as a natural mixing requirement, ergodicity of the process, i.e.,

$$\|P_t(x,\cdot) - \mu\|_{\mathrm{TV}} \underset{t\to\infty}{\longrightarrow} 0, \quad \forall x \in \mathbb{R}^d.$$

Indeed, if $\boldsymbol{X}$ is ergodic, then dominated convergence shows that $\beta(t) \to 0$ as $t \to \infty$, i.e., stationary, ergodic processes are β-mixing.

Proposition 2.11. *Let $\boldsymbol{X}$ be an ergodic T-process such that* (2.21) *is satisfied for $r_{\mathbb{S}}$ given as in* (2.22). *Then, $\boldsymbol{X}$ is positive Harris recurrent, every compact set is small and $\boldsymbol{X}$ is exponentially β-mixing.*

Proof. For the exponential β-mixing property, it suffices to check that every compact set is small by Proposition 2.10, since ergodicity clearly implies μ-irreducibility of $\boldsymbol{X}$. We prove this property together with positive Harris recurrence at once. To this end, for a given $\varepsilon > 0$, choose a compact set $C \subset \mathbb{R}^d$ such that $\mu(C) \geq 1 - \varepsilon$. Then, for fixed $x \in \mathbb{R}^d$, ergodicity guarantees that $\lim_{t\to\infty} \mathbb{P}^x(X_t \in C) \geq 1 - \varepsilon$, and hence $\boldsymbol{X}$ is bounded in probability on average as defined on p. 495 of [130]. Since $\boldsymbol{X}$ is an irreducible T-process, Theorem 3.2 and Theorem 4.1 of the same paper yield Harris recurrence and petiteness of compact sets. It remains to show that small and petite sets coincide for the given process. The reverse implication of Theorem 6.1 in [130] guarantees that there exists an irreducible skeleton $\boldsymbol{X}^\Delta = (X_{n\Delta})_{n\in\mathbb{N}_0}$ for some $\Delta > 0$ thanks to ergodicity and positive Harris recurrence of $\boldsymbol{X}$. Proposition 6.1 in [130] therefore implies equivalence of small and petite sets, which finishes the proof. ∎

Markovian statistics under mixing assumptions

3

There exist various probabilistic concepts that permit the investigation of quantitative ergodic properties of Markov processes, providing a number of approaches to analyzing the rate of convergence of the process to equilibrium. Such results actually present precious tools for an adequate statistical modeling of complex systems. Markov models, especially of (jump) diffusion-type, find numerous applications in biology, chemistry, natural resource management, computer vision, Bayesian inference in machine learning, cloud computing and many more [8, 36, 75, 76, 81, 88, 161, 165], and ergodicity can usually be seen as some kind of minimum requirement for the development of a fruitful statistical theory. While the probabilistic picture of quantitative ergodic properties is now quite clear, there are still open questions regarding the statistical implications. With this chapter, we want to contribute to closing this gap, paying particular attention to a general Markovian multidimensional setting.

In contrast to the highly-developed statistical theory for scalar diffusion processes, there are relatively few references for nonparametric or high-dimensional general Markov models. To not let sampling effects obscure the statistical implications, it is natural to base the statistical analysis in this context on a continuous observation scheme (i.e., one assumes that a complete trajectory of the process is available). A substantial point of reference for a thorough statistical analysis of ergodic multivariate diffusion processes is provided by the article [59] where the fundamental question of asymptotic statistical equivalence is investigated. Apart from its principal central statement, the work also nicely demonstrates the implications of probabilistic properties of processes on quantitative statistical results. Specifically, heat kernel bounds and the spectral gap inequality are used to prove tight variance bounds for integral functionals which in turn provide fast convergence rates for the specific problem of invariant density estimation. Similar techniques can be used for the in-depth analysis of other statistical questions such as (adaptive) estimation of the drift vector of an ergodic diffusion (cf. [160], [159]). The results in [59, 159, 160] are developed for diffusion processes with drift of gradient-type and unit diffusion matrix. While in this specific case the reversibility assumption is directly verified, the condition of symmetry of the process presents a significant constraint, in particular for solutions of SDEs with jump noise.

More recently, a Bayesian approach to drift estimation of multivariate diffusion processes is undertaken in [133] and [87]. Whilst [87] work in a reversible setting since their approach relies on placing a Gaussian prior on the potential B of the drift $b = \nabla B$ instead of tackling the drift directly, [133] approach drift estimation for non-reversible diffusions by employing PDE techniques to a penalized likelihood estimator. This opens up an excitingly different viewpoint on the statistical handling of multivariate diffusion processes and in case of [133] avoids the need for reversibility, but both approaches restrict the setting to assumed periodicity of the drift coefficient. While this assumption (similar to reversibility) can certainly be justified for specific applications, the approach does not yet provide an answer to the question of how to conduct a statistical analysis of multidimensional Markov processes without strong structural constraints on the coefficients. From a different perspective, the current preprint [12] yields the remarkable observation that quantitatively similar statistical results as in the reversible diffusion case can also be proven for jump diffusions with Lévy-driven jump part, without the need to rely on a reversible or periodic setting, by focusing on assumptions on the characteristics of the process which guarantee exponential ergodicity as the driving force of the statistical approach.

Another branch of the literature that does not consider specific structural assumptions on the process is based on the so called Castellana–Leadbetter condition or variations thereof [34, 39, 113], which imposes finiteness of the integrated uniform distance between the density of the bivariate law of (X_0, X_t) of a stationary Markov process $\boldsymbol{X}$ with stationary density ρ and the product density $\rho \otimes \rho$. This assumption yields dimension independent parametric estimation rates of the invariant density and is thus not suitable for our goal to extend the dimension dependent minimax optimal estimation rates for continuous diffusion processes to more general classes of multidimensional Markov processes, introduced below.

Throughout, we suppose that $(\boldsymbol{X}, (\mathbb{P}^x)_{x\in\mathbb{R}^d})$ is a non-explosive Borel right Markov process with state space $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ and semigroup $(P_t)_{t\geq 0}$ defined by

$$P_t(x,B) \coloneqq \mathbb{P}^x(X_t \in B), \quad x \in \mathbb{R}^d, B \in \mathcal{B}(\mathbb{R}^d).$$

Under regularity assumptions on the coefficients, the exemplary class of (jump) diffusion processes that we study in detail later on belongs to the class of Feller processes and hence falls into our general probabilistic regime. Moreover, as discussed in Section 2.1, Borel right Markov processes are the object of stability analysis of time-continuous Markov processes pioneered by Meyn and Tweedie in the 1990s [74, 127, 130, 132], in which the long-time behavior is quantitatively associated with Lyapunov drift criteria. This approach is central to our probabilistic modeling. We therefore work, as a minimal requirement for stability, in an ergodic setting for $\boldsymbol{X}$ throughout the chapter. That is, the following assumption is in place:

($\mathscr{A}$0) The marginal laws of $\boldsymbol{X}$ are absolutely continuous, i.e., for any $t > 0$ and $x \in \mathbb{R}^d$, there exists a measurable function $p_t \colon \mathbb{R}^d \times \mathbb{R}^d \longrightarrow \mathbb{R}_+$ such that

$$P_t(x,B) = \int_B p_t(x,y)\,\mathrm{d}y, \quad B \in \mathcal{B}(\mathbb{R}^d),$$

and, moreover, $\boldsymbol{X}$ admits a unique absolutely continuous invariant probability measure μ, i.e., there exists a density $\rho \colon \mathbb{R}^d \longrightarrow \mathbb{R}_+$ such that $\mathrm{d}\mu = \rho\,\mathrm{d}\boldsymbol{\lambda}$ and

$$\mathbb{P}^\mu(X_t \in B) \coloneqq \int_{\mathbb{R}^d} P_t(x,B)\,\mu(\mathrm{d}x) = \int_{\mathbb{R}^d}\int_B p_t(x,y)\rho(x)\,\mathrm{d}y\,\mathrm{d}x = \int_B \rho(x)\,\mathrm{d}x = \mu(B)$$

for any Borel set B.

We abbreviate $\mathbb{P}^\mu = \mathbb{P}$, $\mathbb{E}^\mu = \mathbb{E}$ and denote $\mu(g) = \int g\,\mathrm{d}\mu$ for $g \in L^1(\mu)$ or $g \geq 0$. Note that in ($\mathscr{A}$0) existence of a density ρ of the invariant distribution μ is not an additional requirement on $\boldsymbol{X}$, but is guaranteed by the Radon–Nikodym theorem thanks to the definition of invariance and the existence of densities for the transition operators.

Turning away from Lyapunov criteria for general ergodic Markov processes, the long-time behavior of Markovian semigroups is also known to be linked to functional inequalities. The most familiar setting is the L^2 framework with its equivalence between the corresponding Poincaré inequalities and exponential decay of the Markovian semigroup. The relation between both approaches in terms of quantifying ergodic properties of Markov processes is studied in [18].

We want to understand the interaction between the probabilistic concepts and *statistical* properties. In order to obtain a clear picture and benchmark results that are not distorted by discretization errors, we consider a statistical framework including the standing assumption that a continuous observation of a trajectory $\boldsymbol{X}^T = (X_t)_{t\in[0,T]}$ of $\boldsymbol{X}$ is available. For the analysis of statistical methods (e.g., for estimating the characteristics of $\boldsymbol{X}$), variance bounds and deviation inequalities are of central importance. Section 3.1 focuses on the analysis of the variance of additive functionals of the form $\int_0^t f(X_s)\,\mathrm{d}s$ for the ergodic process $\boldsymbol{X}$. We introduce sets of general assumptions on transition and invariant density which allow to prove tight variance bounds (cf. Propositions 3.1 and 3.6). Here, we consider an on-diagonal heat kernel bound to regulate the short-time transitional behavior of the process and either local uniform transition density convergence to the invariant distribution at sufficient speed for any dimension $d \in \mathbb{N}$ or exponential β-mixing in dimension $d \geq 2$ to obtain tight controls on the long-time transitions of the process. The combination of heat kernel bound and local uniform transition density convergence can be interpreted as a localized version of the Castellana–Leadbetter condition that separates the short- and long time effects and considerably weakens the inherent assumptions on the speed at which the law of X_t approaches a singular distribution as $t \downarrow 0$ in higher dimensions. We give a detailed analysis of this condition. We demonstrate how total variation convergence at sufficient speed implies the local uniform transition density assumption and argue that in case of μ-a.s. exponential ergodicity of the process, exponential β-mixing and local uniform transition density convergence are essentially equivalent, giving a homogeneous picture of our different sets of assumptions.

In Section 3.2 we proceed by showing how the β-mixing property of $\boldsymbol{X}$—which is satisfied for a wide range of Markov processes appearing in applied and theoretical probability theory—is reflected in uniform moment bounds on empirical processes associated to integral functionals of $\boldsymbol{X}$. More precisely, for countable classes $\mathcal{G}$ of bounded measurable functions g, we establish an upper bound on

$$\left(\mathbb{E}\Big[\sup_{g\in\mathcal{G}}\Big|\frac{1}{T}\int_0^T g(X_s)\,\mathrm{d}s - \int g\,\mathrm{d}\mu\Big|^p\Big]\right)^{1/p}, \quad p \geq 1,$$

(cf. Theorem 3.7) stated in terms of entropy integrals related to $\mathcal{G}$ and the variance of the integral functionals. This result holds for β-mixing Borel right processes on general state spaces without any assumptions on the existence of transition densities, i.e., Assumption $(\mathscr{A}0)$ is diminished to stationarity which further increases the applicability of our findings for future investigations. Such moment bounds and associated uniform deviation inequalities are generally the focal point for efficient implementation of adaptive estimation procedures, both for the L^2 and the sup-norm risk. In our concrete estimation context, we use the uniform moment bounds together with the variance bounds from Section 3.1 to establish oracle-type deviation inequalities for the sup-norm risk of a kernel invariant density estimator that is essential for the adaptive estimation scheme considered in Section 3.3 that we describe below. Our motivation to study sup-norm estimation is not only rooted in the higher degree of intepretability of such statements compared to the pointwise L^2 risk and the general usefulness of mathematical results obtained along the way, but also comes from the observation that certain problems from applied probability can only be handled with statistical tools, when sup-norm estimation bounds of a quantitiy of interest are available. This point is highlighted in Chapter 5, where the general framework presented in this chapter is implemented for the development of data-driven stochastic optimal control strategies for diffusions and Lévy processes.

Making the mixing behavior of the process a cornerstone of the statistical analysis is completely natural when comparing to discrete time theory. For discrete observations it is well-established by in the field of weak dependence that different sets of mixing assumptions (e.g., α-mixing or β-mixing) and relaxations thereof can produce variance bounds and deviation inequalities that hold up to analogous results from i.i.d. observations to yield sharp nonparametric estimation results, see [63, 142] for an overview. Statistically, it is therefore fundamentally interesting whether analysing a continuous time mixing Markov process based on full observations in our framework yields better estimation rates compared to partial observations corresponding to a weakly dependent observation sequence.

Indeed, in presence of the additional analytic tool provided by the heat-kernel bound, we establish in Section 3.3 that the stationary density of exponentially β-mixing Markov processes can be estimated in any dimension at optimal rates both wrt. sup-norm risk and pointwise L^2 risk—where optimality is understood relative to the benchmark minimax rates known for continuous reversible diffusion processes that are faster than the nonparametric rate for well-behaved discretely sampled data. We go even further by showing that in dimension $d \geq 3$—where the optimal bandwidth choice depends on the typically unknown degree of Hölder smoothness β— a Lepski type adaptive bandwidth selection scheme proposed in [85] for i.i.d. data fitted to our needs provides optimal estimation rates up to iterated log-factors (see also [115] for an adaptive scheme for anisotropic sup-norm estimation for i.i.d. observations). More precisely, our main result Theorem 3.11 shows that given a kernel estimator $\widehat{\rho}_{h,T}$ for the unknown invariant density ρ with bandwidth choice

$$h \equiv h(T) \sim \begin{cases} \log^2 T/\sqrt{T}, & d = 1, \\ \log T/T^{1/4}, & d = 2, \\ (\log T/T)^{1/(2\beta+d-2)}, & d \geq 3, \end{cases}$$

we have for any $p \geq 1$ and a bounded open domain D,

$$\mathbb{E}\Big[\sup_{x\in D}\big|\widehat{\rho}_{h,T}(x) - \rho(x)\big|^p\Big]^{1/p} \in \begin{cases} \mathrm{O}\big(\sqrt{\log T/T}\big), & d = 1, \\ \mathrm{O}\big(\log T/\sqrt{T}\big), & d = 2, \\ \mathrm{O}\big((\log T/T)^{\frac{\beta}{2\beta+d-2}}\big), & d \geq 3. \end{cases}$$

If for $d \geq 3$ we replace the smoothness-dependent bandwidth choice $h(T)$ by the adaptive selector $\widehat{h}_T \equiv \widehat{h}_T^{(k)}$ introduced in (3.34) and the order of the kernel is sufficiently large, then for $\log_{(k)} T$ denoting the k-th iterated logarithm,

$$\mathbb{E}\Big[\sup_{x\in D}\big|\widehat{\rho}_{\widehat{h}_T,T}(x) - \rho(x)\big|\Big] \in \mathrm{O}\bigg(\bigg(\frac{\log_{(k)} T \log T}{T}\bigg)^{\frac{\beta}{2\beta+d-2}}\bigg),$$

where $k \in \mathbb{N}$ can, in principle, be chosen arbitrarily large—which however decreases the size of the set of candidate bandwidths for the adaptive selection procedure given a finite oberservation horizon. We emphasize that the logarithmic gap could be avoided if constants appearing in the uniform deviation inequality from Section 3.2 were explicitely calculated. This, however, requires exact knowledge of the ergodic and short time behavior of the process, contradicting a truly adaptive nature of the approach.

Such sup-norm adaptive multivariate estimation results are completely new and complement adaptive L^2 estimation procedures considered in [54] for discrete time mixing chains based on model selection and in [12] for Lévy driven jump-diffusions. We emphasize that [54] also consider estimation of continuous time mixing processes in terms of their sampled skeletons. However, our improved adaptive estimation rates in presence of heat kernel bounds demonstrate that such approach can be considerably improved by not taking a Markov chain viewpoint under partial observations but by exploiting continuous time probabilistic structures under full observations.

As a concrete example, we investigate multidimensional SDEs with Lévy-driven jump part, i.e., Markov processes associated to the solution of

$$\mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t + \gamma(X_{t-})\,\mathrm{d}Z_t, \quad X_0 = x \in \mathbb{R}^d, \tag{3.1}$$

where $\boldsymbol{W}$ is d-dimensional Brownian motion and $\boldsymbol{Z}$ is a pure jump Lévy process independent of $\boldsymbol{W}$. In Section 3.4.1, we investigate Lévy driven Ornstein–Uhlenbeck processes as the basic class of Lévy driven jump diffusions with unbounded drift coefficient. In presence of non-trivial Gaussian part and very mild moment assumptions on the Lévy measure, we infer optimal sup-norm and pointwise L^2 invariant density estimation results in any dimension. In this case, an adaptive estimation procedure is not necessary, since the invariant density is a smooth function. In Section 3.4.2 we allow for more flexible dispersion and jump coefficients σ, γ with the price to be paid being boundedness of the drift b. By considering solutions $\boldsymbol{X}$ to (3.1) under appropriate assumptions on the coefficients b, σ, γ and the jump measure associated to $\boldsymbol{Z}$ we can apply our general statistical results to invariant density estimation for $\boldsymbol{X}$, thus establishing new results on sup-norm adaptive invariant density estimation for such general jump processes.

In the sequel, we concentrate on guiding the reader through our framework and the accompanied mathematical results.

Basic notation. A set $B \in \mathcal{B}(\mathbb{R}^d)$ is called μ-full if $\mu(B) = 1$. We say that the Borel right Markov process $\boldsymbol{X}$ is μ-a.s. V-ergodic at speed Ξ if, for some μ-full set Λ,

$$\|P_t(x,\cdot) - \mu\|_{\mathrm{TV}} \leq CV(x)\Xi(t), \quad t \geq 0, x \in \Lambda, \tag{3.2}$$

where $V\colon \mathbb{R}^d \to [0,\infty]$ with $V\mathbb{1}_\lambda(x) < \infty$ and, for a signed measure ν, $\|\nu\|_{\mathrm{TV}} := \sup_{|f|\leq 1}|\nu(f)|$ denotes its total variation norm. If (3.2) holds with $\Xi(t) = (1+t)^{-\alpha}$ for some $\alpha > 0$, we say that $\boldsymbol{X}$ is μ-a.s. V-polynomially ergodic of degree α. If $\Xi(t) = \mathrm{e}^{-\kappa t}$ for some $\kappa > 0$, then $\boldsymbol{X}$ is called μ-a.s. exponentially ergodic. When $\Lambda = \mathbb{R}^d$ and $V(x) < \infty$ for any $x \in \mathbb{R}^d$, we just say that $\boldsymbol{X}$ is ergodic at speed Ξ (resp., polynomially ergodic and exponentially ergodic).

For any multi-index $\alpha \in \mathbb{N}^d$ and $x \in \mathbb{R}^d$, set $|\alpha| = \sum_{i=1}^d \alpha_i$ and $x^\alpha = \prod_{i=1}^d x_i^{\alpha_i}$. For $\llbracket\beta\rrbracket$ denoting the largest integer *strictly* smaller than β, introduce the Hölder class on an open domain $D \subset \mathbb{R}^d$

$$\mathcal{H}_D(\beta,\mathrm{L}) = \left\{f \in \mathcal{C}^{\llbracket\beta\rrbracket}(D,\mathbb{R}) : \max_{|\alpha|=\llbracket\beta\rrbracket} \sup_{x,y\in D, x\neq y} \frac{|f^{(\alpha)}(x) - f^{(\alpha)}(y)|}{|x-y|^{\alpha-\llbracket\alpha\rrbracket}} \leq \mathrm{L}, \sup_{x\in D}|f(x)| \leq \mathrm{L}\right\}, \tag{3.3}$$

where $f^{(\alpha)} := \frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1}\ldots\partial x_d^{\alpha_d}}$. Recall that a kernel function $K\colon \mathbb{R}^d \to \mathbb{R}$ is said to be of order $\ell \in \mathbb{N}$ if, for any $\alpha \in \mathbb{N}^d$ with $|\alpha| \leq \ell$, $x \mapsto x^\alpha K(x)$ is integrable and, moreover, $\int_{\mathbb{R}^d} K(x)\,\mathrm{d}x = 1$, $\int_{\mathbb{R}^d} K(x)x^\alpha\,\mathrm{d}x = 0$, for $\alpha \in \mathbb{N}^d$, $|\alpha| \in \{1,\ldots,\ell\}$.

3.1 BASIC FRAMEWORK AND VARIANCE ANALYSIS

This first section focuses on the analysis of the variance of integral functionals of the form $\int_0^t f(X_s)\,\mathrm{d}s$ for the ergodic process $\boldsymbol{X} = (X_s)_{0\leq s\leq t}$ under different sets of general assumptions on $\boldsymbol{X}$ that will carry us through the rest of the chapter. Such variance bounds are indispensable tools for statistical applications since (as we will see in Section 3.2) the variance of integral functionals naturally appears in associated deviation inequalities and related moment bounds and thus requires tight estimates.

3.1.1 *Variance analysis under assumptions on transition and invariant density*

Recall the definition of Assumption ($\mathscr{A}0$) from the introduction and let us say that a set $B \in \mathcal{B}(\mathbb{R}^d)$ is μ-full if $\mu(B) = 1$. We start by working under the following set of additional assumptions:

($\mathscr{A}1$) In case $d = 1$, there exists a non-negative, measurable function $\alpha\colon (0,1] \longrightarrow \mathbb{R}_+$ such that, for any $t \in (0,1]$,

$$\sup_{x,y\in\mathbb{R}} p_t(x,y) \leq \alpha(t) \quad \text{and} \quad \int_{0+}^1 \alpha(t)\,\mathrm{d}t = c_1 < \infty,$$

and, in case $d \geq 2$, there exists $c_2 > 0$ such that the following on-diagonal heat kernel estimate holds true:

$$\forall t \in (0,1]: \sup_{x,y\in\mathbb{R}^d} p_t(x,y) \leq c_2 t^{-d/2}. \tag{3.4}$$

($\mathscr{A}2$) There exists a μ-full set Λ such that for any compact set $\mathcal{S} \subset \mathbb{R}^d$, there exists a non-negative, measurable function $r_{\mathcal{S}}\colon (0,\infty) \longrightarrow \mathbb{R}_+$ such that

$$\forall t > 1: \sup_{x\in\mathcal{S}\cap\Lambda, y\in\mathcal{S}} |p_t(x,y) - \rho(y)| \leq r_{\mathcal{S}}(t) \text{ with } \int_1^\infty r_{\mathcal{S}}(t)\,\mathrm{d}t =: c_{\mathcal{S}} < \infty. \tag{3.5}$$

An essential aspect of the statistical analysis of stochastic processes is the influence of the dimension of the underlying process. It is known that certain phenomena (as compared, e.g., to estimation based on i.i.d. observations) occur in the one-dimensional case. However, these phenomena can usually only be detected by means of specific techniques that take advantage of the unique probabilistic characteristics of scalar processes such as local time for one-dimensional diffusion processes. A "standardized" statistical framework which covers all dimensions with similar conditions cannot capture these phenomena. Our assumptions may therefore be understood as an attempt to find general conditions that make no reference to dimension or process specific phenomena, yet yield variance bounds which are tight enough to allow proving optimal convergence rates for nonparametric procedures.

In this regard, they should be compared to the Castellana–Leadbetter condition [39] requiring that

$$\int_{(0,\infty)} \sup_{x,y\in\mathbb{R}^d} |\rho(x)p_t(x,y) - \rho(x)\rho(y)|\,\mathrm{d}t < \infty, \tag{3.6}$$

and which allows L^2 estimation of the invariant density via a kernel estimator at parametric (or *superoptimal* [33]) rate $1/T$ in *any* dimension $d \geq 1$. Since ($\mathscr{A}1$) implies that ρ is bounded, ($\mathscr{A}2$)

can be understood as a localized, unweighted alternative to (3.6) away from 0, which captures the mixing behavior of the process as we discuss below. Our assumption $(\mathscr{A}1)$ corresponds to the integral part of (3.6) close to 0 and guarantees that the distribution of X_t is not too close to a singular distribution. However, in dimension $d \geq 2$ this assumption is much milder than (3.6) since heat-kernel bounds on the transition density are quite common for many multdimensional Markov processes such as strong solutions of (jump) SDEs. On the other hand, (3.6) is too strong for such Markov processes, since, e.g., the minimax optimal L^2 rate for multivariate diffusions processes is known to be worse than $1/T$ and hence the variance bound implied by (3.6) cannot be achieved.

Also note that the transition density bounds formulated in $(\mathscr{A}1)$ are exceptionally weak compared to related literature dealing with statistical estimation of jump processes. E.g., [12] construct their assumptions on the coefficients and the jump measure of a d-dimensional Lévy-driven jump diffusion to guarantee a heat kernel-type estimate of the form

$$p_t(x,y) \lesssim t^{-d/2} \mathrm{e}^{-\lambda \frac{\|y-x\|^2}{t}} + \frac{t}{|\sqrt{t} + \|y-x\||^{d+\alpha}}, \quad x,y \in \mathbb{R}^d, t \in (0,T],$$

for the estimation horizon $T > 0$, where $\alpha \in (0,2)$ is the self-similarity index of a strictly α-stable Lévy process whose Lévy measure is assumed to dominate the Lévy measure governing the jumps of the SDE. Clearly, this condition is stronger than what we require and is fitted to the concrete probabilistic setting. The reason for this specific choice becomes apparent from Corollary 3.16 in Section 3.4.2, but our approach reveals that $(\mathscr{A}1)$ is sufficient to obtain tight variance bounds in a general multivariate setting. Let us now give the variance bounds implied in our framework.

PROPOSITION 3.1. *Suppose that $(\mathscr{A}1)$ and $(\mathscr{A}2)$ are satisfied, and let f be a bounded function with compact support $\mathcal{S}$ fulfilling $\boldsymbol{\lambda}(\mathcal{S}) < 1$. Then, there exists a constant $C > 0$, such that, for any $T > 0$,*

$$\operatorname{Var}\left(\int_0^T f(X_t)\,\mathrm{d}t\right) \leq C(1 \vee c_{\mathcal{S}}) T \|f\|_\infty^2 \boldsymbol{\lambda}(\mathcal{S}) \mu(\mathcal{S}) \psi_d^2(\boldsymbol{\lambda}(\mathcal{S})), \text{ with } \psi_d(x) := \begin{cases} 1, & d = 1, \\ \sqrt{1 + \log(1/x)}, & d = 2, \\ x^{\frac{1}{d} - \frac{1}{2}}, & d \geq 3, \end{cases} \tag{3.7}$$

where the variance is taken with respect to $\mathbb{P}$.

Proof. Without loss of generality, let $T \geq 1$ be fixed. Then, using the Markov property and the invariance of μ, for any $\delta \in [0,1]$,

$$\begin{aligned}
\operatorname{Var}\left(\int_0^T f(X_s)\,\mathrm{d}s\right) &= \mathbb{E}\left[\left(\int_0^T (f(X_s) - \mathbb{E}f(X_0))\,\mathrm{d}s\right)^2\right] \\
&= 2\mathbb{E}\left[\int_0^T \int_0^u (f(X_0) - \mathbb{E}f(X_0))(f(X_{u-s}) - \mathbb{E}f(X_0))\,\mathrm{d}s\,\mathrm{d}u\right] \\
&= 2\int_0^T \int_0^u \left(\mathbb{E}[f(X_0)f(X_{u-s})] - (\mathbb{E}f(X_0))^2\right) \mathrm{d}s\,\mathrm{d}u \\
&= 2\int_0^T \int_0^u \left[\iint_{\mathbb{R}^{d\times d}} f(x)f(y)p_{u-s}(x,y)\,\mathrm{d}y\mu(\mathrm{d}x) - \int f(x)\,\mu(\mathrm{d}x) \int f(y)\rho(y)\,\mathrm{d}y\right] \mathrm{d}s\,\mathrm{d}u \\
&= 2\int_0^T \int_0^u \int_\Lambda \int_{\mathbb{R}^d} f(x)f(y)(p_{u-s}(x,y) - \rho(y))\,\mathrm{d}y\,\mu(\mathrm{d}x)\,\mathrm{d}s\,\mathrm{d}u
\end{aligned}$$

$$= 2(\mathcal{I}(0,\delta) + \mathcal{I}(\delta,1) + \mathcal{I}(1,T)),$$

with (substituting $v = u - s$)

$$\mathcal{I}(a,b) := \int_a^b (T-v) \int_{\mathbb{R}^d} \int_\Lambda f(x)f(y)(p_v(x,y) - \rho(y))\, \mu(\mathrm{d}x)\,\mathrm{d}y\,\mathrm{d}v, \quad 0 \le a < b \le T.$$

It follows from the assumption on the convergence of the transition density in (3.5) that

$$\begin{aligned}\mathcal{I}(1,T) \le & \int_1^T (T-v) \sup_{x\in\mathcal{S}\cap\Lambda, y\in\mathcal{S}} |p_v(x,y) - \rho(y)|\,\mathrm{d}v \iint_{\mathbb{R}^d\times\mathbb{R}^d} f(x)f(y)\,\mu(\mathrm{d}x)\,\mathrm{d}y \\ \le & \; T\|f\|_\infty^2 \boldsymbol{\lambda}(\mathcal{S})\mu(\mathcal{S}) \int_1^T r_{\mathcal{S}}(v)\,\mathrm{d}v \;\le\; c_{\mathcal{S}} T\|f\|_\infty^2 \boldsymbol{\lambda}(\mathcal{S})\mu(\mathcal{S}).\end{aligned}$$

It remains to consider the first parts of the integral. We now restrict to dimension $d \ge 3$; the remaining cases are handled with analogous arguments. Note first that

$$\mathcal{I}(0,\delta) \;\le\; T\|f\|_\infty^2 \int_0^\delta \iint_{\mathcal{S}\times\mathbb{R}^d} p_v(x,y)\,\mu(\mathrm{d}x)\,\mathrm{d}y\,\mathrm{d}v \;=\; T\|f\|_\infty^2 \mu(\mathcal{S})\delta. \tag{3.8}$$

On the other hand, the heat kernel bound (3.4) gives for any $x, y \in \mathbb{R}^d$,

$$\int_\delta^1 p_v(x,y)\,\mathrm{d}v \le c_2 \int_\delta^1 v^{-d/2}\,\mathrm{d}v = c_2' \delta^{1-d/2},$$

where $c_2' = 2/(d-2)c_2$. Letting $\delta = (\boldsymbol{\lambda}(S))^{2/d}$ and exploiting that $\boldsymbol{\lambda}(\mathcal{S}) < 1$, it follows

$$\mathcal{I}(\delta,1) \;\le\; T\|f\|_\infty^2 \int_\delta^1 \iint_{\mathcal{S}^2} p_v(x,y)\,\mu(\mathrm{d}x)\,\mathrm{d}y\,\mathrm{d}v \;\le\; c_2' T\|f\|_\infty^2 \mu(\mathcal{S})(\boldsymbol{\lambda}(\mathcal{S}))^{\frac{2}{d}}.$$

■

Remark 3.2 (comparison to spectral gap approach). Let f be a bounded function with compact support $\mathcal{S}$ of Lebesgue measure $\boldsymbol{\lambda}(\mathcal{S}) < 1$, and denote by $f_c := f - \int f\,\mathrm{d}\mu$ its centered version. Assuming stationarity and symmetry of the process $\boldsymbol{X}$, the proof of Proposition 1 in [59] starts from the representation

$$\mathrm{Var}\left(\int_0^T f(X_s)\,\mathrm{d}s\right) \;=\; 2\int_0^T (T-u)\mathbb{E}[f_c(X_0)f_c(X_u)]\,\mathrm{d}u \;\le\; 2T\int_0^T \langle f_c, P_u f_c\rangle\,\mathrm{d}u.$$

Fix $D \in (0,T]$. Heat kernel bounds of the form (3.4) yield an upper bound on the integral from 0 to D, in dimension $d = 1$ specified as

$$\int_0^D \langle f_c, P_u f_c\rangle\,\mathrm{d}u \;\lesssim\; \mu(\mathcal{S})\boldsymbol{\lambda}(\mathcal{S})\|f\|_\infty^2 \left(2D + \frac{2}{3-\alpha} D^{\frac{3-\alpha}{2}}\right). \tag{3.9}$$

For controlling the remaining part of the integral, [59] use the spectral gap inequality,

$$\exists \varpi > 0 \text{ s.t. } \left\|P_t f - \int f\,\mathrm{d}\mu\right\|_{L^2(\mu)} \;\le\; \mathrm{e}^{-t\varpi}\|f\|_{L^2(\mu)}. \tag{3.10}$$

Consequently, one obtains

$$\int_D^T \langle f_c, P_u f_c\rangle\, u \;\leq\; \int_D^T \|f_c\|_{L^2(\mu)}\|P_u f_c\|_{L^2(\mu)}\,\mathrm{d}u \;\leq\; \|f\|^2_{L^2(\mu)}\int_D^T \mathrm{e}^{-u\varpi}\,\mathrm{d}u \;\leq\; \frac{1}{\varpi}\mathrm{e}^{-\varpi D}\mu(\mathcal{S})\|f\|^2_\infty. \tag{3.11}$$

In order to derive variance bounds of order $T\boldsymbol{\lambda}(\mathcal{S})\mu(\mathcal{S})$, one needs to balance the upper bounds (3.9) and (3.11) by choosing D in a suitable way. Precisely, for $D := -\varpi^{-1}\log(\boldsymbol{\lambda}(\mathcal{S}))\vee 1$, (3.9) and (3.11) then imply that

$$\operatorname{Var}\left(\int_0^T f(X_s)\,\mathrm{d}s\right) \;\lesssim\; \mu(\mathcal{S})\boldsymbol{\lambda}(\mathcal{S})\Big(\log(\boldsymbol{\lambda}(\mathcal{S})) + (\log(\boldsymbol{\lambda}(\mathcal{S})))^{\frac{3-\alpha}{2}} + 1\Big)\|f\|^2_\infty.$$

Note that the *uniform* control of the supremum in (3.5) immediately allows to derive bounds of the required order. Consequently, it is not necessary to balance both error estimates carefully and to introduce a bound of integration depending on the support. Such an approach, however, is crucial for utilizing the exponential decay in (3.10).

To get an impression of the usefulness of the above result, let us discuss the relation of the local uniform transition density convergence assumption $(\mathscr{A}2)$ to more general and often conveniently verifiable stability conditions on $\boldsymbol{X}$. In [166], conditions on the characteristic function $\varphi^x_{X_t}(\lambda) := \mathbb{E}^x[\exp(\mathrm{i}\langle X_t,\lambda\rangle)]$ of X_t and the Fourier transform $\{\mathcal{F}\mu\}(\lambda) = \int_{\mathbb{R}^d}\mathrm{e}^{\mathrm{i}\langle x,\lambda\rangle}\,\mu(\mathrm{d}x)$ were formulated in the scalar setting $d=1$ that imply finiteness of the integral part away from 0 in the Castellana–Leadbetter condition (3.6). A straightforward adaption to our multivariate localized setting yields the following result, with the proof being omitted.

Lemma 3.3. *Suppose that $\boldsymbol{X}$ is V-polynomially ergodic of degree $\gamma_1 > q/(q-1)$ for some locally bounded function V and $q > 1$. If there exists $\gamma_2 > qd$ and a locally bounded function V such that*

$(\mathscr{V}1)$ $|\varphi^x_{X_t}(\lambda) - \{\mathcal{F}\mu\}(\lambda)| \leq V(x)(1+t)^{-\gamma_1}, \quad t\geq 1, x,\lambda\in\mathbb{R}^d,$

$(\mathscr{V}2)$ $|\varphi^x_{X_t}(\lambda)| \vee |\{\mathcal{F}\mu\}(\lambda)| \lesssim (1+\|\lambda\|)^{-\gamma_2}, \quad x,\lambda\in\mathbb{R}^d, t\geq 1,$

then $(\mathscr{A}2)$ is satisfied with $\Lambda = \mathbb{R}^d$, $r_{\mathcal{S}}(t) \sim \sup_{x\in\mathcal{S}} V(x)(1+t)^{-\gamma_1}$ for compacts $\mathcal{S}$.

Note that $(\mathscr{V}2)$ implies that the Fourier transforms of $P_t(x,\cdot)$ and μ are integrable and hence the Fourier inversion theorem guarantees that continuous bounded transition and invariant densities exist. Moreover, as remarked in [166], $(\mathscr{V}1)$ is fulfilled whenever $\boldsymbol{X}$ is V-polynomially ergodic with rate $\gamma_1 > 1$.

Condition $(\mathscr{V}2)$ is quite natural in a statistical estimation context since it essentially encodes a certain amount of smoothness of the transition and stationary density. However, the following simple observation demonstrates that the additional growth conditions on the characteristic function are not needed in presence of sufficiently fast total variation convergence.

Lemma 3.4. *Suppose that $\|p_1\|_\infty < \infty$ and that $\boldsymbol{X}$ is μ-a.s. V-ergodic at speed Ξ such that $V\mathbb{1}_\Lambda$ is locally bounded and $\int_0^\infty \Xi(t) < \infty$. Then, $(\mathscr{A}2)$ holds with*

$$r_{\mathcal{S}}(t) = 2C\|p_1\|_\infty \sup_{x\in\mathcal{S}\cap\Lambda} V(x)\Xi(t-1), \quad t > 1.$$

Proof. By the semigroup property of $(P_t)_{t\geq 0}$ and invariance of μ we have for any $t > 1$ and $y \in \mathbb{R}^d$ and μ-a.e. $x \in \mathbb{R}^d$,

$$\begin{aligned}|p_t(x,y) - \rho(y)| &\leq \int_{\mathbb{R}^d} p_1(z,y)|p_{t-1}(x,z) - \rho(z)|\,\mathrm{d}z \\ &\leq \|p_1\|_\infty \int_{\mathbb{R}^d} |p_{t-1}(x,z) - \rho(z)|\,\mathrm{d}z \\ &= 2\|p_1\|_\infty \|P_{t-1}(x,\cdot) - \mu\|_{\mathrm{TV}} \leq 2\|p_1\|_\infty C V(x)\Xi(t-1),\end{aligned}$$

where the equality follows from Scheffé's theorem, see [163, Lemma 2.1]. Thus for any compact set $\mathcal{S}$ and $r_{\mathcal{S}}(t) = 2C\|p_1\|_\infty\|_\infty) \sup_{x\in\mathcal{S}\cap\Lambda} V(x)\Xi(t-1)$ it follows that

$$\int_1^\infty \sup_{x\in\mathcal{S}\cap\Lambda, y\in\mathcal{S}} |p_t(x,y) - \rho(y)|\,\mathrm{d}t \leq \int_1^\infty r_{\mathcal{S}}(t)\,\mathrm{d}t \lesssim \sup_{x\in\mathcal{S}\cap\Lambda} V(x) \int_0^\infty \Xi(t) < \infty,$$

by local boundedness of $V\mathbb{1}_\Lambda$ and the convergence assumption on Ξ, which yields $(\mathscr{A}2)$. ■

Concerning the specific set of assumptions $(\mathscr{A}0)$–$(\mathscr{A}2)$, it is established with this result in Section 3.4.1 that they are satisfied, e.g., for a large class of multivariate Lévy-driven Ornstein–Uhlenbeck processess.

Recall that the stationary Markov process $\boldsymbol{X}$ is said to be β-mixing if

$$\beta(t) := \int_{\mathbb{R}^d} \|P_t(x,\cdot) - \mu(\cdot)\|_{\mathrm{TV}}\,\mu(\mathrm{d}x) \underset{t\to\infty}{\longrightarrow} 0.$$

If there exist constants $\kappa, c_\kappa > 0$ such that $\beta(t) \leq c_\kappa \mathrm{e}^{-\kappa t}$ for any $t > 0$, then $\boldsymbol{X}$ is said to be exponentially β-mixing, which is always the case for μ-a.s. V-exponentially ergodic Markov processes provided $\mu(V) < \infty$. Here, $\mu(V) < \infty$ is not a restriction since V and $\kappa > 0$ can always be chosen such that $\mu(V) < \infty$, which follows from a straightforward extension of [134, Theorem 6.14.(iii)] to the continuous time case. By the same theorem, the converse is also true, i.e., if $\boldsymbol{X}$ is exponentially β-mixing, then $\boldsymbol{X}$ is μ-a.s. V-exponentially ergodic. See also [46, Lemma 8.9] for these statements. Exponential β-mixing is formulated as assumption $(\mathscr{A}\beta)$ in the next section and will be one of the pillars of our statistical analysis for the sup-norm risk. It is therefore critical for us to understand the exact relationship between exponential β-mixing and $(\mathscr{A}2)$. To this end, as a partial converse to Lemma 3.4, we explored in Section 3.1.1 under which additional (quite natural) conditions, $(\mathscr{A}2)$ implies the exponential β-mixing property of $\boldsymbol{X}$. Our main findings, taking account of Lemma 3.4, Section 3.1.1 and the developments in Section 3.1.2, are summarized in Figure 3.1.

A clear picture is drawn, demonstrating that local uniform transition density convergence at exponential speed is intimately connected with exponential β-mixing of the process—both concepts having exponential ergodicity as the driving force behind them in most concrete applications. Both conditions $(\mathscr{A}2)$ and $(\mathscr{A}\beta)$ gain substantial additional statistical power via the smoothing assumption $(\mathscr{A}1)$, which allows obtaining tight variance bounds that yield superior estimation properties under continuous observations compared to incomplete information via sampling procedures, as will be demonstrated in Section 3.3. Moreover, the slightly more specific localized Castellana–Leadbetter condition provides the advantage of optimal estimation also in the scalar case $d = 1$ and wrt. the L^2 risk under less restrictive assumptions on the speed of convergence of the process (polynomial is sufficient) in any dimension, which justifies us studying this concept separately from exponential β-mixing.

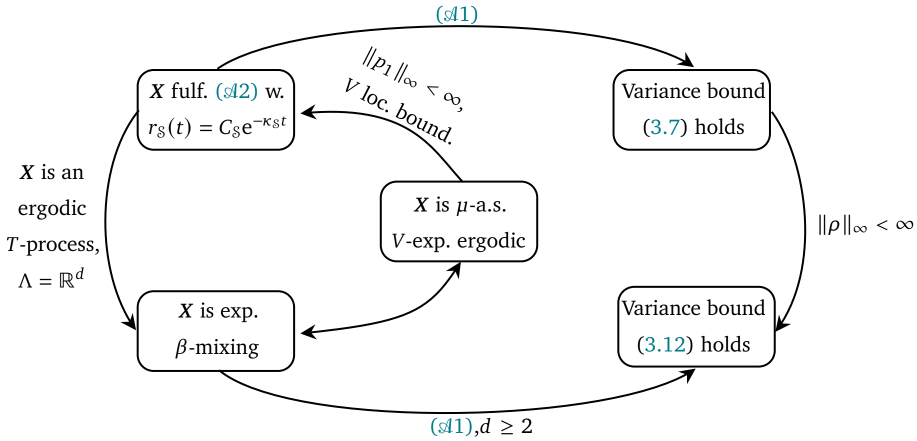

Figure 3.1: Overview of interplay between variance bound results, assumptions and stability concepts

3.1.2 *Variance analysis under exponential β-mixing*

In this subsection, we specify our study to multidimensional stochastic processes by restricting the analysis to dimension $d \geq 2$. While we further assume that the on-diagonal heat kernel bound on the transition density (3.4) from ($\mathscr{A}$1) still holds, we drop the transition density rate assumption ($\mathscr{A}$2) and instead impose exponential β-mixing of $\boldsymbol{X}$. Note that this is implied by ($\mathscr{A}$2) under suitable technical conditions on $\boldsymbol{X}$ (see Figure 3.1 and Propositions 2.10 and 2.11 in Section 2.3).

($\mathscr{A}\beta$) The process $\boldsymbol{X}$ started in the invariant measure μ is exponentially β-mixing, i.e., there exist constants $c_\kappa, \kappa > 0$, such that

$$\int \|P_t(x,\cdot) - \mu(\cdot)\|_{\mathrm{TV}}\, \mu(\mathrm{d}x) \;\leq\; c_\kappa \mathrm{e}^{-\kappa t}, \quad t \geq 0.$$

Let us emphasize that in presence of the heat kernel bound ($\mathscr{A}$1), Lemma 3.5 below shows that Assumption ($\mathscr{A}$0) is strengthened to the existence of a *bounded* invariant density since the transition density of any skeleton chain is uniformly bounded for fixed $t > 0$. That is, the following assumption is in place.

($\mathscr{A}$0+) Assumption ($\mathscr{A}$0) holds and the invariant density has a bounded version ρ, i.e., $\|\rho\|_\infty < \infty$.

Lemma 3.5. *Assume that $\boldsymbol{X}$ has an invariant distribution μ and that there is some $\Delta > 0$ such that the transition density p_Δ exists and $\sup_{x,y\in\mathbb{R}^d} p_\Delta(x,y) \leq c$ for some constant $c > 0$. Then, μ admits a bounded density.*

Proof. Let $B \in \mathcal{B}(\mathbb{R}^d)$ such that $\boldsymbol{\lambda}(B) = 0$. Then, it holds that

$$\mu(B) = \int_{\mathbb{R}^d}\int_B p_\Delta(x,y)\,\mathrm{d}y\,\mu(\mathrm{d}x) = 0,$$

which yields the existence of a Lebesgue density ρ of μ by the Radon–Nikodym theorem. Now, let $B \in \mathcal{B}(\mathbb{R}^d)$ such that $\boldsymbol{\lambda}(B) > 0$. Arguing as above and using boundedness of p_Δ, we get

$$\frac{\int_B \rho(x)\,\boldsymbol{\lambda}(\mathrm{d}x)}{\boldsymbol{\lambda}(B)} \leq c.$$

Now the Lebesgue differentiation theorem yields $\operatorname{ess\,sup} \rho \leq c$, and defining

$$\rho_b(x) = \rho(x)\mathbf{1}_{[0,c]}(\rho(x)), \quad x \in \mathbb{R}^d,$$

we have $\rho = \rho_b$ almost everywhere and $\rho_b \leq c$. ■

The next result gives a tight variance bound on the integral $\int_0^T f(X_t)\,\mathrm{d}t$ under β-mixing. Its effectiveness for sup-norm estimation of general Markov processes will be demonstrated in Section 3.3. Note in particular that, using boundedness of ρ under $(\mathscr{A}0)$ and $(\mathscr{A}1)$, the same rate can be obtained under $(\mathscr{A}2)$ from Proposition 3.1. Recall the definition of $\psi_d\colon (0,\mathrm{e}) \to \mathbb{R}_+$ in (3.7).

PROPOSITION 3.6. *Grant assumptions $(\mathscr{A}1)$ and $(\mathscr{A}\beta)$, and let f be a bounded function with compact support $\mathcal{S}$ fulfilling $\boldsymbol{\lambda}(\mathcal{S}) < 1$. Then, for any $d \geq 2$, there exists a constant $C > 0$ not depending on f such that, for any $T > 0$,*

$$\operatorname{Var}\left(\int_0^T f(X_t)\,\mathrm{d}t\right) \leq CT\|f\|_\infty^2\|\rho\|_\infty\boldsymbol{\lambda}^2(\mathcal{S})\psi_d^2(\boldsymbol{\lambda}(\mathcal{S})). \tag{3.12}$$

Proof. Let $0 < \delta < 1 \leq D$. Analogously to the proof of Proposition 3.1, one can compute that

$$\begin{aligned}
&\operatorname{Var}\left(\int_0^T f(X_t)\,\mathrm{d}t\right)\\
&= 2\int_0^T (T-v)\iint_{\mathbb{R}^{d\times d}} f(x)f(y)(p_v(x,y)-\rho(y))\,\mu(\mathrm{d}x)\,\mathrm{d}y\,\mathrm{d}v\\
&\leq 2T\|f\|_\infty^2\left(\int_0^D\iint_{\mathcal{S}^2} p_v(x,y)\,\mu(\mathrm{d}x)\,\mathrm{d}y\,\mathrm{d}v + \int_D^T\iint_{\mathcal{S}^2}(p_v(x,y)-\rho(y))\,\mu(\mathrm{d}x)\,\mathrm{d}y\,\mathrm{d}v\right)\\
&= 2T\|f\|_\infty^2(\mathcal{I}_\delta + \mathcal{I}_D + \mathcal{I}_T),
\end{aligned}$$

where $\mathcal{I}_\delta := \int_0^\delta\iint_{\mathcal{S}^2} p_v(x,y)\mu(\mathrm{d}x)\,\mathrm{d}y\,\mathrm{d}v$, $\mathcal{I}_D := \int_\delta^D\iint_{\mathcal{S}^2} p_v(x,y)\mu(\mathrm{d}x)\,\mathrm{d}y\,\mathrm{d}v$ and

$$\mathcal{I}_T \;:=\; \int_D^T\int_{\mathcal{S}}(P_v(x,\mathcal{S})-\mu(\mathcal{S}))\,\mu(\mathrm{d}x)\,\mathrm{d}v.$$

As before (see (3.8)) and under our additional assumption that ρ is bounded, it holds

$$\mathcal{I}_\delta \leq \mu(\mathcal{S})\delta \leq \|\rho\|_\infty\boldsymbol{\lambda}(\mathcal{S})\delta. \tag{3.13}$$

Furthermore, exploiting the mixing property of $\boldsymbol{X}$,

$$\mathcal{I}_T \;\leq\; \int_D^T\int \|P_v(x,\cdot)-\mu(\cdot)\|_{\mathrm{TV}}\mu(\mathrm{d}x)\,\mathrm{d}v \;\leq\; c_\kappa\int_D^T \mathrm{e}^{-\kappa v}\,\mathrm{d}v \;\leq\; \frac{c_\kappa}{\kappa}\mathrm{e}^{-\kappa D}\mathbb{1}_{(D,\infty)}(T). \tag{3.14}$$

By assumption ($\mathscr{A}$1), $p_v(x,y) \le c_2 v^{-d/2}$, for $0 < t \le 1$. Hence, we have $p_{1/2}(x,y) \le c_2 2^{d/2} =: c_p$ which implies

$$p_t(x,y) = \int p_{t-1/2}(x,z) p_{1/2}(z,y)\,\mathrm{d}z \le c_p,$$

for all $t > 1/2$. Since $\delta < 1 \le D$, it follows

$$\int_\delta^D p_v(x,y)\,\mathrm{d}v \le c_2 \int_\delta^1 v^{-d/2}\,\mathrm{d}v + c_p D\mathbf{1}_{(1,\infty)}(D) \le c_{\delta,D}\Big(\int_\delta^1 v^{-d/2}\,\mathrm{d}v + D\mathbf{1}_{(1,\infty)}(D)\Big)$$

for $c_{\delta,D} \coloneqq c_2 + c_p$. For $d \ge 3$, this implies

$$\begin{aligned}\int_\delta^D p_v(x,y)\,\mathrm{d}v &\le c_{\delta,D}\Big(\int_\delta^1 v^{-d/2}\,\mathrm{d}v + D\mathbf{1}_{(1,\infty)}(D)\Big)\\ &\le c_{\delta,D}\Big((d/2-1)^{-1}\delta^{1-d/2} + D\mathbf{1}_{(1,\infty)}(D)\Big)\\ &\le c'_{\delta,D}\Big(\delta^{1-d/2} + D\mathbf{1}_{(1,\infty)}(D)\Big),\end{aligned}\tag{3.15}$$

where $c'_{\delta,D} \coloneqq 2c_{\delta,D}$. Letting $\delta = \boldsymbol{\lambda}(\mathcal{S})^{2/d}$, $D = (1 \vee -\frac{2}{\kappa}\log(\boldsymbol{\lambda}(\mathcal{S}))) \wedge T$, (3.15) and $\boldsymbol{\lambda}(\mathcal{S}) < 1$ imply

$$\begin{aligned}\int_\delta^D p_v(x,y)\,\mathrm{d}v &\le c'_{\delta,D}\Big(\boldsymbol{\lambda}(\mathcal{S})^{2/d-1} + \tfrac{2}{\kappa}\log(\boldsymbol{\lambda}(\mathcal{S})^{-1})\Big) \le c'_{\delta,D}\Big(\boldsymbol{\lambda}(\mathcal{S})^{2/d-1} + \tfrac{2}{\kappa(1-2/d)}\boldsymbol{\lambda}(\mathcal{S})^{2/d-1}\Big)\\ &\le c''_{\delta,D}\boldsymbol{\lambda}(\mathcal{S})^{2/d-1}, \qquad \text{for } c''_{\delta,D} \coloneqq c'_{\delta,D}(1 + \tfrac{2}{\kappa(1-2/d)}),\end{aligned}$$

where we have used the well-known inequality $\log(x) \le nx^{1/n}$, $x, n > 0$. Using Fubini's theorem, this directly implies

$$\mathcal{I}_D = \int_\delta^D \iint_{\mathcal{S}^2} p_v(x,y)\mu(\mathrm{d}x)\,\mathrm{d}y\,\mathrm{d}v \le c''_{\delta,D}\mu(\mathcal{S})\boldsymbol{\lambda}(\mathcal{S})^{2/d} \le c''_{\delta,D}\|\rho\|_\infty \boldsymbol{\lambda}(\mathcal{S})^{2/d+1}\tag{3.16}$$

for $d \ge 3$. Noting that our choice of δ and D implies by (3.13) and (3.14) that

$$\mathcal{I}_\delta \le \|\rho\|_\infty \boldsymbol{\lambda}(\mathcal{S})^{2/d+1} \quad \text{and} \quad \mathcal{I}_T \le \frac{c_\kappa}{\kappa}\boldsymbol{\lambda}(\mathcal{S})^2 \le \frac{c_\kappa}{\kappa}\boldsymbol{\lambda}(\mathcal{S})^{2/d+1},$$

(3.12) follows for any $d \ge 3$ by combining these estimates with (3.16). The case $d = 2$ is treated by similar arguments. ∎

Notation. Throughout the sequel, we denote by $\boldsymbol{\Sigma}$ the class of non-explosive, exponentially β-mixing Borel right Markov processes $\boldsymbol{X}$ such that assumptions ($\mathscr{A}$0) and ($\mathscr{A}$1) hold (and hence ($\mathscr{A}$0+) is in place, i.e., the invariant density ρ is bounded). Moreover, in dimension $d = 1$ we assume that ($\mathscr{A}$2) is in place with a rate function $r_\mathcal{S}$ which is monotone wrt the compact sets $\mathcal{S}$ in the sense that

$$\mathcal{S}_1 \subset \mathcal{S}_2 \implies c_{\mathcal{S}_1} = \int_1^\infty r_{\mathcal{S}_1}(t)\,\mathrm{d}t \le \int_1^\infty r_{\mathcal{S}_2}(t)\,\mathrm{d}t = c_{\mathcal{S}_2} < \infty.\tag{3.17}$$

Alternatively, if we do not want to restrict to exponentially β-mixing processes, consider the class of processes $\boldsymbol{\Theta}$ consisting of d-dimensional non-explosive Borel right processes such that ($\mathscr{A}$0)–($\mathscr{A}$2) hold, where again the constants $c_\mathcal{S}$ appearing in ($\mathscr{A}$2) satisfy (3.17). Note that if $\widetilde{\boldsymbol{\Theta}}$ is the restriction of $\boldsymbol{\Theta}$ containing the class of processes $\boldsymbol{X}$ satisfying the assumptions of Proposition 2.10 or Proposition 2.11, then $\widetilde{\boldsymbol{\Theta}} \subset \boldsymbol{\Sigma}$.

3.2 Uniform moment bounds and deviation inequalities

Uniform moment bounds for integral functionals of $\boldsymbol{X}$ are the subject of this section. These are intimately connected with Bernstein-type tail inequalities, which due to their crucial importance for many probabilistic and statistical applications—such as the derivation of limit theorems or upper bound statements for nonparametric estimation procedures—have been excessively studied in the literature (see Section 1.1 of [82] for an overview). Both a Lyapunov function method and a functional inequalities approach can be used for deriving results on the concentration behavior of additive functionals of $\boldsymbol{X}$. [40] establish non-asymptotic deviation bounds for

$$\mathbb{P}\left(\left|\int_0^t f(X_s)\,\mathrm{d}s - \int f\,\mathrm{d}\mu\right| \geq r\right), \qquad f \in L^1(\mu),$$

using different moment assumptions for f and regularity conditions for μ, "regularity" referring to the condition that μ may satisfy various functional inequalities (F-Sobolev, generalized Poincaré, etc.). In a symmetric Markovian setting and assuming a spectral gap, [116] uses Kato's perturbation theory for proving Bernstein-type concentration inequalities for empirical means of the form $\int_0^t f(X_s)\,\mathrm{d}s$, the upper bound depending on the asymptotic variance of f. Amongst other methods, [82] exploit both a Lyapunov function method and a functional inequalities approach for extending Lezaud's result to inequalities for possibly unbounded f. Going beyond the symmetric case, Lyapunov-type conditions can also be used for verifying exponential mixing properties, paving the way to generalizing concentration results based on independent observations to the dependent case. For corresponding results for discrete random (Markov) sequences under different mixing or ergodicity assumptions, we refer to [1, 2, 23, 53, 64, 114, 126, 145]

3.2.1 *General framework*

Our main focus in this subsection is on deriving corresponding uniform moment inequalities of empirical processes and (in Section 3.3) proving their efficiency in a concrete statistical application, using merely the previously introduced assumptions (in particular, the β-mixing property), and without introducing any additional conditions on the process. We emphasize that for this section no assumption on the existence of transition or invariant densities is needed, but that we only work within an ergodic β-mixing framework. Moreover, the results are established for β-mixing Markov processes with arbitrary topological state space $\mathcal{X}$, not necessarily equal to $\mathbb{R}^d$, and general mixing rate. That is, we suppose in this section that

$$\beta(t) = \int_{\mathcal{X}} \|P_t(x,\cdot) - \mu\|_{\mathrm{TV}}\,\mu(\mathrm{d}x) \leq \Xi(t),$$

for some rate function $\Xi(t)$ decreasing to 0 as $t \longrightarrow \infty$.

We aim to prove moment bounds for suprema of the form

$$\sup_{g\in\mathcal{G}} |\mathbb{G}_t(g)| =: \|\mathbb{G}_t\|_{\mathcal{G}}, \qquad \text{for } \mathbb{G}_t(g) := \frac{1}{\sqrt{t}}\int_0^t g(X_s)\,\mathrm{d}s,$$

where the supremum is taken over entire (possibly infinite-dimensional) function classes $\mathcal{G} \subset \mathcal{B}_b(\mathcal{X})$ of μ-centered measurable bounded functions on $\mathcal{X}$. Such results are indispensable tools, e.g., for the analysis of nonparametric adaptive estimation procedures and for applications in statistical learning theory. Similarly to [20] and [68], we apply the generic chaining device for the derivation of our result. The basic strategy of the proof is splitting the integral into blocks of length m_t, construct an independent Berbee coupling based on the β-mixing property as described in Viennet [170], and then use the classical Bernstein inequality for i.i.d. random variables for the coupled integral blocks to drive the chaining procedure from [68]. The use of Berbee's coupling lemma is a well-established method for studying empirical processes of discrete β-mixing sequences, see [143, Chapter 8], and has recently been employed in [12] for establishing L^2 oracle bounds for an adaptive estimator of the invariant density of a class of exponentially β-mixing Lévy driven SDEs. Our final moment bound on the supremum of the process $\mathbb{G}_t$ is stated in terms of entropy integrals of the indexing function class $\mathcal{G}$. In many applications, the corresponding assumption is straightforward to verify. For any given $\varepsilon > 0$, denote by $\mathcal{N}(\varepsilon, \mathcal{G}, d)$ the covering number of $\mathcal{G}$, i.e., the smallest number of balls of d-radius ε needed to cover $\mathcal{G}$. Furthermore, given $f, g \in \mathcal{G}$, let $d_\infty(f,g) := \|f-g\|_\infty$ and

$$d^2_{\mathbb{G},t}(f,g) := \sigma_t^2(f-g), \text{ where } \sigma_t^2(f) := \operatorname{Var}\left(\frac{1}{\sqrt{t}}\int_0^t f(X_s)\,\mathrm{d}s\right).$$

THEOREM 3.7. *Suppose that $\boldsymbol{X}$ is β-mixing with rate function $\Xi(t)$. Let $\mathcal{G}$ be a countable class of bounded real-valued functions with $\mu(g) = 0$ and let $m_t \in (0, t/4]$. Then, there exist $\tau \in [m_t, 2m_t]$ and constants $\widetilde{C}_1, \widetilde{C}_2 > 0$ such that, for any $1 \le p < \infty$,*

$$\begin{aligned}\left(\mathbb{E}\left[\|\mathbb{G}_t\|_{\mathcal{G}}^p\right]\right)^{1/p} &\le \widetilde{C}_1 \int_0^\infty \log \mathcal{N}\left(u, \mathcal{G}, \tfrac{2m_t}{\sqrt{t}} d_\infty\right) \mathrm{d}u + \widetilde{C}_2 \int_0^\infty \sqrt{\log \mathcal{N}(u, \mathcal{G}, d_{\mathbb{G},\tau})}\,\mathrm{d}u \\ &\quad + 4 \sup_{g \in \mathcal{G}} \left(\frac{2m_t}{\sqrt{t}} \|g\|_\infty \widetilde{c}_1 p + \|g\|_{\mathbb{G},\tau} \widetilde{c}_2 \sqrt{p} + \frac{1}{2}\|g\|_\infty \sqrt{t}\,\Xi(m_t)^{1/p}\right),\end{aligned} \tag{3.18}$$

for positive constants $\widetilde{c}_1, \widetilde{c}_2$ defined in (3.21).

Proof. We start by splitting the process $(X_s)_{0 \le s \le t}$ with Borel state space $\mathcal{X}$ into $2n_t$ parts of length m_t, where $t = 2n_t m_t$, $n_t \in \mathbb{N}$, $m_t \in \mathbb{R}_+$. More precisely, for $j \in \{1, \dots, n_t\}$, define the processes

$$X^{j,1} := (X_s)_{s \in [2(j-1)m_t, (2j-1)m_t]}, \quad X^{j,2} := (X_s)_{s \in [(2j-1)m_t, 2jm_t]}.$$

Since $\boldsymbol{X}$ is a stationary Markov process, the β-mixing assumption is equivalent to

$$\Xi(s) \ge \int_{\mathbb{R}^d} \|P_s(x,\cdot) - \mu\|_{\mathrm{TV}}\,\mu(\mathrm{d}x) = \mathbb{E}\left[\|\mathbb{P}(\cdot|\mathcal{F}_0) - \mathbb{P}\|_{\mathrm{TV}|\overline{\mathcal{F}}_s}\right] = \mathbb{E}\left[\|\mathbb{P}(\cdot|\mathcal{F}_t) - \mathbb{P}\|_{\mathrm{TV}|\overline{\mathcal{F}}_{t+s}}\right],$$

for any $s, t > 0$, see Proposition 1 in [62]. Here, $(\mathcal{F}_t = \sigma(X_s, s \le t))_{t \ge 0}$ denotes the natural filtration of $\boldsymbol{X}$, $(\overline{\mathcal{F}}_t = \sigma(X_s, s \ge t))_{t \ge 0}$ the filtration of the future of $\boldsymbol{X}$ and, for a signed measure μ and a sub-σ-algebra $\mathcal{A}$ on a measure space $(\Omega, \mathcal{F})$, $\|\mu\|_{\mathrm{TV}|\mathcal{A}}$ denotes the total variation norm of μ restricted to $\mathcal{A}$. As demonstrated in [173, Lemma 1.4],

$$\mathbb{E}\left[\|\mathbb{P}(\cdot|\mathcal{F}_t) - \mathbb{P}\|_{\mathrm{TV}|\overline{\mathcal{F}}_{t+s}}\right] = \beta(\mathcal{F}_t, \overline{\mathcal{F}}_{t+s}),$$

where for two sub-σ-algebras $\mathcal{A}, \mathcal{B} \subset \mathcal{G}$ and a probability measure $\mathbb{P}$ on $(\Omega, \mathcal{G})$, the classical β-mixing coefficient $\beta(\mathcal{A}, \mathcal{B})$ is given by

$$\beta(\mathcal{A}, \mathcal{B}) = \sup_{C \in \mathcal{A} \otimes \mathcal{B}} \left|\mathbb{P}|_{\mathcal{A} \otimes \mathcal{B}}(C) - \mathbb{P}|_{\mathcal{A}} \otimes \mathbb{P}|_{\mathcal{B}}(C)\right|.$$

Here, $\mathbb{P}|_{\mathcal{A}\otimes\mathcal{B}}$ is the restriction to $(\Omega\times\Omega, \mathcal{A}\otimes\mathcal{B})$ of the image measure of $\mathbb{P}$ under the canonical injection $\iota(\omega)=(\omega,\omega)$. Clearly, if $\mathcal{A}_1\subset\mathcal{A}_2$, we have $\beta(\mathcal{A}_1,\mathcal{B})\leq\beta(\mathcal{A}_2,\mathcal{B})$. Observe that $X^{j,1}$, as a mapping from Ω to $\mathcal{X}^{[2(j-1)m_t,(2j-1)m_t]}$, is both $\mathcal{F}_{(2j-1)m_t}$-measurable and $\overline{\mathcal{F}}_{2(j-1)m_t}$-measurable. It now follows from the above discussion for $j,k\in\{1,\dots,n_t\}$, $j<k$, that

$$\begin{aligned}\beta(X^{j,1},X^{k,1}) \coloneqq \beta(\sigma(X^{j,1}),\sigma(X^{k,1})) &\leq \beta(\mathcal{F}_{(2j-1)m_t},\overline{\mathcal{F}}_{2(k-1)m_t})\\ &\leq \Xi((2(k-j)-1)m_t)\leq\Xi((k-j)m_t).\end{aligned}$$

In the same way, we obtain $\beta(X^{j,2},X^{k,2})\leq\Xi((k-j)m_t)$. Arguing as in the proof of Proposition 5.1 of [170], we can then construct a process $(\widehat{X}_s)_{0\leq s\leq t}$ by Berbee's coupling method, such that for $k=1,2$,

1. $X^{j,k}\overset{(\mathrm{d})}{=}\widehat{X}^{j,k}$, for all $j\in\{1,\dots,n_t\}$,
2. $\mathbb{P}(X^{j,k}\neq\widehat{X}^{j,k})\leq\Xi(m_t)$ for all $j\in\{1,\dots,n_t\}$,
3. $\widehat{X}^{1,k},\dots,\widehat{X}^{n_t,k}$ are independent,

where $\widehat{X}^{j,k}$ is defined analogously to $X^{j,k}$ for $j\in\{1,\dots,n_t\}$ and $k=1,2$. In order to ease the notation, define for $j\in\{1,\dots,n_t\}$

$$I_g(X^{j,1})\coloneqq\int_{2(j-1)m_t}^{(2j-1)m_t}g(X_s)\,\mathrm{d}s,\quad I_g(X^{j,2})\coloneqq\int_{(2j-1)m_t}^{2jm_t}g(X_s)\,\mathrm{d}s,$$

and, analogously, define $I_g(\widehat{X}^{j,k})$ for $k=1,2$, $j\in\{1,\dots,n_t\}$. Fix $p\geq1$. Then,

$$\begin{aligned}&\left(\mathbb{E}\left[\|\mathbb{G}_t\|_{\mathcal{G}}^p\right]\right)^{1/p}\\ &\leq\left(\mathbb{E}\left[\sup_{g\in\mathcal{G}}\left|\frac{1}{\sqrt{t}}\int_0^t g(\widehat{X}_s)\,\mathrm{d}s\right|^p\right]\right)^{1/p}+\left(\mathbb{E}\left[\sup_{g\in\mathcal{G}}\left|\frac{1}{\sqrt{t}}\int_0^t(g(X_s)-g(\widehat{X}_s))\,\mathrm{d}s\right|^p\right]\right)^{1/p}\\ &=\left(\mathbb{E}\left[\sup_{g\in\mathcal{G}}\left|\frac{1}{\sqrt{t}}\sum_{k=1}^2\sum_{j=1}^{n_t}I_g(\widehat{X}^{j,k})\right|^p\right]\right)^{1/p}+\left(\mathbb{E}\left[\sup_{g\in\mathcal{G}}\left|\frac{1}{\sqrt{t}}\sum_{k=1}^2\sum_{j=1}^{n_t}(I_g(X^{j,k})-I_g(\widehat{X}^{j,k}))\right|^p\right]\right)^{1/p}.\end{aligned}\tag{3.19}$$

The classical Bernstein inequality implies for $u>0$ that

$$\mathbb{P}\left(\left|\frac{1}{\sqrt{t}}\sum_{j=1}^{n_t}I_g(\widehat{X}^{j,k})\right|>\sqrt{\frac{2n_t\mathrm{Var}\left(\int_0^{m_t}g(X_s)\,\mathrm{d}s\right)u}{t}}+\frac{m_t\|g\|_\infty u}{\sqrt{t}}\right)\leq 2\mathrm{e}^{-u},$$

which in combination with $2n_t/t=1/m_t$ yields

$$\mathbb{P}\left(\left|\frac{1}{\sqrt{t}}\sum_{j=1}^{n_t}I_g(\widehat{X}^{j,k})\right|>u\right)\leq 2\exp\left(-\frac{u^2}{2\left(\mathrm{Var}\left(\frac{1}{\sqrt{m_t}}\int_0^{m_t}g(X_s)\,\mathrm{d}s\right)+\frac{m_t}{\sqrt{t}}\|g\|_\infty u\right)}\right),\quad u>0,\tag{3.20}$$

see Theorem 2.10 and Corollary 2.11 in [35]. Consequently, denoting

$$\widetilde{c}_1 := 2\mathrm{e}^{1/(2\mathrm{e})}\sqrt{2\pi}\mathrm{e}^{-11/12}, \quad \widetilde{c}_2 := 2(2\mathrm{e})^{-1/2}\mathrm{e}^{1/(2\mathrm{e})}\sqrt{\pi}\mathrm{e}^{1/6}, \tag{3.21}$$

Lemma A.2 in [68] gives, for $k \in \{1,2\}$,

$$\left(\mathbb{E}\left[\left|\frac{1}{\sqrt{t}}\sum_{j=1}^{n_t} I_g(\widehat{X}^{j,k})\right|^p\right]\right)^{1/p} \leq \|g\|_\infty \frac{m_t}{\sqrt{t}}\widetilde{c}_1 p + \sqrt{\operatorname{Var}\left(\frac{1}{\sqrt{m_t}}\int_0^{m_t} g(X_s)\,\mathrm{d}s\right)}\widetilde{c}_2\sqrt{p}, \tag{3.22}$$

where we used again $2n_t/t = 1/m_t$. In addition, Theorem 3.5 in [68] implies that there exist positive constants $\widetilde{C}_1, \widetilde{C}_2$ such that

$$\begin{aligned}\left(\mathbb{E}\left[\sup_{g\in\mathcal{G}}\left|\frac{1}{\sqrt{t}}\sum_{j=1}^{n_t} I_g(\widehat{X}^{j,k})\right|^p\right]\right)^{1/p} &\leq \frac{\widetilde{C}_1}{2}\int_0^\infty \log\mathcal{N}\big(u,\mathcal{G},\tfrac{m_t}{\sqrt{t}}d_\infty\big)\,\mathrm{d}u + \frac{\widetilde{C}_2}{2}\int_0^\infty \sqrt{\log\mathcal{N}(u,\mathcal{G},d_{\mathbb{G},m_t})}\,\mathrm{d}u \\ &\quad + 2\sup_{g\in\mathcal{G}}\left(\mathbb{E}\left[\left|\frac{1}{\sqrt{t}}\sum_{j=1}^{n_t} I_g(\widehat{X}^{j,k})\right|^p\right]\right)^{1/p}.\end{aligned} \tag{3.23}$$

Here, we bounded the γ_α-functionals appearing in the original statement of the theorem by the corresponding entropy integrals. Note further that the last term on the rhs of (3.19) is upper bounded by

$$\begin{aligned}\left(\mathbb{E}\left[\sup_{g\in\mathcal{G}}\left|\frac{1}{\sqrt{t}}\sum_{k=1}^{2}\sum_{j=1}^{n_t}\Big(I_g(X^{j,k}) - I_g(\widehat{X}^{j,k})\Big)\cdot\mathbf{1}_{X^{j,k}\neq\widehat{X}^{j,k}}\right|^p\right]\right)^{1/p} &\leq \frac{4n_t m_t}{\sqrt{t}}\sup_{g\in\mathcal{G}}\|g\|_\infty\big(\mathbb{P}\big(X^{j,k}\neq\widehat{X}^{j,k}\big)\big)^{1/p} \\ &\leq 2\sup_{g\in\mathcal{G}}\|g\|_\infty\sqrt{t}\,\Xi(m_t)^{1/p}.\end{aligned} \tag{3.24}$$

Plugging the upper bounds (3.22), (3.23) and (3.24) into (3.19) yields

$$\begin{aligned}\left(\mathbb{E}\left[\sup_{g\in\mathcal{G}}|\mathbb{G}_t(g)|^p\right]\right)^{1/p} &\leq \widetilde{C}_1\int_0^\infty \log\mathcal{N}\big(u,\mathcal{G},\tfrac{m_t}{\sqrt{t}}d_\infty\big)\,\mathrm{d}u + \widetilde{C}_2\int_0^\infty\sqrt{\log\mathcal{N}(u,\mathcal{G},d_{\mathbb{G},m_t})}\,\mathrm{d}u \\ &\quad + 4\sup_{g\in\mathcal{G}}\left(\frac{m_t}{\sqrt{t}}\|g\|_\infty\widetilde{c}_1 p + \|g\|_{\mathbb{G},m_t}\widetilde{c}_2\sqrt{p} + \frac{1}{2}\|g\|_\infty\sqrt{t}\,\Xi(m_t)^{1/p}\right).\end{aligned} \tag{3.25}$$

For general $m_t \in (0, \frac{t}{4}]$, let $\widetilde{n}_t = \lfloor\frac{t}{2m_t}\rfloor$, where $\lfloor x\rfloor$ denotes the largest integer smaller or equal to $x \geq 1$. Then, for $\widetilde{m}_t := \frac{t}{2\widetilde{n}_t}$, we have $m_t \leq \widetilde{m}_t$, and from $\widetilde{n}_t \geq \frac{t}{2m_t} - 1 = \frac{t-2m_t}{2m_t}$ and $m_t \leq \frac{t}{4}$, we get

$$\widetilde{m}_t = \frac{t}{2\widetilde{n}_t} \leq \frac{tm_t}{t-2m_t} \leq 2m_t.$$

Since $\widetilde{n}_t \in \mathbb{N}$, (3.25) holds with $\tau = \widetilde{m}_t \in [m_t, 2m_t]$ and m_t being replaced by $\widetilde{m}_t$, and combining this with the computations above yields

$$\left(\mathbb{E}\left[\sup_{g\in\mathcal{G}}|\mathbb{G}_t(g)|^p\right]\right)^{1/p} \leq \widetilde{C}_1\int_0^\infty \log\mathcal{N}\big(u,\mathcal{G},\tfrac{\tau}{\sqrt{t}}d_\infty\big)\,\mathrm{d}u + \widetilde{C}_2\int_0^\infty\sqrt{\log\mathcal{N}(u,\mathcal{G},d_{\mathbb{G},\tau})}\,\mathrm{d}u$$

$$
\begin{aligned}
&+ 4 \sup_{g\in\mathcal{G}} \Big(\frac{\tau}{\sqrt{t}} \|g\|_\infty \widetilde{c}_1 p + \|g\|_{\mathbb{G},\tau} \widetilde{c}_2 \sqrt{p} + \frac{1}{2}\|g\|_\infty \sqrt{t}\,\Xi(\tau)^{1/p} \Big) \\
\le\ & \widetilde{C}_1 \int_0^\infty \log \mathcal{N}\big(u, \mathcal{G}, \tfrac{2m_t}{\sqrt{t}} d_\infty\big)\,\mathrm{d}u + \widetilde{C}_2 \int_0^\infty \sqrt{\log \mathcal{N}(u, \mathcal{G}, d_{\mathbb{G},\tau})}\,\mathrm{d}u \\
&+ 4 \sup_{g\in\mathcal{G}} \Big(\frac{2m_t}{\sqrt{t}} \|g\|_\infty \widetilde{c}_1 p + \|g\|_{\mathbb{G},\tau} \widetilde{c}_2 \sqrt{p} + \frac{1}{2}\|g\|_\infty \sqrt{t}\,\Xi(m_t)^{1/p} \Big),
\end{aligned}
$$

which completes the proof. ∎

Consider $p = 1$ and the specific choice of $m_t = \kappa^{-1} \log t$ in case of exponential β-mixing rate $\Xi(t) = c_\kappa \exp(-\kappa t)$. Then, the above result implies that

$$
\mathbb{E}[\|\mathbb{G}_t\|_{\mathcal{G}}] \lesssim \int_0^\infty \log \mathcal{N}\big(u, \mathcal{G}, \tfrac{\log t}{\sqrt{t}} d_\infty\big)\,\mathrm{d}u + \int_0^\infty \sqrt{\log \mathcal{N}(u, \mathcal{G}, d_{\mathbb{G},\tau})}\,\mathrm{d}u + \sup_{g\in\mathcal{G}} \Big(\frac{\log t}{\sqrt{t}} \|g\|_\infty + \|g\|_{\mathbb{G},\tau} \Big). \tag{3.26}
$$

If we considered the related discrete time problem of finding uniform moment bounds for additive functionals $\frac{1}{\sqrt{n}} \sum_{k=0}^n g(X_k)$ of a Markov chain $(X_n)_{n\in\mathbb{N}_0}$ and assumed exponential ergodicity of the chain, using the state of the art Bernstein inequality given in [1, Theorem 6] (see also [114]) for the generic chaining procedure would yield an analogous result with an asymptotic version of the variance norm. In particular, the log-scaling of the sup-norm is also present in the discrete time case as a consequence of exponential ergodicity, whereas in the i.i.d. case this factor would disappear. Our direct coupling approach therefore yields optimal uniform moment bounds and makes the contribution of the mixing term transparent, which also naturally paves the way for studying nonparametric implications of sub-exponential mixing rates for sup-norm estimation problems in continuous time in future work.

To get a first taste of the consequences of Theorem 3.7 consider the trivial situation where $\mathcal{G}$ is a singleton set to study rates for the L^p-version of von Neumann's ergodic theorem[1] for continuous time ergodic Markov processes which states that for $g \in L^p(\mu)$,

$$
\frac{1}{T}\int_0^T g(X_s)\,\mathrm{d}s \underset{t\to\infty}{\longrightarrow} \mu(g), \quad \text{in } L^p(\mathbb{P}).
$$

Note that indeed, β-mixing implies strong mixing and hence ergodicity of $\boldsymbol{X}$.

COROLLARY 3.8. *Suppose that $\boldsymbol{X}$ is exponentially β-mixing. Then, there exists a constant $C > 0$ such that, for any $T > 0$, $1 \le p < \infty$ and any bounded, measurable function g,*

$$
\Big\| \frac{1}{T}\int_0^T g(X_t)\,\mathrm{d}t - \mu(g) \Big\|_{L^p(\mathbb{P})} \le Cp\|g\|_\infty \frac{1}{\sqrt{T}}.
$$

If $\boldsymbol{X}$ is polynomially mixing of degree $\alpha > 1$, i.e., $\Xi(t) \lesssim t^{-\alpha}$, then for any $p \ge 1$ and $T \ge 4^{(\alpha+p)/\alpha}$ we have

$$
\Big\| \frac{1}{T}\int_0^T g(X_t)\,\mathrm{d}t - \mu(g) \Big\|_{L^p(\mathbb{P})} \lesssim \|g\|_\infty T^{-\left(\frac{1}{2}\wedge\frac{\alpha}{\alpha+p}\right)}.
$$

[1]Not referring to the L^p-statement as Birkhoff's ergodic theorem is not without reason, see [177].

Proof. In case of exponential β-mixing we obtain, similarly to the proof of Proposition 3.6, for any $t > 0$,

$$\|g\|_{\mathbb{G},t}^2 = \frac{1}{t}\mathrm{Var}\Big(\int_0^t g(X_s)\,\mathrm{d}s\Big) \le 2\|g\|_\infty^2 \int_0^t \int \|P_s(x,\cdot) - \mu\|_{\mathrm{TV}}\,\mu(\mathrm{d}x)\,\mathrm{d}s \le 2\|g\|_\infty^2 \frac{c_\kappa}{\kappa}.$$

Choosing $m_T = \sqrt{T}$ and plugging this into (3.18) therefore yields the assertion for the exponential mixing case. For the α-polynomial case we obtain the assertion similarly by the minimizing choice $m_T = T^{p/(\alpha+p)}$, where $T \ge 4^{(\alpha+p)/\alpha}$ guarantees that $m_T \le T/4$ and the assumption $\alpha > 1$ is needed for uniform boundedness of $\|g\|_{\mathbb{G},t}^2$ in t. ∎

3.2.2 *Deviation inequalities for suprema of empirical Markov processes*

Theorem 3.7 provides a foundation for the derivation of deviation inequalities, as they are needed, for example, for bounding the sup-norm risk of estimators and for the convergence analysis of adaptive estimation procedures. We will focus on the question of invariant density estimation for Borel right Markov processes, introduced and discussed in Section 3.1. Recall the definition of $\mathbf{\Sigma}$ and $\mathbf{\Theta}$ at the end of that section. Given the observation $(X_s)_{0\le s\le T}$, a natural kernel estimator for the invariant density ρ on a domain D of a Markov process $\boldsymbol{X} \in \mathbf{\Sigma} \cup \mathbf{\Theta}$ is given by

$$\widehat{\rho}_{h,T}(x) = \frac{1}{T}\int_0^T K_h(x - X_s)\,\mathrm{d}s, \quad x \in \mathbb{R}^d, \quad \text{where } K_h(\cdot) := h^{-d}K(\cdot/h), \quad h > 0, \tag{3.27}$$

for some smooth, Lipschitz continuous kernel function $K\colon \mathbb{R}^d \to \mathbb{R}$ with compact support $[-1/2, 1/2]^d$. The knowledge of the invariant density is not only a question of its own interest, but is also needed, among other things, for the implementation of drift estimation procedures or data-driven methods of stochastic control. Furthermore, this specific estimation problem can be regarded as an acid test for the quality of the statistical analysis: It is known that the invariant density of (possibly multidimensional) diffusion processes can be estimated with a faster convergence rate than is feasible in the classical discrete i.i.d. or weak dependency context. However, these superior convergence rates can only be verified with sufficiently tight estimates in the proof of the upper bound, more precisely, for the stochastic error part. Indeed, denoting $\mathbb{H}_{h,T}(x) := \widehat{\rho}_{h,T}(x) - \mathbb{E}[\widehat{\rho}_{h,T}(x)]$, we have the decomposition

$$\widehat{\rho}_{h,T}(x) - \rho(x) = \mathbb{H}_{h,T}(x) + (\rho * K_h - \rho)(x). \tag{3.28}$$

While the bias part is bounded using standard arguments, tight upper bounds on (the supremum of) the stochastic error require specific probabilistic tools. For bounding $\mathbb{E}[\sup_{x\in D}|\mathbb{H}_{h,T}(x)|^p]$, we want to apply Theorem 3.7 to the function class

$$\mathcal{G} := \left\{\overline{K}((x-\cdot)/h) : x \in D\cap\mathbb{Q}^d\right\}, \quad \text{where } \overline{K}((x-\cdot)/h) = K((x-\cdot)/h) - \mu(K((x-\cdot)/h)), \tag{3.29}$$

for some kernel function K with Lipschitz constant L wrt to the sup-norm $\|\cdot\|_\infty$, and the bandwidth h chosen in $(0,1)$. The following uniform deviation result is central for this purpose. The proof is given in Appendix 3.A along with useful bounds on the entropy integrals for processes from the class $\mathbf{\Sigma} \cup \mathbf{\Theta}$. Recall that if $\boldsymbol{X} \in \mathbf{\Sigma}$, then by definition, $\boldsymbol{X}$ is exponentially β-mixing, i.e., β-mixing with rate function $\Xi(t) = c_\kappa \mathrm{e}^{-\kappa t}$ for some constants $c_\kappa, \kappa > 0$.

LEMMA 3.9. *Suppose that $\mathbf{X} \in \mathbf{\Theta} \cup \mathbf{\Sigma}$ and additionally assume in case $\mathbf{X} \in \mathbf{\Theta}$ that $\mathbf{X}$ is β-mixing with strictly decreasing rate function $\Xi(t)$. Then, for any $u_T \geq 1$ such that $\Xi^{-1}(T^{-u_T}) \in \mathrm{o}(T)$ and $T^{-2} \leq h = h_T \in \mathrm{o}(1)$, there exists a constant $c^* > 0$ such that for large enough T*

$$\mathbb{P}\left(\left\|\widehat{\rho}_{h,T} - \mathbb{E}\widehat{\rho}_{h,T}\right\|_{L^\infty(D)} \geq c^*\left(\frac{u_T + \log T}{Th^d}\Xi^{-1}(T^{-u_T}) + T^{-\frac{1}{2}}\psi_d(h^d)\sqrt{u_T \vee \log(h^{-1})}\right)\right) \leq \mathrm{e}^{-u_T}.$$

In particular, when $\mathbf{X} \in \mathbf{\Sigma}$, for any $\gamma > 0$ and $u_T \in [1, \gamma \log T]$ there exists a constant $c_\gamma > 0$ such that for large enough T

$$\mathbb{P}\left(\left\|\widehat{\rho}_{h,T} - \mathbb{E}\widehat{\rho}_{h,T}\right\|_{L^\infty(D)} \geq c_\gamma \Upsilon_{h,T}(u_T)\right) \leq \mathrm{e}^{-u_T},$$

where

$$\Upsilon_{h,T}(u) := \frac{u(\log T)^2}{Th^d} + T^{-\frac{1}{2}}\psi_d(h^d)\sqrt{u \vee \log(h^{-1})}, \quad u \geq 1. \tag{3.30}$$

3.3 SUP-NORM ADAPTIVE ESTIMATION OF THE STATIONARY DENSITY FOR GENERAL MARKOV PROCESSES

In this section, we demonstrate the effectiveness of our previous results and probabilistic tools in a concrete statistical application. We already introduced the general form of the kernel invariant density estimator in (3.27). In order to quantify the speed of convergence, we will now analyse its convergence behavior under standard Hölder smoothness assumptions, i.e., we focus on the problem of estimating the invariant density ρ on a domain D of a Markov process $\mathbf{X} \in \mathbf{\Sigma} \cup \mathbf{\Theta}$ with $\rho|_D \in \mathcal{H}_D(\beta, \mathrm{L})$ (as introduced in (3.3)). For stating our statistical results, we define

$$\Phi_{d,\beta}(T) := \begin{cases} 1/\sqrt{T}, & d = 1, \\ \sqrt{\frac{\log T}{T}}, & d = 2, \\ T^{-\frac{\beta}{2\beta+d-2}}, & d \geq 3, \end{cases} \quad \text{and} \quad \Psi_{d,\beta}(T) := \begin{cases} \sqrt{\frac{\log T}{T}}, & d = 1, \\ \frac{\log T}{\sqrt{T}}, & d = 2, \\ \left(\frac{\log T}{T}\right)^{\frac{\beta}{2\beta+d-2}}, & d \geq 3. \end{cases} \tag{3.31}$$

Throughout, K denotes a $\|\cdot\|_\infty$-Lipschitz kernel of order ℓ and with Lipschitz constant L that is supported on $[-1/2, 1/2]^d$.

3.3.1 *General framework*

Depending on the concrete application, one might be interested in quantifying the accuracy of estimators in terms of different risk measures. Our findings from Section 3.1 immediately imply an upper bound on the classical mean squared error at some fixed point $x \in \mathbb{R}^d$.

COROLLARY 3.10. *Suppose that $\mathbf{X} \in \mathbf{\Sigma} \cup \mathbf{\Theta}$. For $x \in \mathbb{R}^d$ such that there exists an open neighbourhood $D \subset \mathbb{R}^d$ of x such that $\rho|_D \in \mathcal{H}_D(\beta, \mathrm{L})$, $\beta \in (0, \ell + 1]$, it holds for the kernel estimator*

$$\mathbb{E}\left[\left(\widehat{\rho}_{h,T}(x) - \rho(x)\right)^2\right] \in \mathrm{O}\left(\Phi^2_{d,\beta}(T)\right), \quad \text{if } h = h(T) \sim \begin{cases} T^{-1/\gamma}, & d \leq 2, \gamma \in (0, \beta], \\ T^{-1/(2\beta+d-2)}, & d \geq 3. \end{cases}$$

Proof. Fix x such that there exists an open neighbourhood D of x such that $\rho|_D \in \mathcal{H}_D(\beta, \mathrm{L})$. The usual bias-variance decomposition gives

$$\mathbb{E}\left[\left(\widehat{\rho}_{h,T}(x) - \rho(x)\right)^2\right] = (\rho * K_h(x) - \rho(x))^2 + \mathrm{Var}\left(\widehat{\rho}_{h,T}(x)\right). \tag{3.32}$$

For the bias term, since $\lfloor\beta\rfloor \leq \ell$, there exists a universal constant $M > 0$ such that

$$|(\rho * K_h - \rho)(x)| = \left|h^{-d}\int K\left(\frac{x-y}{h}\right)(\rho(y) - \rho(x))\,\mathrm{d}y\right| \leq Mh^\beta, \tag{3.33}$$

see Proposition 1.2 in [163] for the case $d = 1$ and the analogous estimator for discrete observations, which can be extended to the general multivariate case under continuous observations without much effort. Moreover, for any dimension d and $\boldsymbol{X} \in \boldsymbol{\Sigma} \cup \boldsymbol{\Theta}$, it follows from (3.47) that for any $h \in (0, 1)$

$$\mathrm{Var}\left(\frac{1}{T}\int_0^T K_h(x - X_t)\,\mathrm{d}t\right) \lesssim T^{-1}\|K\|_\infty^2\|\rho\|_\infty\psi_d^2(h^d).$$

The claim follows by plugging the specific choice of h into (3.33) and (3.47) and using (3.32). ■

We now turn our focus to the technically significantly more involved problem of sup-norm adaptive invariant density estimation for processes from the class Σ having Hölder continuous invariant densities. We demonstrate that optimal estimation rates in any dimension are achieved by the kernel estimator for a suitable bandwidth choice. While in dimension $d = 1, 2$ the optimal bandwidth has the remarkable property of being independent of the (typically unknown) order β of Hölder smoothness, this is not the case in higher dimensions $d \geq 3$. In order to remove β from the bandwidth choice, we need to find a data-driven substitute for the upper bound on the bias in the balancing process. Heuristically, this is the idea behind the Lepski-type selection procedure suggested now:

1. Specify the discrete set of candidate bandwidths

$$\mathcal{H}_T \equiv \mathcal{H}_T^{(k)} := \left\{h_l = \eta^{-l} :\ l \in \mathbb{N}_0,\ \eta^{-l} > \left(\frac{\log_{(k)} T(\log T)^5}{T}\right)^{\frac{1}{d+2}}\right\}, \quad \eta > 1 \text{ arbitrary,}$$

   for arbitrarily chosen $k \in \mathbb{N}$, and denote by $h_{\min}$ the smallest element in the grid $\mathcal{H}_T$. Here, $\log_{(k)} T$ denotes the k-th iterated logarithm, iteratively specified by $\log_{(k)} T := \log\log_{(k-1)} T$ and $\log_{(0)} T = T$, which is well-defined for T large enough.

2. Define $\widehat{h}_T \equiv \widehat{h}_T^{(k)}$ by letting

$$\widehat{h}_T := \max\left\{h \in \mathcal{H}_T : \left\|\widehat{\rho}_{h,T} - \widehat{\rho}_{g,T}\right\|_{L^\infty(D)} \leq \sqrt{\|\widehat{\rho}_{h_{\min},T}\|_{L^\infty(D)}}\,\sigma(g, T)\ \forall g \leq h,\ g \in \mathcal{H}_T\right\}, \tag{3.34}$$

   where

$$\sigma(h, T) := \frac{\log_{(k)} T(\log T)^2}{Th^d}\log(h^{-1}) + \psi_d(h^d)\sqrt{\frac{\log_{(k)} T\log(h^{-1})}{T}}, \quad h \in \mathcal{H}_T. \tag{3.35}$$

Letting $\|\cdot\|_{L^\infty(D)}$ denote the restriction of the sup-norm to a domain $D \subset \mathbb{R}^d$, we obtain the following result.

THEOREM 3.11. *Suppose that $\mathbf{X} \in \Sigma$. Let $D \subset \mathbb{R}^d$ be open and bounded. Suppose that $\rho|_D \in \mathcal{H}_D(\beta, \mathrm{L})$ with $\beta \in (1, \ell+1]$ for $d = 1$ and $\beta \in (2, \ell+1]$ for $d \geq 2$. Then, for any $p \geq 1$,*

$$\left(\mathbb{E}\left[\left\|\widehat{\rho}_{h,T} - \rho\right\|^p_{L^\infty(D)}\right]\right)^{1/p} \in \mathrm{O}\big(\Psi_{d,\beta}(T)\big), \quad \text{if } h = h(T) \sim \begin{cases} \log^2 T/\sqrt{T}, & d = 1, \\ \log T/T^{1/4}, & d = 2, \\ (\log T/T)^{1/(2\beta+d-2)}, & d \geq 3. \end{cases}$$

For the adaptive bandwidth scheme, let $\widehat{h}_T = \widehat{h}_T^{(k)}$ be selected according to (3.34) for some $k \in \mathbb{N}$. Then, if $\rho|_D \in \mathcal{H}_D(\beta, \mathrm{L})$ with $\beta \in (2, \ell+1]$, we have in any dimension $d \geq 3$,

$$\mathbb{E}\left[\left\|\widehat{\rho}_{\widehat{h}_T,T} - \rho\right\|_{L^\infty(D)}\right] \in \mathrm{O}\left(\left(\frac{\log_{(k)} T \log T}{T}\right)^{\frac{\beta}{2\beta+d-2}}\right). \tag{3.36}$$

Proof. Fix $p \geq 1$, and recall the decomposition (3.28). By the assumption on the order of the kernel K, the bias term $\rho * K_h - \rho$ is bounded by $B(h) := Mh^\beta$ for some universal constant $M > 0$ as in the pointwise case (see (3.33)), while the upper bound on the stochastic error $\mathbb{H}_{h,T}$ relies on a suitable specification on the upper bound in (3.50). For $d \geq 3$, set $h = h(T) = (\log T/T)^{1/(2\beta+d-2)}$ and $m_T = p \log T/\kappa$ such that

$$\frac{1}{\sqrt{T}}\psi_d(h^d) \in \mathrm{O}\big(T^{-\beta/(2\beta+d-2)}\big) \quad \text{and} \quad \frac{m_T}{Th^d} = \left(\frac{\log T}{T}\right)^{\frac{2(\beta-1)}{2(\beta-1)+d}}.$$

Upon noting that $\beta > 2$ implies $2(\beta-1) > \beta$, it follows from (3.50) that

$$\left(\mathbb{E}\left[\sup_{x\in D} |\mathbb{H}_{h,T}(x)|^p\right]\right)^{1/p} \in \mathrm{O}\left(\left(\frac{\log T}{T}\right)^{\beta/(2\beta+d-2)}\right). \tag{3.37}$$

Since $h^\beta = (\log T/T)^{\beta/(2\beta+d-2)}$, (3.28), (3.33) and (3.37) finally give $\mathbb{E}[\|\widehat{\rho}_{h,T} - \rho\|^p_{L^\infty(D)}]^{1/p} \in \mathrm{O}\big(\Psi_{d,\beta}(T)\big)$ for $d \geq 3$. For $d = 1$ and $d = 2$, the assertion follows by analogous arguments.

We now proceed with the proof of the convergence rate of the adaptive scheme for $d \geq 3$. For the variance, we obtain from (3.50) that, for $m_T := 2\log_{(k)} T(\log T)^2/\kappa$ and whenever $h \leq \mathrm{e}^{-2}L\mathrm{diam}(D)/\|K\|_\infty$, there exists some constant $\mathfrak{C} > 0$ such that

$$\mathbb{E}\left[\left\|\widehat{\rho}_{h,T} - \mathbb{E}\widehat{\rho}_{h,T}\right\|^2_{L^\infty(D)}\right] = \mathbb{E}\left[\sup_{x\in D} |\mathbb{H}_{h,T}(x)|^2\right] \leq \mathfrak{C}^2\sigma^2(h,T),$$

where $\sigma^2(\cdot,\cdot)$ is defined according to (3.35). Define h_ρ by the balance equation

$$h_\rho := \max\left\{h \in \mathcal{H}_T : B(h) \leq \frac{1}{4}\sqrt{0.8\mathcal{M}}\sigma(h,T)\right\}, \quad \text{where } \mathcal{M} := \|\rho\|_{L^\infty(D)}.$$

This definition implies that $B(h_\rho) \simeq \sqrt{0.8\mathcal{M}}\sigma(h_\rho,T)/4$ and, since $\mathcal{H}_T \ni h_\rho > \left(\frac{\log_{(k)} T(\log T)^5}{T}\right)^{\frac{1}{d+2}}$,

$$h_\rho^{2\beta+d-2} \simeq \frac{\log_{(k)} T \log T}{T} \quad \text{and} \quad \sigma(h_\rho,T) \simeq \left(\frac{\log_{(k)} T \log T}{T}\right)^{\frac{\beta}{2\beta+d-2}}.$$

To justify this, define $h_0 := (\log_{(k)} T \log T/T)^{1/(2\beta+d-2)}$. For large enough T, we have the estimate $\log(\log_{(k)} T \log T) \leq (\log T)/2$ and hence

$$\begin{aligned}\sigma(h_0,T) &= \frac{\log_{(k)} T(\log T)^2}{Th_0^d}\log(h_0^{-1}) + \psi_d(h_0^d)\sqrt{\frac{\log_{(k)} T\log(h_0^{-1})}{T}}\\ &\geq \sqrt{\frac{\log_{(k)} T\log T}{2(2\beta+d-2)T}}\psi_d(h_0^d) = \sqrt{\frac{1}{2(2\beta+d-2)}}\left(\frac{\log_{(k)} T\log T}{T}\right)^{\frac{\beta}{2\beta+d-2}} = \mathcal{L}^{-1}B(h_0),\end{aligned}$$

for $\mathcal{L} = \sqrt{2(2\beta+d-2)M^2}$. Additionally, we get, since $\beta > 2$,

$$\sigma(h_0,T) = \frac{\log_{(k)} T(\log T)^2}{Th_0^d}\log(h_0^{-1}) + \psi_d(h_0^d)\sqrt{\frac{\log_{(k)} T\log(h_0^{-1})}{T}} \simeq \left(\frac{\log_{(k)} T\log T}{T}\right)^{\frac{\beta}{2\beta+d-2}}.$$

In particular, it holds that $h_0 \lesssim h_\rho$, which is clear if $\mathcal{L} \leq \frac{1}{4}\sqrt{0.8\mathcal{M}}$, and else follows by the fact that, for any $0 < \lambda < 1$,

$$B(\lambda h_0) = \lambda^\beta B(h_0) \leq \lambda^\beta \mathcal{L}\sigma(h_0,T) \leq \lambda^\beta\mathcal{L}\sigma(\lambda h_0,T).$$

Lastly, we show $h_\rho \lesssim h_0$ by proving $h_\rho^{2\beta+d-2}h_0^{-(2\beta+d-2)} \in \mathrm{O}(1)$. Indeed, by the definition of h_ρ,

$$\begin{aligned}h_\rho^{2\beta+d-2} &\lesssim h_\rho^{d-2}\sigma^2(h_\rho,T)\\ &\lesssim h_\rho^{d-2}\left(\frac{\log_{(k)} T(\log T)^3}{T}h_\rho^{-d} + \psi_d(h_\rho^d)\sqrt{\frac{\log_{(k)} T\log T}{T}}\right)^2\\ &\lesssim \frac{(\log_{(k)} T)^2(\log T)^6}{T^2}h_\rho^{-(2+d)} + h_\rho^{d-2}\psi_d^2(h_\rho^d)\frac{\log_{(k)} T\log T}{T},\end{aligned}$$

and thus it holds that

$$h_\rho^{2\beta+d-2}h_0^{-(2\beta+d-2)} \lesssim \frac{\log_{(k)} T(\log T)^5}{T}h_\rho^{-(2+d)} + h_\rho^{d-2}\psi_d^2(h_\rho^d) \in \mathrm{O}(1),$$

thanks to $h_\rho > (\log_{(k)} T(\log T)^5/T)^{1/(d+2)}$.

Case 1: We first consider the case where $\widehat{h}_T \geq h_\rho$. To shorten notation, denote $\widetilde{\mathcal{M}} := \|\widehat{\rho}_{h_{\min},T}\|_{L^\infty(D)}$. Then, exploiting the definition of $\widehat{h}_T$ according to (3.34) and the bias and variance bounds,

$$\begin{aligned}&\mathbb{E}\left[\|\widehat{\rho}_{\widehat{h}_T,T} - \rho\|_{L^\infty(D)}\cdot\mathbf{1}_{\{\widehat{h}_T\geq h_\rho\}\cap\{\widetilde{\mathcal{M}}\leq 1.2\mathcal{M}\}}\right]\\ &\quad\leq \mathbb{E}\left[\left(\|\widehat{\rho}_{\widehat{h}_T,T} - \widehat{\rho}_{h_\rho,T}\|_{L^\infty(D)} + \|\widehat{\rho}_{h_\rho,T} - \mathbb{E}\widehat{\rho}_{h_\rho,T}\|_{L^\infty(D)} + B(h_\rho)\right)\mathbf{1}_{\{\widehat{h}_T\geq h_\rho\}\cap\{\widetilde{\mathcal{M}}\leq 1.2\mathcal{M}\}}\right]\\ &\quad\leq \sqrt{1.2\mathcal{M}}\sigma(h_\rho,T) + \mathfrak{C}\sigma(h_\rho,T) + \frac{1}{4}\sqrt{0.8\mathcal{M}}\sigma(h_\rho,T) \in \mathrm{O}(\sigma(h_\rho,T)).\end{aligned}$$

Similarly,

$$\begin{aligned}
&\mathbb{E}\Big[\|\widehat{\rho}_{\widehat{h}_T,T}-\rho\|_{L^\infty(D)}\cdot\mathbf{1}_{\{\widehat{h}_T\geq h_\rho\}\cap\{\widetilde{\mathcal{M}}>1.2\mathcal{M}\}}\Big]\\
&\quad\leq\sum_{h\in\mathcal{H}_T:h\geq h_\rho}\mathbb{E}\Big[\big(\|\widehat{\rho}_{h,T}-\mathbb{E}\widehat{\rho}_{h,T}\|_{L^\infty(D)}+B(h)\big)\cdot\mathbf{1}_{\{\widehat{h}_T=h\}\cap\{\widetilde{\mathcal{M}}>1.2\mathcal{M}\}}\Big]\\
&\quad\lesssim\log T\big(\mathfrak{C}\sigma(h_\rho,T)+B(1)\big)\sqrt{\mathbb{P}(\widetilde{\mathcal{M}}>1.2\mathcal{M})}.
\end{aligned}$$

Now, for any T large enough,

$$\begin{aligned}
\mathbb{P}\Big(\big|\widetilde{\mathcal{M}}-\mathcal{M}\big|>0.2\|\rho\|_{L^\infty(D)}\Big)&=\mathbb{P}\big(\big|\|\widehat{\rho}_{h_{\min},T}\|_{L^\infty(D)}-\|\rho\|_{L^\infty(D)}\big|>0.2\mathcal{M}\big)\\
&\leq\mathbb{P}\Big(\big\|\widehat{\rho}_{h_{\min},T}-\rho\big\|_{L^\infty(D)}>0.2\|\rho\|_{L^\infty(D)}\Big)\\
&\leq\mathbb{P}\Big(\big\|\widehat{\rho}_{h_{\min},T}-\mathbb{E}\widehat{\rho}_{h_{\min},T}\big\|_{L^\infty(D)}>0.2\|\rho\|_{L^\infty(D)}-B(h_{\min})\Big)\\
&\leq\mathbb{P}\Big(\big\|\widehat{\rho}_{h_{\min},T}-\mathbb{E}\widehat{\rho}_{h_{\min},T}\big\|_{L^\infty(D)}>0.1\|\rho\|_{L^\infty(D)}\Big)\\
&\leq\mathbb{P}\Big(\big\|\widehat{\rho}_{h_{\min},T}-\mathbb{E}\widehat{\rho}_{h_{\min},T}\big\|_{L^\infty(D)}>\Upsilon_{h_{\min},T}(\log T)\Big)\\
&\leq T^{-1},
\end{aligned}\tag{3.38}$$

where, for the function $\Upsilon_{h_{\min},T}(\cdot)$ defined according to (3.30), the last inequality follows from Lemma 3.9 and the last but one inequality holds since there exists some constant C such that

$$\begin{aligned}
\Upsilon_{h_{\min},T}(\log T)&\leq CT^{-\frac{2}{d+2}}\Big((\log T)^{\frac{6-2d}{d+2}}(\log_{(k)}T)^{-\frac{d}{d+2}}+(\log T)^{\frac{6-3d}{d+2}}(\log_{(k)}T)^{\frac{2-d}{2(d+2)}}\Big)\\
&\leq 0.2\|\rho\|_{L^\infty(D)},
\end{aligned}$$

for T sufficiently large. Thus, we can conclude that $\mathbb{E}\Big[\|\widehat{\rho}_{\widehat{h}_T,T}-\rho\|_{L^\infty(D)}\cdot\mathbf{1}_{\{\widehat{h}_T\geq h_\rho\}}\Big]\in\mathrm{O}(\sigma(h_\rho,T))$.

Case 2: For the case $\widehat{h}_T<h_\rho$, note first that the previous bias and variance bounds together with (3.38) imply that

$$\begin{aligned}
&\mathbb{E}\Big[\|\widehat{\rho}_{\widehat{h}_T,T}-\rho\|_{L^\infty(D)}\cdot\mathbf{1}_{\{\widehat{h}_T<h_\rho\}\cap\{\widetilde{\mathcal{M}}<0.8\mathcal{M}\}}\Big]\\
&\quad\leq\sum_{h\in\mathcal{H}_T:h<h_\rho}\mathbb{E}\Big[\big(\|\widehat{\rho}_{h,T}-\mathbb{E}\widehat{\rho}_{h,T}\|_{L^\infty(D)}+B(h)\big)\cdot\mathbf{1}_{\{\widehat{h}_T=h\}\cap\{\widetilde{\mathcal{M}}<0.8\mathcal{M}\}}\Big]\\
&\quad\lesssim\log T\big(\mathfrak{C}\sigma(h_{\min},T)+B(h_\rho)\big)\sqrt{\mathbb{P}(\widetilde{\mathcal{M}}<0.8\mathcal{M})}=O(\sigma(h_\rho,T)).
\end{aligned}$$

On the other hand,

$$\begin{aligned}
&\mathbb{E}\Big[\|\widehat{\rho}_{\widehat{h}_T,T}-\rho\|_{L^\infty(D)}\cdot\mathbf{1}_{\{\widehat{h}_T<h_\rho\}\cap\{0.8\mathcal{M}\leq\widetilde{\mathcal{M}}\}}\Big]\\
&\quad\leq\sum_{h\in\mathcal{H}_T:h<h_\rho}\mathbb{E}\Big[\big(\|\widehat{\rho}_{h,T}-\mathbb{E}\widehat{\rho}_{h,T}\|_{L^\infty(D)}+B(h)\big)\cdot\mathbf{1}_{\{\widehat{h}_T=h\}\cap\{0.8\mathcal{M}\leq\widetilde{\mathcal{M}}\}}\Big]\\
&\quad\leq\sum_{h\in\mathcal{H}_T:h<h_\rho}\sqrt{\mathbb{E}\Big[\|\widehat{\rho}_{h,T}-\mathbb{E}\widehat{\rho}_{h,T}\|^2_{L^\infty(D)}\Big]}\sqrt{\mathbb{E}\Big[\mathbf{1}_{\{\widehat{h}_T\geq h_\rho\}\cap\{0.8\mathcal{M}\leq\widetilde{\mathcal{M}}\}}\Big]}+B(h_\rho)
\end{aligned}$$

$$\leq \sum_{h\in\mathcal{H}_T : h<h_\rho} \mathfrak{C}\sigma(h,T)\sqrt{\mathbb{P}\Big(\{\widehat{h}_T \geq h_\rho\}\cap\{0.8\mathcal{M}\leq \widetilde{\mathcal{M}}\}\Big)} + \mathrm{O}(\sigma(h_\rho,T)).$$

Pick any $h \in \mathcal{H}_T$ such that $h < h_\rho$ and denote $h^+ := \min\{g \in \mathcal{H}_T : g > h\} = \eta h$. It is then shown as in the proof of Theorem 2 in [85] that the verification of the fact that the first sum on the rhs of the last display is of order $\mathrm{O}(\sigma(h_\rho,T))$ boils down to proving that

$$\sum_{h\in\mathcal{H}_T:h<h_\rho}\sigma(h,T)\Bigg(\sum_{g\in\mathcal{H}_T:g\leq h}\mathbb{P}\Big(\big\|\widehat{\rho}_{h^+,T}-\widehat{\rho}_{g,T}\big\|_{L^\infty(D)}>\sqrt{0.8\mathcal{M}}\sigma(g,T)\Big)\Bigg)^{1/2}\in\mathrm{O}(\sigma(h_\rho,T)).$$

Following again the lines of [85], we obtain

$$\begin{aligned}\mathbb{P}\Big(\big\|\widehat{\rho}_{h^+,T}-\widehat{\rho}_{g,T}\big\|_{L^\infty(D)}>\sqrt{0.8\mathcal{M}}\sigma(g,T)\Big) &\leq \mathbb{P}\Big(\big\|\widehat{\rho}_{h^+,T}-\mathbb{E}\widehat{\rho}_{h^+,T}\big\|_{L^\infty(D)}>\frac{1}{4}\sqrt{0.8\mathcal{M}}\sigma(h^+,T)\Big)\\ &\quad+\mathbb{P}\Big(\big\|\widehat{\rho}_{g,T}-\mathbb{E}\widehat{\rho}_{g,T}\big\|_{L^\infty(D)}>\frac{1}{4}\sqrt{0.8\mathcal{M}}\sigma(g,T)\Big).\end{aligned}$$

Let $\gamma \geq 1$. Clearly, by definition of $\sigma(g,T)$, there exists $T(\gamma) > 0$ such that, for any $T \geq T(\gamma)$ and any $g \leq h_\rho$, $g \in \mathcal{H}_T$, we have

$$\frac{1}{4}\sqrt{0.8\mathcal{M}}\sigma(g,T)\geq c_\gamma\Upsilon_{g,T}(\gamma\log(g^{-1}))=c_\gamma\frac{\gamma\log(g^{-1})(\log T)^2}{Tg^d}+\psi_d(g^d)\sqrt{\frac{\gamma\log(g^{-1})}{T}},$$

where c_γ is the constant appearing in Lemma 3.9. Thus, using Lemma 3.9, we obtain for $T \geq T(\gamma)$ that

$$\mathbb{P}\Big(\big\|\widehat{\rho}_{g,T}-\mathbb{E}[\widehat{\rho}_{g,T}]\big\|_{L^\infty(D)}>\frac{1}{4}\sqrt{0.8\mathcal{M}}\sigma(g,T)\Big)\leq \mathrm{e}^{-\gamma\log(g^{-1})}=g^\gamma=:\iota_\gamma(g)$$

and hence

$$\sum_{g\in\mathcal{H}_T:g\leq h}\mathbb{P}\Big(\big\|\widehat{\rho}_{h^+,T}-\widehat{\rho}_{g,T}\big\|_{L^\infty(D)}>\sqrt{0.8\mathcal{M}}\sigma(g,T)\Big)\leq\sum_{g\in\mathcal{H}_T:g\leq h}(\iota_\gamma(g)+\iota_\gamma(h^+))\leq 2\iota_\gamma(h)\log T.$$

Thus, choosing γ large enough demonstrates that

$$\begin{aligned}&\sum_{h\in\mathcal{H}_T:h<h_\rho}\sigma(h,T)\Bigg(\sum_{g\in\mathcal{H}_T:g\leq h}\mathbb{P}\Big(\big\|\widehat{\rho}_{h^+,T}-\widehat{\rho}_{g,T}\big\|_{L^\infty(D)}>\sqrt{0.8\mathcal{M}}\sigma(g,T)\Big)\Bigg)^{1/2}\\ &\qquad\leq\sum_{h\in\mathcal{H}_T:h<h_\rho}\sigma(h,T)\sqrt{2\iota_\gamma(h)\log T}\leq\sqrt{2h_\rho^\gamma(\log T)^3}\sigma(h_{\min},T)\in\mathrm{O}(\sigma(h_\rho,T)),\end{aligned}$$

as desired. ∎

The convergence rates introduced in (3.31) clearly reflect the fact that the invariant density of stochastic processes can be estimated faster than in the classical context of nonparametric density estimation based on i.i.d. observations. While this is well-known for ergodic continuous diffusion processes (see [59, 158]), we will show in the following section that the result is fulfilled for a much larger class of stochastic processes. The additional log-factor occurring in the definition of $\Psi_{d,\beta}(\cdot)$ represents the common price to be paid when switching from the pointwise error control (described by $\Phi_{d,\beta}(\cdot)$) to bounding the sup-norm risk.

Remark 3.12. (a) The conditions on the Hölder index β stated in Theorem 3.11 are due to two different reasons: On the one hand, in dimension $d \leq 2$, we chose a bandwidth not depending on β which still achieves the optimal balance between bias and stochastic error. By choosing a bandwidth dependent on β (as in Corollary 3.10), restrictions on β could be avoided. However, for the implementation of estimators it is advantageous to be able to choose a bandwidth independent of the typically unknown smoothness β. On the other hand, in dimension $d \geq 3$, the assumption on β is an unavoidable effect. The coupling error leaves us no other choice but to select the interval block length m_T in the decomposition of (3.27) of order $\log T$, which forces $\beta > 2$ to balance out bias and stochastic sensitivity of the estimator. We emphasize that this is not an artifact of our proof strategy since the additional log-factor also appears in the optimal Bernstein inequalities for geometrically ergodic Markov chains in [1, 114]. The restriction on β can therefore be considered as a price that must be paid for the generality of our exponential β-mixing assumption.

(b) The logarithmic gap (of arbitrary iterative order k) between the adaptive rate (see (3.36)) and the optimal rate $\Psi_{d,\beta}$ in dimension $d \geq 3$ (see (3.31)) is *not* a consequence of suboptimality of arguments used in the proof. Rather, it is a deliberate choice motivated by our desire to introduce a truly adaptive selection procedure that does not rely on the specification of obscure constants. To be more precise, a key step in the proof of the upper bound for the adaptive approach requires quantifying the concentration of the estimator $\widehat{\rho}_{h,T}$ around the variance proxy $\sigma(h,T)$ from (3.35), which is handled with the deviation inequality from Lemma 3.9 involving the term $\Upsilon_{h,T}(\gamma \log T)$ (see (3.30)). If we remove the factor $\log_{(k)} T$ in the variance proxy $\sigma(h,T)$, we obtain

$$\frac{(\log T)^2}{Th^d}\log(h^{-1}) + \psi_d(h^d)\sqrt{\frac{\log(h^{-1})}{T}} \simeq \Upsilon_{h,T}(\gamma \log T).$$

In this case, an exact quantification of the constant c_γ from Lemma 3.9 is mandatory, which would then be included as an additional factor in the specification of $\widehat{h}_T$ in (3.34). Together with an adjustment of the candidate bandwidths $\mathcal{H}_T$, this would allow us to close the logarithmic gap and hence obtain optimal rates for the adaptive procedure.

However, c_γ is of the form $\gamma \times C(D, L, \kappa, c_\kappa, c_2)$—where we recall that c_κ, κ determine the mixing coefficient and c_2 is a constant appearing in the heat kernel bound from Assumption ($\mathscr{A}$1)—and therefore can only be bounded with explicit knowledge/assumptions on the process. We avoid this fundamental problem in our procedure to not shift the problem from unknown exact smoothness to unknown exact ergodic and small time behavior, with the price to be paid being a logarithmic loss. In this regard, our approach differs from the bandwidth selection procedure for the L^2 risk in [12], which relies on the choice of a "sufficiently large" constant k that cannot be exactly specified or efficiently chosen in a data-driven way.

3.4 Examples

Our previous results rely on the very general conditions ($\mathscr{A}$0) and ($\mathscr{A}$1) as well as assumptions related to the speed of convergence to the invariant distribution, ($\mathscr{A}$2) and ($\mathscr{A}\beta$). For statistical purposes, however, it is essential to derive results under conditions on the coefficients of the underlying process as easily verifiable as possible. For this reason, this section is devoted to investigating specific classes of jump diffusion processes and explicit conditions on their underlying characteristics such that the above assumptions are satisfied and hence statistical conclusions can be drawn from our general theory.

3.4.1 *Lévy-driven Ornstein–Uhlenbeck processes*

As a first example we discuss estimation rates of d-dimensional Lévy-driven Ornstein–Uhlenbeck processes as representatives of Lévy-driven jump diffusions with unbounded drift coefficient by establishing assumptions on the characteristics of the Lévy process that guarantee $\boldsymbol{X} \in \boldsymbol{\Sigma} \cup \boldsymbol{\Theta}$. Let $\boldsymbol{Z}$ be a d-dimensional Lévy process with generating triplet (a, Q, ν), where $a \in \mathbb{R}^d$, $Q \in \mathbb{R}^{d\times d}$ is a symmetric positive semidefinite matrix and ν is a measure on $\mathbb{R}^d$ satisfying $\nu(\{0\}) = 0$ and $\int_{\mathbb{R}^d}(1 \wedge \|x\|^2)\,\nu(\mathrm{d}x) < \infty$ such that $\mathbb{E}^0[\exp(\mathrm{i}\langle Z_1, \theta\rangle)] = \exp(\psi(\theta))$ with

$$\psi(\theta) = \mathrm{i}\langle a, \theta\rangle - \frac{1}{2}\langle Q\theta, \theta\rangle + \int_{\mathbb{R}^d\setminus\{0\}} \left(\mathrm{e}^{\mathrm{i}\langle x,\theta\rangle} - 1 - \mathrm{i}\langle x, \theta\rangle \mathbb{1}_{B(0,1)}(x)\right) \nu(\mathrm{d}x), \quad \theta \in \mathbb{R}^d,$$

where $B(0,1) = \{x \in \mathbb{R}^d : \|x\| < 1\}$. Then, given some matrix $B \in \mathbb{R}^{d\times d}$, a Lévy driven Ornstein–Uhlenbeck process $\boldsymbol{X}$ is a solution to the SDE

$$\mathrm{d}X_t = -BX_t\,\mathrm{d}t + \mathrm{d}Z_t,$$

given by

$$X_t = \mathrm{e}^{-tB}X_0 + \int_0^t \mathrm{e}^{-(t-s)B}\,\mathrm{d}Z_s, \quad t \geq 0.$$

We suppose that the real parts of all eigenvalues of B are positive, implying that $\mathrm{e}^{-tB} \longrightarrow \mathbb{0}_{d\times d}$ as $t \longrightarrow \infty$, and assume the following moment condition

$$\int_{\|z\|>2} \log\|z\|\,\nu(\mathrm{d}z) < \infty. \tag{3.39}$$

Then, $\boldsymbol{X}$ is a Markov process on $\mathbb{R}^d$ with invariant distribution μ such that

$$\{\mathfrak{F}\mu\}(u) = \exp\left(\int_0^\infty \psi\big(\mathrm{e}^{-sB^\top}u\big)\,\mathrm{d}s\right), \quad u \in \mathbb{R}^d,$$

$$\text{and } \varphi_{X_t}^x(u) = \exp\left(\mathrm{i}\langle x, \mathrm{e}^{-tB^\top}u\rangle + \int_0^t \psi\big(\mathrm{e}^{-sB^\top}u\big)\,\mathrm{d}s\right), \quad u, x \in \mathbb{R}^d, t > 0,$$

see [148, Theorem 3.1, Theorem 4.1]. Let us now introduce the following conditions.

($\mathfrak{O}1$) $\operatorname{rank}(Q) = d$;

($\mathfrak{O}2$) $\int_{\{\|x\|>1\}}\|x\|^p\,\nu(\mathrm{d}x) < \infty$ for some $p > 0$.

($\mathfrak{O}3$) $\int_{\{\|x\|>1\}}(\log\|x\|)^\alpha\,\nu(\mathrm{d}x) < \infty$ for some $\alpha > 2$;

These assumptions are borrowed from [125], [123] and [103], where (sub-)exponential ergodicity and exponential β-mixing of OU-processes are investigated. ($\mathfrak{O}1$) guarantees the strong

Feller property of $\boldsymbol{X}$ and the existence of a $\mathcal{C}_b^\infty$-density for $P_t(x,\cdot)$, $x \in \mathbb{R}^d$ ([125, Theorem 3.1]). Similar arguments to the ones in [125, Theorem 3.2] also show that under $(\mathfrak{S}1)$, μ admits a $\mathcal{C}_b^\infty$-density ρ. $(\mathfrak{S}2)$ and $(\mathfrak{S}3)$ are moment assumptions on $\boldsymbol{Z}$, where $(\mathfrak{S}3)$ in absence of $(\mathfrak{S}2)$ corresponds to an extremely heavy tailed distribution and represents a minor strengthening of the necessary and sufficient criterion (3.39) for stationarity of $\boldsymbol{X}$.

Based on the results from [103, 123, 125] together with our investigations in Sections 3.1 and 3.3, we can obtain the following result.

THEOREM 3.13. *Suppose that $(\mathfrak{S}1)$ holds. Then, in any dimension $d \in \mathbb{N}$, $(\mathscr{A}1)$ holds with*

$$\sup_{x,y\in\mathbb{R}^d} p_t(x,y) \lesssim t^{-d/2}, \quad t \in (0,1]. \tag{3.40}$$

If additionally,

(i) $(\mathfrak{S}2)$ holds for some $p > 0$, then for any $d \geq 1$, $\boldsymbol{X} \in \boldsymbol{\Sigma} \cap \boldsymbol{\Theta}$;

(ii) $(\mathfrak{S}3)$ holds, then for $d = 1$, $\boldsymbol{X} \in \boldsymbol{\Theta}$.

Let $d \geq 1$ in scenario (i) and $d = 1$ in scenario (ii). Then, for arbitrary $\beta > 0$ we obtain for any $x \in \mathbb{R}^d$, that

$$\mathbb{E}\left[\left(\widehat{\rho}_{h,T}(x) - \rho(x)\right)^2\right] \in \mathrm{O}\left(\Phi^2_{d,\beta}(T)\right), \quad \textit{if } h = h(T) \sim \begin{cases} T^{-1}, & d \leq 2, \\ T^{-1/(2\beta+d-2)}, & d \geq 3. \end{cases}$$

and for any bounded, open domain $D \subset \mathbb{R}^d$ and $p \geq 1$ that in scenario (i)

$$\mathbb{E}\left[\left\|\widehat{\rho}_{h,T} - \rho\right\|^p_{L^\infty(D)}\right]^{1/p} \in \mathrm{O}\left(\Psi_{d,\beta}(T)\right), \quad \textit{if } h = h(T) \sim \begin{cases} \log^2 T/\sqrt{T}, & d = 1, \\ \log T/T^{1/4}, & d = 2, \\ (\log T/T)^{1/(2\beta+d-2)}, & d \geq 3. \end{cases}$$

Remark 3.14. (i) Since we can choose $\beta > 0$ arbitrarily large, we make the remarkable observation that in the scenarios described above, for any $\varepsilon > 0$ we can obtain the almost superoptimal rates $T^{-(1+\varepsilon)}$ and $(\log T/T)^{1/(2(1+\varepsilon))}$ in any dimension $d \geq 3$ for the pointwise L^2 and sup-norm risk, respectively. Moreover, in any dimension, an adaptive choice of the bandwidth is not necessary.

(ii) We emphasize that this result demonstrates that even under much less stringent assumptions (logarithmic moments and unbounded drift) compared to the class of processes studied in the next section, there are examples of Lévy driven jump diffusions for which optimal estimation results are feasible. It is therefore an interesting question for future research to determine more general coefficient assumptions based on a linear growth condition on the drift that yield optimal estimation properties.

Proof of Theorem 3.13. Let us first verify that under $(\mathfrak{S}1)$ the heat kernel bound $(\mathscr{A}1)$ holds. Arguing as in the proof of Theorem 3.2 of Masuda [125], we see that $\mathcal{F}\mu$ and $\varphi^x_{X_t}$ are integrable for any $x \in \mathbb{R}$ and $t > 0$ and hence we can obtain the invariant density ρ and the transition density p_t of $\boldsymbol{X}$ via inverse Fourier transformation through

$$\rho(y) = \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d} \mathrm{e}^{-\mathrm{i}\langle y,\lambda\rangle}\{\mathcal{F}\mu\}(\lambda)\,\mathrm{d}\lambda, \quad y \in \mathbb{R}^d,$$

and
$$p_t(x,y) = \frac{1}{(2\pi)^d}\int_{\mathbb{R}} \mathrm{e}^{-\mathrm{i}\langle y,\lambda\rangle}\varphi^x_{X_t}(\lambda)\,\mathrm{d}\lambda, \quad x,y\in\mathbb{R}^d, t>0.$$
Again, as in the proof of Theorem 3.2 in [125], it follows that under $(\mathfrak{O}1)$,
$$\big|\varphi^x_{X_t}(\lambda)\big| \leq \exp\Big(-\frac{1}{2}\lambda^\top\Big(\int_0^t \mathrm{e}^{-sB}Q\mathrm{e}^{-sB^\top}\,\mathrm{d}s\Big)\lambda\Big), \quad x,\lambda\in\mathbb{R}^d, t>0. \tag{3.41}$$
Thus, using the characterization of the multivariate normal distribution, we obtain
$$\begin{aligned} p_t(x,y) &\leq \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d}\exp\Big(-\frac{1}{2}\lambda^\top\Big(\int_0^t \mathrm{e}^{-sB}Q\mathrm{e}^{-sB^\top}\,\mathrm{d}s\Big)\lambda\Big)\,\mathrm{d}\lambda \\ &= \frac{1}{(2\pi)^{d/2}}\Big(\det\Big(\int_0^t \mathrm{e}^{-sB}Q\mathrm{e}^{-sB^\top}\,\mathrm{d}s\Big)\Big)^{-1/2}. \end{aligned}$$
Observing that
$$\begin{aligned} \lim_{t\downarrow 0} t^{d/2}\Big(\det\Big(\int_0^t \mathrm{e}^{-sB}Q\mathrm{e}^{-sB^\top}\,\mathrm{d}s\Big)\Big)^{-1/2} &= \Big(\det\Big(\lim_{t\downarrow 0}\frac{1}{t}\int_0^t \mathrm{e}^{-sB}Q\mathrm{e}^{-sB^\top}\,\mathrm{d}s\Big)\Big)^{-1/2} \\ &= \det(Q)^{-1/2} < \infty, \end{aligned}$$
where finiteness is a consequence of invertibility of Q by $(\mathfrak{O}1)$, it follows that for any $d\geq 1$, there exists a constant $c>0$ such that
$$\sup_{x,y\in\mathbb{R}^d} p_t(x,y) \leq ct^{-d/2}, \quad t\in(0,1].$$
Thus indeed, for any dimension $d\in\mathbb{N}$, $(\mathscr{A}1)$ holds. Next, in scenario (i), [125, Theorem 4.3] gives the exponential β-mixing property and the proof of Theorem 2.6 in [123] along with [123, Proposition 3.8] yields V-exponential ergodicity with $V(x)\sim(1+\|x\|^p)$. This together with (3.40) entails that in scenario (i), we have $\boldsymbol{X}\in\boldsymbol{\Sigma}\cap\boldsymbol{\Theta}$. Finally, $\boldsymbol{X}\in\boldsymbol{\Theta}$ in scenario (ii) follows from the considerations above and Lemma 3.4 due to the fact that the combination of $(\mathfrak{O}1)$ and the logarithmic moment condition imply that every compact set is small and hence petite since $\boldsymbol{X}$ is strong Feller and by [99, Theorem 3.1] ergodic (see Proposition 2.11) and hence $(\mathfrak{O}3)$ implies V-polynomial ergodicity of degree $\alpha-1>1$ with $V(x)=C(\log|x|)^\alpha$ in dimension $d=1$ by [103, Corollary 1]. The statements on the estimation rates are now an immediate consequence of Corollary 3.10 and Theorem 3.11 and the fact that $\rho\in\mathcal{C}_b^\infty$ has arbitrary Hölder smoothness. ∎

3.4.2 *Non-reversible Lévy-driven jump diffusion processes*

The goal of this section is to show that solutions of the d-dimensional SDE, $d\in\mathbb{N}$,
$$X_t = X_0 + \int_0^t b(X_s)\,\mathrm{d}s + \int_0^t \sigma(X_s)\,\mathrm{d}W_s + \int_0^t\int_{\mathbb{R}^d}\gamma(X_{s-})z\,\widetilde{N}(\mathrm{d}s,\mathrm{d}z) \tag{3.42}$$
satisfy assumptions $(\mathscr{A}0)$, $(\mathscr{A}1)$ and $(\mathscr{A}\beta)$ which then allows using Theorem 3.11 to bound the sup-norm risk of the kernel invariant density estimator. Here, $\sigma\colon\mathbb{R}^d\to\mathbb{R}^{d\times d}$, $\gamma\colon\mathbb{R}^d\to\mathbb{R}^{d\times d}$, $b\colon\mathbb{R}^d\to\mathbb{R}^d$, W denotes an $\mathbb{R}^d$-valued Brownian motion, N is a Poisson random measure

on $[0,\infty)\times\mathbb{R}^d\backslash\{0\}$ with intensity measure $\mu(\mathrm{d}s,\mathrm{d}z)=\mathrm{d}s\otimes\nu(\mathrm{d}z)$, and $\widetilde{N}$ denotes the compensated Poisson random measure. Moreover, ν is a Lévy measure and we assume that N,W and X_0 are independent. Note that, if $z\mapsto\gamma(x)z$ is in $L^1(\mathbb{R}^d\backslash\{B_1\},\nu)$ for all $x\in\mathbb{R}^d$, (3.42) is equivalent to

$$\begin{aligned} X_t = X_0 &+ \int_0^t b^*(X_s)\,\mathrm{d}s + \int_0^t \sigma(X_s)\,\mathrm{d}W_s \\ &+ \int_0^t\int_{\|z\|\leq 1}\gamma(X_{s-})z\,\widetilde{N}(\mathrm{d}s,\mathrm{d}z) + \int_0^t\int_{|z|>1}\gamma(X_{s-})z\,N(\mathrm{d}s,\mathrm{d}z), \end{aligned} \tag{3.43}$$

with $b^*(x):=b(x)-\int_{\|z\|>1}\gamma(x)z\,\nu(dz)$ and $B_1:=\{z\in\mathbb{R}^d:\|z\|\leq 1\}$. We assume the following.

($\mathfrak{J}$1) The functions b,γ,σ are globally Lipschitz continuous, b and γ are bounded, and, for $\mathbb{I}_{d\times d}$ denoting the $d\times d$-identity matrix, there exists a constant $c\geq 1$ such that

$$c^{-1}\mathbb{I}_{d\times d}\leq\sigma\sigma^\top\leq c\mathbb{I}_{d\times d},$$

where the ordering is in the sense of Loewner for positive semi-definite matrices.

($\mathfrak{J}$2) ν is absolutely continuous wrt. the Lebesgue measure and, for an $\alpha\in(0,2)$,

$$(x,z)\mapsto\|\gamma(x)z\|^{d+\alpha}\nu(z)$$

is bounded and measurable, where, by abuse of notation, we denoted the density of ν also by ν. Furthermore, if $\alpha=1$,

$$\int_{r<\|\gamma(x)z\|\leq R}\gamma(x)z\,\nu(\mathrm{d}z)=0,\qquad\text{for any } 0<r<R<\infty,\ x\in\mathbb{R}^d.$$

($\mathfrak{J}$3) There exist $c_1,c_2>0$ and $\eta_0>0$ such that

$$\langle x,b(x)\rangle\leq -c_1\|x\|,\quad\forall x:\|x\|\geq c_2,\quad\text{and}\quad\int_{\mathbb{R}^d}\|z\|^2\mathrm{e}^{\eta_0\|z\|}\nu(\mathrm{d}z)<\infty.$$

In [12], the authors also investigate L^2 invariant density estimation for jump diffusions and use a similar approach for formulating requirements on the diffusion coefficients which imply their respective heat kernel bound and mixture assumptions. The conditions however are more restrictive and, in particular, the case of continuous diffusions cannot be handled within their framework since it requires $\operatorname{supp}(\nu)=\mathbb{R}^d$ and $\det(\gamma(x))>c$ for some constant $c>0$ and all $x\in\mathbb{R}^d$. In [13], the authors improve the L^2 rate for dimension $d=1$ from [12] to the parametric rate $1/T$ by imposing an additional smoothness restriction on the jump measure. Our main contribution in this section is to show that under the less stringent assumptions above, optimal convergence rates can be achieved not only wrt. the L^2 risk but even wrt. sup-norm risk in any dimension.

Note that ($\mathfrak{J}$1) and ($\mathfrak{J}$3) directly imply $\gamma(x)z\in L^1(\mathbb{R}^d\backslash\{B_1\},\nu)$, so (3.42) and (3.43) are equivalent. The subsequent lemma shows that, under the given assumptions, there exists a pathwise unique strong solution for (3.42) and that the conditions of Corollary 1.5 of [47] hold, implying the heat kernel bound (3.44). All proofs can be found in Appendix 3.B.

Lemma 3.15. *Let* $(\mathscr{J}1)$–$(\mathscr{J}3)$ *hold. Then,* (3.42) *admits a càdlàg, non-explosive, pathwise unique, strong solution possessing the strong Markov property, and the assumptions* $(\mathbf{H}^{\alpha})$ *and* $(\mathbf{H}^{\kappa})$ *of [47] hold.*

Let $\boldsymbol{X}$ be the unique solution of (3.42) described in Lemma 3.15.

Corollary 3.16. *Let* $(\mathscr{J}1)$–$(\mathscr{J}3)$ *hold. Then, transition densities* $(p_t)_{t>0}$ *exist and there are constants* $C, \lambda > 1$ *such that the solution* $\boldsymbol{X}$ *of* (3.42) *satisfies the following heat kernel estimate for all* $x, y \in \mathbb{R}^d, 0 < t \leq 1$,

$$\begin{aligned} &C^{-1}(t^{-d/2}\exp(-\lambda\|x-y\|^2/t) + (\inf_{x\in\mathbb{R}^d}\operatorname{ess\,inf}_{z\in\mathbb{R}^d}\kappa_\alpha(x,z))t(\|x-y\|+t^{1/2})^{-d-\alpha}) \\ &\quad \leq p_t(x,y) \leq C(t^{-d/2}\exp(-\|x-y\|^2/(\lambda t)) + \|\kappa_\alpha\|_\infty t(\|x-y\|+t^{1/2})^{-d-\alpha}), \end{aligned} \tag{3.44}$$

where $\kappa_\alpha(x,z) = \|\gamma(x)z\|^{d+\alpha}\nu(z)$*. In particular, assumption* $(\mathscr{A}1)$ *is satisfied.*

Now our goal is to show that the solution $\boldsymbol{X}$ of (3.42) fulfills the fundamental assumption $(\mathscr{A}0+)$ and exponential ergodicity along with the mixing property $(\mathscr{A}\beta)$. First, observe that $(\mathscr{J}1)$ implies that $b \in \mathcal{C}_b(\mathbb{R}^d;\mathbb{R}^d)$ and $\sigma, \gamma \in \mathcal{C}_b(\mathbb{R}^d;\mathbb{R}^{d\times d})$ and hence Theorem 6.7.4 in [14] guarantees that the unique càdlàg Markov process $\boldsymbol{X}$ solving (3.42) is Feller and therefore Borel right. Further, Corollary 3.16 in particular implies the existence of bounded transition densities and thus, by Lemma 3.5, it suffices to show the existence of an invariant distribution. This will be done as a byproduct while proving exponential ergodicity and the exponential mixing property $(\mathscr{A}\beta)$. For this, we will employ results of Masuda [123] which are again based on the theory of stability of continuous-time Markov processes of Meyn and Tweedie [132]. These lead us to the following proposition.

Proposition 3.17. *Grant assumptions* $(\mathscr{J}1)$–$(\mathscr{J}3)$*. Then, an invariant distribution exists,* $\boldsymbol{X}$ *is* V*-exponentially ergodic with locally bounded* V *and the process* $\boldsymbol{X}$ *started in the invariant distribution* μ *is exponentially* β*-mixing.*

Gathering the results of Corollary 3.16 and Proposition 3.17 and employing Lemma 3.4 now yields that $(\mathscr{A}0)$–$(\mathscr{A}2)$ and $(\mathscr{A}\beta)$ are fulfilled for the solution $\boldsymbol{X}$ of (3.42), i.e., $\boldsymbol{X} \in \boldsymbol{\Sigma} \cap \boldsymbol{\Theta}$. In particular, the results from Section 3.3 can be applied.

Theorem 3.18. *Let* $D \subset \mathbb{R}^d$ *be open and bounded and assume* $(\mathscr{J}1)$–$(\mathscr{J}3)$*. If* $\rho|_D \in \mathcal{H}_D(\beta, \mathrm{L})$ *with* $\beta \in (1, \ell+1]$ *for* $d = 1$ *and* $\beta \in (2, \ell+1]$ *for* $d \geq 2$*, then, the* sup*-norm risk of the kernel estimator defined in* (3.27) *is of order*

$$\mathbb{E}\Big[\|\widehat{\rho}_{h,T}-\rho\|^p_{L^\infty(D)}\Big]^{1/p} \in \begin{cases} \mathrm{O}(\sqrt{\log T/T}), & d=1,\\ \mathrm{O}(\log T/\sqrt{T}), & d=2,\\ \mathrm{O}\big((\log T/T)^{\beta/(2\beta+d-2)}\big), & d\geq 3, \end{cases} \quad \text{if } h \sim \begin{cases} \log^2 T/\sqrt{T}, & d=1,\\ \log T/T^{1/4}, & d=2,\\ (\log T/T)^{-1/(2\beta+d-2)}, & d\geq 3. \end{cases}$$

for any $p \geq 1$*. If* $\widehat{h}_T \equiv \widehat{h}_T^{(k)}$ *is chosen adaptively according to* (3.34) *for some* $k \in \mathbb{N}$*, then for any* $d \geq 3$,

$$\mathbb{E}\Big[\big\|\widehat{\rho}_{\widehat{h}_T} - \rho\big\|_{L^\infty(D)}\Big] \in \mathrm{O}\left(\left(\frac{\log_{(k)} T \log T}{T}\right)^{\beta/(2\beta+d-2)}\right).$$

Moreover, for any $x \in \mathbb{R}^d$ such that $\rho|_D \in \mathcal{H}_D(\beta, \mathrm{L})$ for some $\beta \in (0, \ell+1]$ and a neighborhood D of x, we have the pointwise L^2 risk estimate

$$\mathbb{E}\Big[\big(\widehat{\rho}_{h,T}(x) - \rho(x)\big)^2\Big] \in \begin{cases} \mathrm{O}(1/T), & d = 1, \\ \mathrm{O}(\log T/T), & d = 2, \\ \mathrm{O}\big(T^{-2\beta/(2\beta+d-2)}\big), & d \geq 3, \end{cases} \quad \text{if } h \sim \begin{cases} T^{-1/\gamma}, & d \leq 2, \gamma \leq \beta, \\ T^{-1/(2\beta+d-2)}, & d \geq 3. \end{cases}$$

3.A PROOF OF LEMMA 3.9

For the proof of the bounds on the stochastic error, we start with the following preparatory lemma that provides bounds of the covering numbers of the function class $\mathcal{G}$ introduced in (3.29) with respect to the norms appearing in Theorem 3.7. By a slight abuse of notation, we do not distinguish notationally between the sup-norm on $\mathbb{R}^d$ and the function space $\mathcal{B}_b(\mathbb{R}^d)$.

LEMMA 3.19. *Let $D \subset \mathbb{R}^d$ be a bounded set and, given some Lipschitz continuous kernel K with Lipschitz constant L and compact support $[-1/2, 1/2]^d$, define the function class $\mathcal{G}$ according to (3.29). Then, for any $\varepsilon > 0$,*

$$\mathcal{N}(\varepsilon, \mathcal{G}, \|\cdot\|_{d_\infty}) \leq \Big(\frac{4L\mathrm{diam}(D)}{\varepsilon h}\Big)^d,$$

and if moreover $\mathbf{X} \in \mathbf{\Sigma} \cup \mathbf{\Theta}$, then there exists a constant $\mathbb{A} > 0$ such that, for any $\varepsilon > 0$ and $t > 0$,

$$\mathcal{N}(\varepsilon, \mathcal{G}, \|\cdot\|_{\mathbb{G},t}) \leq \Big(\frac{2L\mathrm{diam}(D)\sqrt{\mathbb{A}\|\rho\|_\infty} h^{d-1}\psi_d(h^d)}{\varepsilon}\Big)^d.$$

Proof. For $x \in \mathbb{R}^d$, we obtain by Lipschitz continuity of K that

$$\begin{aligned} B_{d_\infty}\big(\overline{K}\big((x-\cdot)/h\big), \varepsilon\big) &= \big\{\overline{K}\big((y-\cdot)/h\big) : y \in \mathbb{R}^d, \big\|\overline{K}\big((x-\cdot)/h\big) - \overline{K}\big((y-\cdot)/h\big)\big\|_\infty \leq \varepsilon\big\} \\ &\supset \big\{\overline{K}\big((y-\cdot)/h\big) : y \in \mathbb{R}^d, \|x-y\|_\infty \leq \varepsilon h/(2L)\big\}. \end{aligned} \tag{3.45}$$

Let $Q \supset D$ be a cube of side length $\mathrm{diam}(D) < \infty$ and choose for

$$\overline{n} := \Big(\Big\lfloor \frac{2L\mathrm{diam}(D)}{\varepsilon h}\Big\rfloor\Big)^d$$

points $x_1, \ldots, x_{\overline{n}} \in Q$ such that $\big\{B_{d_\infty}(x_i, \varepsilon h/(2L)) : i = 1, \ldots \overline{n}\big\}$ covers Q and therefore D. From (3.45), it follows that $\big\{B_{d_\infty}\big(\overline{K}\big((x_i-\cdot)/h\big), \varepsilon\big) : i = 1, \ldots \overline{n}\big\}$ is an external covering of $\mathcal{G}$. The external covering number $\mathcal{N}_{\mathrm{ext}}(\varepsilon, \mathcal{G}, d_\infty)$ is thus bounded by $(2L\mathrm{diam}(D)/(\varepsilon h))^d$. Hence,

$$\mathcal{N}(\varepsilon, \mathcal{G}, d_\infty) \leq \mathcal{N}_{\mathrm{ext}}(\varepsilon/2, \mathcal{G}, d_\infty) \leq \Big(\frac{4L\mathrm{diam}(D)}{\varepsilon h}\Big)^d.$$

Similarly, for

$$\widetilde{\mathcal{G}} = \{K(x-\cdot)/h) : x \in D \cap \mathbb{Q}^d\}, \tag{3.46}$$

we obtain

$$\mathcal{N}(\varepsilon, \widetilde{\mathcal{G}}, d_\infty) \leq \Big(\frac{2L\mathrm{diam}(D)}{\varepsilon h}\Big)^d.$$

The variance term is bounded by means of Propositions 3.1 and 3.6, respectively. In case $d = 1$ for $\boldsymbol{X} \in \boldsymbol{\Sigma}$ or any dimension for $\boldsymbol{X} \in \boldsymbol{\Theta}$, boundedness of ρ, Proposition 3.1 and (3.17) yield that, for $h \in (0,1)$ and some constant C independent of $\boldsymbol{\lambda}(\mathrm{supp}(K((x-\cdot)/h))) = h^d$,

$$\mathrm{Var}\left(\int_0^T K\left(\frac{x-X_t}{h}\right)\mathrm{d}t\right) \leq C(1 \vee c_{\widetilde{D}})T\|K\|_\infty^2\|\rho\|_\infty h^{2d}\psi_d^2(h^d),$$

where $\widetilde{D}$ is a compact set containing $D+[-1/2,1/2]^d$. Hence, for any dimension d and $\boldsymbol{X} \in \boldsymbol{\Sigma} \cup \boldsymbol{\Theta}$, we obtain together with Proposition 3.6 that there exists some global constant $\mathbb{A}$ independent of h such that for any $h \in (0,1)$, $t > 0$ and $g \in \widetilde{\mathcal{G}}$,

$$\mathrm{Var}\left(\frac{1}{\sqrt{t}}\int_0^t g(X_s)\,\mathrm{d}s\right) \leq \mathbb{A}\|g\|_\infty^2\|\rho\|_\infty h^{2d}\psi_d^2(h^d), \tag{3.47}$$

and hence

$$\|g\|_{\mathbb{G},t} \leq \sqrt{\mathbb{A}\|\rho\|_\infty}h^d\psi_d(h^d)\|g\|_\infty. \tag{3.48}$$

Consequently, with the first part of the proof we obtain

$$\begin{aligned}\mathcal{N}(\varepsilon,\mathcal{G},\|\cdot\|_{\mathbb{G},t}) = \mathcal{N}(\varepsilon,\widetilde{\mathcal{G}},\|\cdot\|_{\mathbb{G},t}) &\leq \mathcal{N}(\varepsilon(\sqrt{\mathbb{A}\|\rho\|_\infty}h^d\psi_d(h^d))^{-1},\widetilde{\mathcal{G}},\|\cdot\|_\infty)\\ &\leq \left(\frac{2L\mathrm{diam}(D)\sqrt{\mathbb{A}\|\rho\|_\infty}h^{d-1}\psi_d(h^d)}{\varepsilon}\right)^d.\end{aligned}$$

■

Proof of Lemma 3.9. Let $\boldsymbol{X} \in \boldsymbol{\Theta} \cup \boldsymbol{\Sigma}$. We start with bounding $\mathbb{E}[\sup_{x\in D}|\mathbb{H}_{h,T}(x)|^p]$. Let $m_T \in (0,T/4]$ and $\tau \in [m_T, 2m_T]$ as in Theorem 3.7. Using (3.48) and $\sup_{f,g\in\widetilde{\mathcal{G}}}\|f-g\|_\infty \leq 2\|K\|_\infty$ for $\widetilde{\mathcal{G}}$ defined in (3.46), we obtain

$$\sup_{f,g\in\widetilde{\mathcal{G}}}\|f-g\|_{\mathbb{G},\tau} \leq \sqrt{\mathbb{A}\|\rho\|_\infty}\sup_{f,g\in\widetilde{\mathcal{G}}}\|f-g\|_\infty h^d\psi_d(h^d) \leq 2\sqrt{\mathbb{A}\|\rho\|_\infty}\|K\|_\infty h^d\psi_d(h^d) =: \mathbb{V}(h), \tag{3.49}$$

such that $\mathcal{N}(u,\widetilde{\mathcal{G}},\|\cdot\|_{\mathbb{G},\tau}) = 1$ for $u \geq \mathbb{V}(h)$. Consequently, using the estimate $\int_0^C\sqrt{\log(M/u)}\,\mathrm{d}u \leq 4C\sqrt{\log(M/C)}$ provided $\log(M/C) \geq 2$, see e.g. p. 592 of Giné and Nickl [85], and the covering number bound from Lemma 3.19, it follows for $h \leq \mathrm{e}^{-2}L\mathrm{diam}(D)/\|K\|_\infty$ that

$$\begin{aligned}\int_0^\infty\sqrt{\log\mathcal{N}(u,\mathcal{G},d_{\mathbb{G},\tau})}\,\mathrm{d}u = \int_0^\infty\sqrt{\log\mathcal{N}(u,\widetilde{\mathcal{G}},d_{\mathbb{G},\tau})}\,\mathrm{d}u &\leq \int_0^{\mathbb{V}(h)}\sqrt{d\,\log\left(\frac{L\mathrm{diam}(D)\mathbb{V}(h)}{uh\|K\|_\infty}\right)}\,\mathrm{d}u\\ &\leq 2\mathbb{V}(h)\sqrt{d\log\left(\frac{L\mathrm{diam}(D)}{\|K\|_\infty h}\right)}.\end{aligned}$$

Moreover, since $\sup_{f,g\in\mathcal{G}}\|f-g\|_\infty \leq 4\|K\|_\infty$, it follows that $\mathcal{N}(u,\mathcal{G},d_\infty) = 1$ for all $u \geq 4\|K\|_\infty$ and hence we obtain by the covering number bound with respect to the sup-norm from Lemma 3.19 and elementary calculations

$$\int_0^\infty \log\mathcal{N}\big(u,\mathcal{G},\tfrac{2m_T}{\sqrt{T}}d_\infty\big)\,\mathrm{d}u = 2\tfrac{m_T}{\sqrt{T}}\int_0^{4\|K\|_\infty}\log\mathcal{N}(u,\mathcal{G},d_\infty)\,\mathrm{d}u \leq 8\tfrac{m_T}{\sqrt{T}}d\|K\|_\infty\left(1+\log\left(\frac{L\mathrm{diam}(D)}{\|K\|_\infty h}\right)\right).$$

Denseness of $\mathbb{Q}^d$ in $\mathbb{R}^d$, continuity of $x \mapsto \mathbb{H}_{h,T}(x)$ and Theorem 3.7 thus imply for $h \leq \mathrm{e}^{-2}L\mathrm{diam}(D)/\|K\|_\infty$

$$\begin{aligned}\left(\mathbb{E}\Big[\sup_{x\in D}|\mathbb{H}_{h,T}(x)|^p\Big]\right)^{1/p} &= \left(\mathbb{E}\Big[\sup_{x\in D\cap\mathbb{Q}^d}|\mathbb{H}_{h,T}(x)|^p\Big]\right)^{1/p}\\ &\leq \frac{1}{\sqrt{Th^d}}\Bigg(8\widetilde{C}_1\frac{m_T}{\sqrt{T}}d\|K\|_\infty\left(1+\log\left(\frac{L\mathrm{diam}(D)}{\|K\|_\infty h}\right)\right)+2\widetilde{C}_2\mathbb{V}(h)\sqrt{d\log\left(\frac{L\mathrm{diam}(D)}{\|K\|_\infty h}\right)}\\ &\quad + 16\frac{m_T}{\sqrt{T}}\|K\|_\infty\widetilde{c}_1 p + 2\mathbb{V}(h)\widetilde{c}_2\sqrt{p}+4\|K\|_\infty\sqrt{T}\Xi(m_T)^{1/p}\Bigg),\end{aligned} \tag{3.50}$$

for $\mathbb{V}(h)$ introduced in (3.49). Now, let $p = u_T \geq 1$ be such that $\Xi^{-1}(T^{-u_T}) \in \mathrm{o}(T)$. Then, for T large enough, (3.50), $h \geq T^{-2}$ and $h \in \mathrm{o}(1)$ imply for the choice $m_T = \Xi^{-1}(T^{-u_T})$ that

$$\begin{aligned}&\mathbb{E}\big[\|\widehat{\rho}_{h,T}-\mathbb{E}\widehat{\rho}_{h,T}\|_{L^\infty(D)}^{u_T}\big]\\ &\leq c^{u_T}\left(\frac{\log T}{Th^d}\Xi^{-1}(T^{-u_T})+T^{-\frac{1}{2}}\psi_d(h^d)\sqrt{\log(h^{-1})}+\frac{u_T}{Th^d}\Xi^{-1}(T^{-u_T})+T^{-\frac{1}{2}}\psi_d(h^d)\sqrt{u_T}+h^{-d}T^{-1}\right)^{u_T}\\ &\leq c^{u_T}\left(\frac{\log T+u_T}{Th^d}\Xi^{-1}(T^{-u_T})+T^{-\frac{1}{2}}\psi_d(h^d)\left(\sqrt{\log(h^{-1})}+\sqrt{u_T}\right)\right)^{u_T},\end{aligned}$$

where the value of the constant c changes from line to line. Hence Markov's inequality implies that there exists some constant $c^* > 0$ such that

$$\mathbb{P}\left(\|\widehat{\rho}_{h,T}-\mathbb{E}\widehat{\rho}_{h,T}\|_{L^\infty(D)} \geq c^*\left(\frac{u_T+\log T}{Th^d}\Xi^{-1}(T^{-u_T})+T^{-\frac{1}{2}}\psi_d(h^d)\sqrt{u_T\vee\log(h^{-1})}\right)\right) \leq \mathrm{e}^{-u_T}. \tag{3.51}$$

Suppose now that $\boldsymbol{X} \in \boldsymbol{\Sigma}$. Then, $\boldsymbol{X}$ is exponentially β-mixing, i.e., $\Xi(t) = c_\kappa \mathrm{e}^{-\kappa t}$, where without loss of generality we may assume that $c_\kappa \geq 1$. Then, for any $\gamma > 0$ and $1 \leq u_T \leq \gamma\log T$, it follows from $\Xi^{-1}(T^{-u_T}) \leq u_T\log T/\kappa$ and (3.51) that there exists some constant $c_\gamma > 0$ such that

$$\mathbb{P}\left(\|\widehat{\rho}_{h,T}-\mathbb{E}\widehat{\rho}_{h,T}\|_{L^\infty(D)} \geq c_\gamma\left(\frac{u_T(\log T)^2}{Th^d}+T^{-\frac{1}{2}}\psi_d(h^d)\sqrt{u_T\vee\log(h^{-1})}\right)\right) \leq \mathrm{e}^{-u_T}.$$

■

3.B PROOFS FOR SECTION 3.4.2

Proof of Lemma 3.15. We will employ Theorem 6.2.9 and Exercise 6.4.7 of [14] to show the first assertion. So first we must verify that condition **(C1)** on page 365 of [14] holds. Since $(\mathscr{J}1)$ holds, we only have to show that there exists a constant $K_1 > 0$ such that, for all $x, y \in \mathbb{R}^d$,

$$\sum_{i,j=1}^d(\sigma_{i,j}(x)-\sigma_{i,j}(y))^2+\int_{\mathbb{R}^d}\|\gamma(x)z-\gamma(y)z\|^2\,\nu(\mathrm{d}z) \leq K_1\|x-y\|^2,$$

where $\sigma_{i,j}(x)$ denotes the components of $\sigma(x) \in \mathbb{R}^{d\times d}$ for any $x \in \mathbb{R}^d$. ($\mathfrak{J}1$) implies that there exists a finite constant $L_{i,j} > 0$ for any $i, j \in \{1, \dots, d\}$, such that $\sigma_{i,j}\colon \mathbb{R}^d \to \mathbb{R}$ is Lipschitz continuous with Lipschitz constant $L_{i,j} > 0$ and hence we have for $x, y \in \mathbb{R}^d$

$$\sum_{i,j=1}^{d} (\sigma_{i,j}(x) - \sigma_{i,j}(y))^2 \le 2d \max_{i,j\in\{1,\dots,d\}} L_{i,j}^2 \|x - y\|^2.$$

Furthermore, we have for $x, y \in \mathbb{R}^d$ by the Lipschitz continuity of γ

$$\int_{\mathbb{R}^d} \|\gamma(x)z - \gamma(y)z\|^2 \,\nu(\mathrm{d}z) \le L_\gamma^2 \|x - y\|^2 \int_{\mathbb{R}^d} \|z\|^2 \,\nu(\mathrm{d}z),$$

where we denote the Lipschitz constant of γ by L_γ. By ($\mathfrak{J}3$), $\int_{\mathbb{R}^d} \|z\|^2 \,\nu(\mathrm{d}z)$ is finite and hence **(C1)** holds. To verify the growth condition **(C2)** on page 366 of [14], we have to show that there exists a constant K_2 such that, for all $x \in \mathbb{R}^d$,

$$\int_{\mathbb{R}^d} \|\gamma(x)z\|^2 \,\nu(\mathrm{d}z) \le K_2(1 + \|x\|^2).$$

Since γ is Lipschitz continuous by ($\mathfrak{J}1$), there exists a constant $K > 0$ such that the linear growth condition $\|\gamma(x)\| \le K(1 + \|x\|)$ holds for all $x \in \mathbb{R}^d$, and thus we have, for $x \in \mathbb{R}^d$,

$$\int_{\mathbb{R}^d} \|\gamma(x)z\|^2 \,\nu(\mathrm{d}z) \le 2K^2(1 + \|x\|^2) \int_{\mathbb{R}^d} \|z\|^2 \,\nu(\mathrm{d}z).$$

Again by ($\mathfrak{J}3$), $\int_{\mathbb{R}^d} \|z\|^2 \,\nu(\mathrm{d}z)$ is finite and hence **(C2)** holds for $K_2 = 2K^2 \int_{\mathbb{R}^d} \|z\|^2 \,\nu(\mathrm{d}z)$. Since Assumption 6.2.8 in [14] is trivially fulfilled, the first assertion follows by Theorem 6.2.9 and Exercise 6.4.7 of [14].
We proceed by showing the second assertion. Equation (1.21) of [47] is in the setting of (3.42) equivalent to $\kappa_\alpha(x, z) = \|\gamma(x)z\|^{d+\alpha}\nu(z) \ge 0$ for all $x \in \mathbb{R}^d$ and almost every $z \in \mathbb{R}^d$. Since ν is a density, this assumption is fulfilled.
For assumption ($\mathbf{H}^a$) of [47] to hold, we only need to show that there exists a $\beta \in (0, 1)$ such that the function $a(x) := \sigma(x)\sigma^\top(x)$ is β-Hölder continuous. However this follows directly from the Lipschitz continuity and the boundedness of σ imposed in ($\mathfrak{J}1$), as can be seen in the proof of Lemma 1 of [12]. Now we note that assumption ($\mathbf{H}^\kappa$) of [47] follows by ($\mathfrak{J}2$). ■

Proof of Corollary 3.16. Since ($\mathfrak{J}1$) and ($\mathfrak{J}3$) imply that b^* is bounded, arguing as in the proof of Lemma 1 of [12] and using Lemma 3.15 entails that b^* belongs to the Kato class $\mathbb{K}_2$ for $d \ge 2$. For the definition of $\mathbb{K}_2$, see (2.28) in [47]. Existence of transition densities and the heat kernel estimate now follow directly from Corollary 1.5 of [47] and Lemma 3.15 for $d \ge 2$ and as described in Lemma 1 of [12], the same conclusions may be drawn for dimension $d = 1$ by adapting the arguments in [47]. Now note that (3.44), $t \le 1$ and $\alpha \in (0, 2)$ imply

$$\begin{aligned} p_t(x, y) &\le C(t^{-d/2} \exp(-\|x - y\|^2/(\lambda t)) + \|\kappa_\alpha\|_\infty t(\|x - y\| + t^{1/2})^{-d-\alpha}) \\ &\le C(t^{-d/2} + t^{1-(d+\alpha)/2}) \le Ct^{-d/2}, \end{aligned}$$

where the value of C changes from line to line. This completes the proof. ■

Proof of Proposition 3.17. To verify the assertion, we show that the solution of (3.42) $\boldsymbol{X}$ satisfies the assumptions of Theorem 2.2 (ii) of [123] which are Assumption 1, 2(a)' and 3* of [123] and [124], respectively. Assumption 1 follows directly from ($\mathcal{J}$1). Now, define $b_u^*(x) := b^*(x) - \int_{u<\|z\|\leq 1}\gamma(x)z\,\nu(\mathrm{d}z) = b(x) - \int_{\|z\|>u}\gamma(x)z\,\nu(\mathrm{d}z)$, and let the diffusion process $Y^u = (Y_t^u)_{t\geq 0}$ be given by

$$Y_t^u = x + \int_0^t b_u^*(x)(Y_s^u)\,\mathrm{d}s + \int_0^t \sigma(Y_s^u)\,\mathrm{d}W_s.$$

For Assumption 2(a)' to be fulfilled, we first have to show that, for any $u \in (0,1)$, there exists $\Delta > 0$ such that $\mathbb{P}_x(Y_\Delta^u \in B) > 0$ for any $x \in \mathbb{R}^d$ and any nonempty open set $B \subset \mathbb{R}^d$. Since Y^u is a continuous diffusion process with bounded and Lipschitz coeffcients b_u^*, σ and $a = \sigma\sigma^T$ is uniformly elliptic, it follows from classical results, see e.g. [153, Theorem A], that for any $x \in \mathbb{R}^d$ and $t > 0$, the transition function $P_t^u(x,\cdot)$ of Y^u has a transition density with full support and hence any Δ-skeleton of Y^u is open set irreducible, showing that Assumtpion 2(a)' is in place. It remains to show that Assumption 3* of [123] is satisfied which is, that there exists a function $V \in Q^*$, where

$$Q^* := \Big\{ f\colon \mathbb{R}^d \to \mathbb{R}_+ : f \in \mathcal{C}^2, f(x) \to \infty \text{ as } \|x\| \to \infty, \text{ and there exists a locally bounded,}$$
$$\text{measurable function } \bar{f}, \text{ such that } \int_{\|z\|>1} f(x+\gamma(x)z)\,\nu(\mathrm{d}z) \leq \bar{f}(x),\ \forall x \in \mathbb{R}^d \Big\},$$

such that there are constants $c_1, c_2 > 0$, for which the Lyapunov drift criterion

$$\mathcal{A}V \leq -c_1 V + c_2 \tag{3.52}$$

holds, where $\mathcal{A}$ denotes the extended generator of $\boldsymbol{X}$ acting on Q^* by

$$\begin{aligned}\mathcal{A}f(x) =& \langle \nabla f(x), b^*(x)\rangle + \tfrac{1}{2}\operatorname{tr}(\nabla^2 f(x)\sigma(x)\sigma^T(x)) + \\ &\int_{\mathbb{R}^d} f(x+\gamma(x)z) - f(x) - \mathbb{1}_{\|z\|\leq 1}\langle \nabla f(x), \gamma(x)z\rangle \nu(\mathrm{d}z), \quad x \in \mathbb{R}^d, f \in Q^*.\end{aligned}$$

Now, for $\eta \in (0, \eta_0 c_\gamma^{-1} \wedge 1)$, where $c_\gamma := \|\gamma\|_\infty$, let V^η be a positive and increasing function in $\mathcal{C}^2(\mathbb{R}^d,\mathbb{R})$ such that $V^\eta = \mathrm{e}^{\eta\|x\|}$ for all $\|x\| > c_V$, where $c_V > 0$. Then, it holds for $i \neq j \in \{1,\ldots,d\}$ and $\|x\| > c_V$,

$$\begin{aligned}\partial_i V^\eta(x) &= \eta \mathrm{e}^{\eta\|x\|}\frac{x_i}{\|x\|}, \\ \partial_{ij}^2 V^\eta(x) &= \eta^2 \mathrm{e}^{\eta\|x\|}\frac{x_i x_j}{\|x\|^2} - \eta \mathrm{e}^{\eta\|x\|}\frac{x_i x_j}{\|x\|^3} + \eta \mathrm{e}^{\eta\|x\|}\|x\|^{-1}\delta_{ij},\end{aligned} \tag{3.53}$$

where δ_{ij} denotes the Kronecker delta. Furthermore, since $V^\eta \in \mathcal{C}^2(\mathbb{R}^d;\mathbb{R})$, for $i,j \in \{1,\ldots,d\}$ the functions $V^\eta, \partial_i V^\eta, \partial_{ij}^2 V^\eta$ are bounded by a constant $c_D > 0$ for $\|x\| \leq c_V$ and hence

$$\begin{aligned}\int_{\|z\|>1} V^\eta(x+\gamma(x)z)\,\nu(\mathrm{d}z) &\leq \int_{\|z\|>1} \left(\mathrm{e}^{\eta\|x+\gamma(x)z\|} + c_D\right)\nu(\mathrm{d}z) \\ &\leq \mathrm{e}^{\eta\|x\|}\int_{\|z\|>1} \mathrm{e}^{c_\gamma \eta\|z\|}\nu(\mathrm{d}z) + c_D \nu(\mathbb{R}^d \setminus B_1),\end{aligned}$$

implying that $V_{\mathbb{S}}^{\eta} \in Q^*$ for all $\eta \le \frac{\eta_0}{c_\gamma}$. This last condition is satisfied by our choice of η. To conclude the proof, the only thing left to show is that there exists $0 < \eta \le \frac{\eta_0}{c_\gamma}$ such that (3.52) holds for V^η. Note that, by the mean value theorem, the definition of b^* and the Cauchy–Schwarz inequality, we have for any $f \in Q^*$

$$
\begin{aligned}
&\mathcal{A}f(x)\\
&\quad = \langle \nabla f(x), b(x)\rangle + \tfrac{1}{2}\operatorname{tr}(\nabla^2 f(x)\sigma(x)\sigma^\top(x)) + \int_{\mathbb{R}^d} f(x+\gamma(x)z) - f(x) - \langle \nabla f(x), \gamma(x)z\rangle \nu(\mathrm{d}z)\\
&\quad \le \langle \nabla f(x), b(x)\rangle + \tfrac{1}{2}\operatorname{tr}(\nabla^2 f(x)\sigma(x)\sigma^\top(x))\\
&\qquad + \int_{\mathbb{R}^d} \sup_{t\in[0,1]} \|\nabla f(x+t\gamma(x)z) - \nabla f(x)\| \|\gamma(x)z\| \nu(\mathrm{d}z)\\
&\quad \le \mathcal{A}_c f(x) + \mathcal{A}_d f(x),
\end{aligned}
$$

where, for $\mathrm{H}^2 f(x)$ denoting the Hessian of f evaluated at x,

$$
\begin{aligned}
\mathcal{A}_c f(x) &:= \langle \nabla f(x), b(x)\rangle + \tfrac{1}{2}\operatorname{tr}(\nabla^2 f(x)\sigma(x)\sigma^T(x)),\\
\mathcal{A}_d f(x) &:= c_\gamma^2 \int_{\mathbb{R}^d} \sup_{t\in[0,1]} \|\mathrm{H}^2 f(x+t\gamma(x)z)\| \|z\|^2\, \nu(\mathrm{d}z).
\end{aligned}
$$

We start by investigating the jump part. By (3.53) and the fact that the operator norm can be bounded by the Frobenius norm $\|\cdot\|_F$, we get for $\|x\| > c_V$

$$
\begin{aligned}
\|\mathrm{H}^2 V^\eta(x)\| \le \|\mathrm{H}^2 \mathrm{e}^{\eta\|x\|}\|_F &= \left(\sum_{i,j=1}^d \left(\eta^2 \mathrm{e}^{\eta\|x\|}\frac{x_i x_j}{\|x\|^2} - \eta \mathrm{e}^{\eta\|x\|}\frac{x_i x_j}{\|x\|^3} + \eta \mathrm{e}^{\eta\|x\|}\|x\|^{-1}\delta_{ij}\right)^2\right)^{\frac{1}{2}}\\
&\le 2\eta \mathrm{e}^{\eta\|x\|}\left(\sum_{i,j=1}^d \left(\eta^2 \frac{x_i^2 x_j^2}{\|x\|^4} + \frac{x_i^2 x_j^2}{\|x\|^6} + \|x\|^{-2}\delta_{ij}\right)\right)^{\frac{1}{2}} \le 2^{3/2}\sqrt{d}\eta \mathrm{e}^{\eta\|x\|}\left(\eta^2 + 2\|x\|^{-2}\right)^{\frac{1}{2}}.
\end{aligned}
$$

Since we can choose c_V to be large, we can without loss of generality assume $c_V \ge \sqrt{2}\eta^{-1}$ and, additionally, $V^\eta \in \mathcal{C}^2$ implies that there exists a real-valued function $c_{\mathrm{H}}(\eta) > 0$ on $(0,\infty)$ such that $\|\mathrm{H}^2 V^\eta(x)\| < c_{\mathrm{H}}(\eta)$ for all $\|x\| \le c_V$. Thus, we have $\|\mathrm{H}^2 V^\eta(x)\| \le 4\sqrt{d}\eta^2 \mathrm{e}^{\eta\|x\|} + c_{\mathrm{H}}(\eta)$, $x \in \mathbb{R}^d$, and we can conclude

$$
\begin{aligned}
\mathcal{A}_d V^\eta(x) &\le 4c_\gamma^2\sqrt{d}\eta^2 \int_{\mathbb{R}^d} \sup_{t\in[0,1]} \mathrm{e}^{\eta\|x+t\gamma(x)z\|}\|z\|^2\nu(\mathrm{d}z) + c_\gamma^2 c_{\mathrm{H}}(\eta)\int_{\mathbb{R}^d}\|z\|^2\nu(\mathrm{d}z)\\
&\le \eta^2 \mathrm{e}^{\eta\|x\|}4c_\gamma^2\sqrt{d}\int_{\mathbb{R}^d}\mathrm{e}^{\eta_0\|z\|}\|z\|^2\nu(\mathrm{d}z) + c_\gamma^2 c_{\mathrm{H}}(\eta)\int_{\mathbb{R}^d}\|z\|^2\nu(\mathrm{d}z) =: c_{d,1}\eta^2\mathrm{e}^{\eta\|x\|} + c_{d,2}(\eta),
\end{aligned}
\tag{3.54}
$$

where $c_{d,1}, c_{d,2}(\eta)$ are positive and finite because of $(\mathfrak{J}3)$ and $\eta < \eta_0 c_\gamma^{-1}$. Now we turn our attention to the continuous part. From now on, without loss of generality, we assume that $c_V \ge c_1$ in $(\mathfrak{J}3)$. Then, for $\|x\| > c_V \ge \eta^{-1}$, we have by $(\mathfrak{J}1)$, $(\mathfrak{J}3)$ and (3.53)

$$
\mathcal{A}_c V^\eta(x) \le -c_1\eta \mathrm{e}^{\eta\|x\|} + \frac{c_2}{2}\sum_{k=1}^d \left|\eta^2 \mathrm{e}^{\eta\|x\|}\frac{x_i^2}{\|x\|^2} + \eta \mathrm{e}^{\eta\|x\|}\|x\|^{-1} - \eta \mathrm{e}^{\eta\|x\|}\frac{x_i^2}{\|x\|^3}\right|
$$

$$\leq \eta \mathrm{e}^{\eta\|x\|}\left(-c_1 + \frac{3c_2 d}{2}\eta\right),$$

and since $V^\eta \in \mathcal{C}^2(\mathbb{R}^d;\mathbb{R})$, there exists a real-valued function $c_c(\eta)$ on $(0,\infty)$ such that $\mathcal{A}_c V^\eta(x) \leq c_c(\eta)$ for all $\|x\| \leq c_V$. Hence, we have

$$\mathcal{A}_c V^\eta(x) \leq \eta \mathrm{e}^{\eta\|x\|}\left(-c_1 + \frac{3c_2 d}{2}\eta\right) + c_C(\eta) + c_1 \mathrm{e}^{c_V} =: \eta \mathrm{e}^{\eta\|x\|}\left(-c_1 + \frac{3c_2 d}{2}\eta\right) + c_{c,1}(\eta), \tag{3.55}$$

where we used that $\eta < 1$, by assumption. Combining (3.54) and (3.55) yields

$$\mathcal{A}V^\eta(x) \leq \eta \mathrm{e}^{\eta\|x\|}\left(-c_1 + \eta\left(\frac{3c_2 d}{2} + c_{d,1}\right)\right) + c_{d,2}(\eta) + c_{c,1}(\eta).$$

Choosing $\eta^* = 1 \wedge \eta_0 c_\gamma^{-1} \wedge \frac{c_1}{3c_2 d + 2c_{d,1}}$ implies

$$\mathcal{A}V^{\eta^*}(x) \leq -\frac{c_1\eta^*}{2}\mathrm{e}^{\eta\|x\|} + c_{d,2}(\eta^*) + c_{c,1}(\eta^*),$$

and thus (3.52) holds for $V^{\eta^*} \in Q^*$. Now, Theorem 2.2 (ii) and Proposition 3.8 of [123] show the required assertion. ■

Stability of overshoots of Markov additive processes

4

In this chapter we prove precise stability results for overshoots of *Markov additive processes*, MAPs in the following, with finite modulating space. Our approach is based on the Markovian nature of overshoots of MAPs whose mixing and ergodic properties are investigated in terms of the characteristics of the MAP. On our way we extend fluctuation theory of MAPs, contributing among others to the understanding of the Wiener–Hopf factorization for MAPs by generalizing Vigon's équations amicales inversés known for Lévy processes. Using the Lamperti transformation the results can be applied to self-similar Markov processes. Among many possible applications, we study the mixing behavior of stable processes sampled at first hitting times as a concrete example.

4.1 Introduction

4.1.1 *Background and motivation*

Overshoots of a Lévy process $\boldsymbol{\xi}$, defined by

$$\mathcal{O}_x = \xi_{T_x} - x, \quad x \geq 0,$$

on $\{T_x < \infty\}$, where $T_x := \inf\{t \geq 0 : \xi_t > x\}$, are classical objects in the study of Lévy processes. Their asymptotic analysis is essentially rooted in renewal theory for random walks and has gained a lot of interest in the past two decades starting with the observation in [24] that classical limit theorems for the residual time chain of renewal processes have a natural analogue in weak convergence of overshoots of subordinators to a non-trivial limiting distribution. Besides applications and extensions in ruin theory for insurance risk processes driven by Lévy processes (see [89, 105, 136]), this observation was used to explain the entrance behavior of positive self-similar Markov processes (pssMps) at the origin. Using the Lamperti transformation for transient pssMps one can show that a pssMp can be started from the origin if and only if the overshoots of the underlying Lévy process converge weakly as the overshoot level x diverges to $+\infty$ (see [28, 44]). This was generalized in [66] to the question of how to start real self-similar Markov processes (rssMps) from the origin. Methods for rssMps are similar to those for pssMps replacing the Lévy processes $\boldsymbol{\xi}$ in the Lamperti transformation by Markov additive processes $(\boldsymbol{\xi}, \boldsymbol{J})$ with finite modulating space $\{-1, 1\}$. The corresponding transformation is usually called Lamperti–Kiu transform. MAPs $(\boldsymbol{\xi}, \boldsymbol{J})$ are also called Markov modulated Lévy processes, due to the *ordinator* $\boldsymbol{\xi}$ behaving as a Lévy process in between jumps of a *modulating chain* $\boldsymbol{J}$, with the Lévy triplet of $\boldsymbol{\xi}$ being determined by the current state of $\boldsymbol{J}$. The limiting behavior of overshoots of MAPs, defined by

$$(\mathcal{O}_x, \mathcal{J}_x) = \big(\xi_{T_x} - x, J_{T_x}\big), \quad x \geq 0,$$

on $\{T_x < \infty\}$, where $T_x := \inf\{t \geq 0 : \xi_t > x\}$, then plays the same role for the entrance law at 0 of rssMps, as do overshoots of Lévy processes for pssMps.

The aim of this chapter is to explore in detail mixing and ergodicity of overshoots of MAPs. We study the convergence in total variation norm, including conditions for polynomial and exponential rates of convergence. Based on fluctuation theory of MAPs developed in [66] we will use the Meyn and Tweedie approach to stability of continuous time Markov processes (see for instance [127, 130, 132, 164]) to demonstrate that overshoot convergence can be much more finely analyzed once we take the perspective on overshoots as a Markov process, where the subsequent spatial levels that are passed by the ordinator $\boldsymbol{\xi}$ serve as time index for the overshoot process $(\mathbf{\mathcal{O}}, \mathbf{\mathcal{J}}) = (\mathcal{O}_t, \mathcal{J}_t)_{t\geq 0}$. This idea is inspired by the observation that for the overshoot process of a Lévy subordinator $\boldsymbol{\sigma}$, inverse local time at 0 is given by $\boldsymbol{\sigma}$ itself [26]. For this special case, this opens the door to powerful results of excursion theory for general Markov processes and allows, among others, to derive explicit formulas for the invariant measure and resolvent of the overshoot process of a Lévy subordinator in terms of its triplet [32, 83]. We generalize these findings to the MAP situation and consequently make use of the analytical tractability of overshoots to analyze their ergodic behavior. For the particular case of Lévy processes, the results can be interpreted as a natural continuous time generalization of results on ergodicity and exponential convergence of the residual time chain belonging to a renewal process, which can be found in the standard references on stability of discrete time Markov chains, Meyn and Tweedie [128] and Nummelin [134]. Extensions of renewal theory for random walks to discrete time MAPs (often called *Markov random walks*) were treated in [6, 43, 97, 102, 111] among others.

Our fine analysis of overshoot stability of MAPs is not only inspired by a theoretical desire to understand their asymptotics, but also by a practical need to develop statistical and numerical procedures to get hold of the *ascending ladder height process* $(\boldsymbol{H}^+, \boldsymbol{J}^+)$ of a given MAP $(\boldsymbol{\xi}, \boldsymbol{J})$. This process is one of the cornerstones of fluctuation theory of MAPs and is theoretically accessible by means of the Wiener–Hopf factorization. However, its explicit analytical characteristics are in general unknown, with a notable exception being the factorization of the MAP associated to an α-stable Lévy process via the Lamperti–Kiu transform, which was found in [108]. Due to its intimate connection with the running supremum of the MAP, observing $(\boldsymbol{\xi}, \boldsymbol{J})$ at first hitting times offers all information needed to determine $(\boldsymbol{H}^+, \boldsymbol{J}^+)$ in numerical or statistical procedures. For a recent account of fluctuation theory of Markov random walks we refer to [7].

The results of this chapter have applications in optimal control problems based on MAPs, see e.g. the recent article [49] for the more particular case of a Lévy driven impulse control problem. There, the generator of the ascending ladder height process is decisive for determining optimal threshold boundaries of a desired reflection strategy. Thus, under uncertainty concerning the underlying Lévy process, efficient statistical estimation of the ascending ladder height process is needed. This problem will be addressed in detail in Chapter 5.

Moreover, parametric estimation becomes feasible for the Lévy system of MAPs—which encodes the jumps of a MAP in analogy to the Lévy measure of a Lévy process—with explicit overshoot distributions based on the MAP observed at first hitting times $(T_{n\Delta})_{n\in\mathbb{N}_0}$ for some step size $\Delta > 0$. Such observation scheme can be described as *stochastic low frequency scheme* as opposed to deterministic low and high frequency schemes usually encountered in parametric inference of stochastic processes (see [22] for an overview in the context of Lévy processes) or the stochastic high-frequency scheme analyzed in [144] for Lévy processes. Furthermore, nonparametric statistical estimation procedures for the ascending ladder height characteristics can be developed based on our observation that under some natural conditions, the overshoot process is exponentially β*-mixing*. As demonstrated in Chapter 3, this property can serve as a central building block to nonparametric statistical analysis of non-reversible ergodic Markov processes. Hence, our results indicate how to include MAPs (which are non-ergodic) in an ergodic statistical setting by considering the space-time transform introduced in form of overshoots.

Due to recent applications of MAPs we also expect applications of our mixing estimates in other fields of probability theory such as planar maps (see for instance [27]). We highlight this point by making use of the the Lamperti–Kiu transform to translate the mixing behavior of MAPs into mixing bounds for self-similar Markov processes sampled at first hitting times. Further applications to nonparametric statistical estimation for MAPs, Lévy processes and equivalently self-similar Markov processes will be subject to future research.

4.1.2 *Overview and main result*

We start in Section 4.2 with formally introducing Markov additive processes and summarizing some results belonging to their fluctuation theory as given in [66]. We then proceed in Section 4.3 with the stability analysis of MAP overshoots, starting with the rigorous description of their Markovian nature and then studying important concepts from the theory of stability for Markov processes such as Harris recurrence, invariant measures and petite sets, which were introduced in Chapter 2. With this setup we come to our primary goal, the ergodicity analysis of overshoots. Our main results in this respect, taking also account of the developments in Section 4.4 described below, can be informally summarized as follows.

THEOREM. *Suppose that the MAP $(\boldsymbol{\xi}, \boldsymbol{J})$ is upward regular, $\boldsymbol{J}$ is irreducible and the ascending ladder height MAP $(\mathbf{H}^+, \boldsymbol{J}^+)$ has a finite first moment. Under mild assumptions on the Lévy system of $(\boldsymbol{\xi}, \boldsymbol{J})$, $(\mathcal{O}_t, \mathcal{J}_t)_{t \geq 0}$ converges in total variation to a unique stationary distribution, which encodes the characteristics of the ascending ladder height MAP. If moreover the jump measures associated to the MAP's Lévy system possess a common (exponential) moment, then the convergence takes place at (exponential) polynomial speed and overshoots are (exponentially) polynomially β-mixing.*

This will be made precise in a sequence of theorems in Section 4.3. In Theorem 4.19 we establish conditions on either the creeping probabilities of the subordinators associated to the ascending ladder height MAP or its Lévy system that guarantee total variation convergence of overshoots. Theorem 4.22 and Theorem 4.25 build on this result, giving exponential/polynomial ergodicity and the exponential/polynomial β-mixing property, respectively.

Section 4.4 is devoted to finding conditions on the Lévy system of the parent MAP, which imply the required assumptions on $(\boldsymbol{H}^+, \boldsymbol{J}^+)$ for the ergodic results of the previous section, thus enhancing significantly our understanding of asymptotics of MAP overshoots. The tool we develop for this purpose is an extension of Vigon‘s équations amicales inversés for Lévy processes given in [172] to MAPs. These equations analytically relate the Lévy systems of $(\boldsymbol{\xi}, \boldsymbol{J})$ and $(\boldsymbol{H}^+, \boldsymbol{J}^+)$, which makes inference of distributional properties of the ascending ladder height process based on the characteristics of the parent MAP possible.

In Section 4.5 we apply our β-mixing result for MAPs to real self-similar Markov processes sampled at first hitting times by exploiting the Lamperti–Kiu transform, which bridges these two classes of processes. As an even more specific application, we then consider the mixing behavior of α-stable Lévy processes and ergodicity of overshoots of the associated Lamperti-stable MAP.

Finally, for the convenience of the reader who is only interested in our results for Lévy processes and to provide an accessible source of reference for this case, we have devoted Appendix 4.B to a short summary of the main implications of our findings for overshoots of Lévy processes. This part can be studied independently of the rest of the chapter since we also briefly recall the main concepts needed from the theory of Lévy processes. Reading Chapter 3 and Appendix 4.B is therefore sufficient if the reader wants to move on directly to Chapter 5, which synthesizes our general statistical β-mixing framework with our contributions to stability theory of overshoots.

4.2 MARKOV ADDITIVE PROCESSES AND THEIR FLUCTUATION THEORY

We start with introducing Markov additive processes with finite modulating space. For the general theory of Markov additive processes the reader may consult the landmark papers of Cinlar [41, 42], a good start for the particular case of finite modulating space is [15, Chapter XI], and a focus on fluctuation theory is given in [66]. Let $\Theta = \{1, \ldots, n\}$ be a finite set and $(\mathbb{R} \times \Theta)_\vartheta$ be the Alexandrov one-point compactification of $\mathbb{R} \times \Theta$ with some isolated state $\vartheta = (\infty, \varpi)$. As usual, we extend a function $f \in \mathcal{B}(\mathbb{R} \times \Theta)$ to a function in $\mathcal{B}((\mathbb{R} \times \Theta)_\vartheta)$ by setting $f(\vartheta) = 0$, which will make notation more convenient. A (killed) Markov additive process (MAP) $(\boldsymbol{\xi}, \boldsymbol{J})$ with finite modulating space Θ is defined as a Feller process with state space $\mathbb{R} \times \Theta$ and cemetery state ϑ, having a possibly finite lifetime ζ and underlying stochastic base $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t \geq 0}, (\mathbb{P}^{x,i})_{(x,i) \in (\mathbb{R} \times \Theta)_\vartheta})$ and which moreover has the characteristic property that given $s, t \geq 0$, $(x, i) \in \mathbb{R} \times \Theta$ and $f \in \mathcal{B}_b((\mathbb{R} \times \Theta)_\vartheta)$ it holds that

$$\mathbb{E}^{x,i}\big[f(\xi_{t+s} - \xi_t, J_{t+s})\mathbb{1}_{\{t<\zeta\}} | \mathcal{F}_t\big] = \mathbb{E}^{0,J_t}[f(\xi_s, J_s)]\mathbb{1}_{\{t<\zeta\}}, \quad \mathbb{P}^{x,i}\text{-a.s.}$$

In other words, conditionally on $\{J_t = i\}$ and no killing before time $t \geq 0$, the pair $(\xi_{t+s} - \xi_t, J_{t+s})_{s \geq 0}$ is independent of the past and has the same distribution as $(\xi_s, J_s)_{s \geq 0}$ under $\mathbb{P}^{0,i}$, which is an equivalent definition for MAPs with finite modulating space often encountered in the literature such as [66]. A straightforward consequence of this property is conditional spatial homogeneity of the process, i.e.

$$\mathbb{E}^{x,i}[f(\boldsymbol{\xi}, \boldsymbol{J})] = \mathbb{E}^{0,i}[f(\boldsymbol{\xi} + x, \boldsymbol{J})]$$

holds for any measurable f on the Skorokhod space $\mathcal{D}(\mathbb{R} \times \Theta)$ of càdlàg functions mapping from $\mathbb{R}_+ = [0, \infty)$ to $\mathbb{R} \times \Theta$ equipped with its Borel σ-algebra (here and for the rest of the chapter we implicitly assume that $(\boldsymbol{\xi}, \boldsymbol{J})$ has exclusively càdlàg paths, which can be easily achieved by either constructing the process as the canonical coordinate process on the Skorokhod space or by a reduction of the probability space and the facts that, by definition, Feller processes have càdlàg paths almost surely and $\mathcal{F}$ is complete). Moreover, $(J_t)_{t \geq 0}$ is a continuous time Markov chain, whose transition function is independent of the initial distribution of $\boldsymbol{\xi}$. Conditional independence of increments and spatial homogeneity of the ordinator $\boldsymbol{\xi}$ already teases an intimate relation of MAPs and Lévy processes. In fact, any MAP can be decomposed into an independent sequence of Lévy processes, whose characteristic triplet depends on the current state of the modulating Markov chain $\boldsymbol{J}$.

More precisely, we suppose that the measurable space $(\Omega, \mathcal{F})$ is rich enough to support a probability measure $\mathbb{P}$ such that $\mathbb{P}^{x,i} = \mathbb{P}(\cdot | \xi_0 = x, J_0 = i)$, i.e. the probabilities underlying the Markov process $(\boldsymbol{\xi}, \boldsymbol{J})$ are given as regular conditional probabilites of $\mathbb{P}$. Then, Proposition 2 in [66] (see also [95, Proposition 2.5] or [42, Theorem 2.23]) gives the following characterization of a MAP, showing that in between jumps of $\boldsymbol{J}$, $\boldsymbol{\xi}$ behaves as a Lévy process with characteristic triplet determined by the current state of $\boldsymbol{J}$ and every jump of $\boldsymbol{J}$ potentially triggers an additional jump of $\boldsymbol{\xi}$.

Proposition 4.1. *A process $(\boldsymbol{\xi}, \boldsymbol{J})$ is an unkilled MAP if and only if there exist sequences of*

- *(killed) Lévy processes $(\xi^{n,i})_{n\in\mathbb{N}_0}$, i.i.d. under $\mathbb{P}$ for fixed $i \in \Theta$,*
- *real random variables $(\Delta^n_{i,j})_{n\in\mathbb{N}}$, i.i.d. under $\mathbb{P}$ for fixed and distinct $i, j \in \Theta$,*

independent of $\boldsymbol{J}$ and of each other under $\mathbb{P}$, such that if σ_n is the n-th jump time of $\boldsymbol{J}$, then under $\mathbb{P}^{x,i}$, $\boldsymbol{\xi}$ can be written almost surely as

$$\xi_t = \begin{cases} x + \xi^{0,i}_t, & t \in [0, \sigma_1), \\ \xi_{\sigma_n-} + \Delta^n_{J_{\sigma_n-}, J_{\sigma_n}} + \xi^{n,J_{\sigma_n}}_{t-\sigma_n}, & t \in [\sigma_n, \sigma_{n+1}), t < \zeta, \\ \xi_t = \infty, & t \geq \zeta, \end{cases}$$

where the lifetime ζ is the first time one of the appearing Lévy processes is killed:

$$\zeta = \inf\left\{t > 0 : \exists n \in \mathbb{N}_0, \sigma_n \leq t \text{ such that } \xi^{n,J_{\sigma_n}} \text{ is killed at time } t - \sigma_n\right\}.$$

In this chapter, we will only deal with MAPs $(\boldsymbol{\xi}, \boldsymbol{J})$ with *infinite lifetime*, i.e. $\zeta = \infty$, $\mathbb{P}^{x,i}$-a.s. for all $(x, i) \in \mathbb{R} \times \Theta$. However, killing is relevant for fluctuation theory of MAPs as described below. Let us define $(\xi^{(i)})_{i\in\Theta}$ as Lévy processes with characteristic triplets (a_i, b_i, Π_i) that have the same law as $(\xi^{0,i})_{i\in\Theta}$ and $(\Delta_{i,j})_{i,j\in\Theta}$ as random variables sharing the same law as the corresponding $(\Delta^1_{i,j})_{i,j\in\Theta}$, with $\Delta_{i,i} := 0$ for all $i \in \Theta$. Moreover, let $F_{i,j}$ be the law of $\Delta_{i,j}$. Then, $(\boldsymbol{\xi}, \boldsymbol{J})$ can be uniquely characterized by the Lévy–Khintchine exponents $\Psi_i(\theta) = \log \mathbb{E}[\exp(\mathrm{i}\theta\xi^{(i)}_1)], i \in \Theta$, the transition rate matrix $\boldsymbol{Q} = (q_{i,j})_{i,j\in\Theta}$ of $\boldsymbol{J}$ and the Fourier transforms of $\Delta_{i,j}$ denoted by $G_{i,j}(\theta) = \mathbb{E}[\exp(\mathrm{i}\Delta_{i,j})], i, j \in \Theta$. For convenience we assume $\Delta_{i,j} = 0$ whenever $q_{i,j} = 0$, which is without loss of generality because Proposition 4.1 shows that these transitional jumps never occur. If we now define the characteristic matrix exponent

$$\boldsymbol{\Psi}(\theta) := \operatorname{diag}(\Psi_1(\theta), \ldots, \Psi_n(\theta)) + \boldsymbol{Q} \odot \boldsymbol{G}(\theta),$$

as an analogue to the Lévy–Khintchine exponent of a Lévy process, then

$$\mathbb{E}^{0,i}\left[\mathrm{e}^{\mathrm{i}\theta\xi_t};\, J_t = j\right] = \left(\mathrm{e}^{t\boldsymbol{\Psi}(\theta)}\right)_{i,j}, \quad i, j \in \Theta, \theta \in \mathbb{R}.$$

Here, $\boldsymbol{G}(\theta) = (G_{i,j}(\theta))_{i,j\in\Theta}$ and $\odot$ denotes the Hadamard product, i.e. pointwise multiplication of matrices of the same dimension. Note that since $\Delta_{i,i} = 0$ we have $G_{i,i}(\theta) = 1$ for all $i \in \Theta$ and hence $(\boldsymbol{Q} \odot \boldsymbol{G}(\theta))_{i,i} = -q_{i,i}$. Let us also define the family of potential measures $(U_{i,j})_{i,j\in\Theta}$ given by

$$U_{i,j}(\mathrm{d}x) = \mathbb{E}^{0,i}\left[\int_0^\infty \mathbb{1}_{\{\xi_t\in\mathrm{d}x, J_t=j\}}\,\mathrm{d}t\right] = \int_0^\infty \mathbb{P}^{0,i}(\xi_t \in \mathrm{d}x, J_t = j)\,\mathrm{d}t, \quad x \in \mathbb{R}, i, j \in \Theta,$$

i.e., $U_{i,j}(A)$ measures the time $\boldsymbol{\xi}$ spends in A when started in i, while the modulator $\boldsymbol{J}$ is in state j. Another important concept in the theory of (general state space) Markov additive processes is the existence of a *Lévy system*, see Cinlar [41], which generalizes the notion of a Lévy measure and becomes explicit for MAPs with finite modulating space thanks to the path decomposition given in Proposition 4.1. We say that $(\boldsymbol{\Pi}, A)$, where $\boldsymbol{\Pi}$ is a kernel on $(\Theta, \mathcal{B}(\mathbb{R} \times \Theta))$ satisfying

$$\boldsymbol{\Pi}(i, \{(0, i)\}) = 0, \qquad \int_{\mathbb{R}} (1 \wedge |y|^2)\, \boldsymbol{\Pi}(i, \mathrm{d}y \times \{i\}) < \infty, \quad i \in \Theta,$$

and A is an increasing continuous additive functional of $(\boldsymbol{\xi}, \boldsymbol{J})$ such that for any $f \in \mathcal{B}_+(\Theta\times\mathbb{R}\times\Theta)$ and $(x,i) \in \mathbb{R}\times\Theta$,

$$\mathbb{E}^{0,i}\Big[\sum_{s\le t} f(J_{s-}, \Delta\xi_s, J_s)\mathbb{1}_{\{\Delta\xi_s\neq 0 \text{ or } J_{s-}\neq J_s\}}\Big] = \mathbb{E}^{0,i}\Big[\int_0^t A_s \int_{\mathbb{R}\times\Theta} \boldsymbol{\Pi}(J_s, \mathrm{d}x, \mathrm{d}y)\, f(J_s, x, y)\Big], \tag{4.1}$$

is a Lévy system for $(\boldsymbol{\xi}, \boldsymbol{J})$. Using Proposition 4.1 and results on expectations of functionals of Poisson random measures, see e.g. Theorem 2.7 in [109], one can demonstrate that $A_t = t \wedge \zeta$ and

$$\boldsymbol{\Pi}(i, \mathrm{d}y \times \{j\}) = \mathbb{1}_{\{i=j\}}\Pi_i(\mathrm{d}y) + \mathbb{1}_{\{i\neq j\}} q_{i,j} F_{i,j}(\mathrm{d}y), \quad i,j \in \Theta,$$

and thus for any $i \in \Theta$,

$$\begin{aligned}\mathbb{E}^{0,i}\Big[\sum_{s\le t} f(J_{s-}, \Delta\xi_s, J_s)\mathbb{1}_{\{\Delta\xi_s\neq 0 \text{ or } J_{s-}\neq J_s\}}\Big] &= \sum_{k=1}^n \Big(\mathbb{E}^{0,i}\Big[\int_0^t \int_{\mathbb{R}\setminus\{0\}} f(k,x,k)\mathbb{1}_{\{J_s=k\}}\,\Pi_k(\mathrm{d}x)\,\mathrm{d}s\Big] \\ &\quad + \sum_{j\neq k} q_{k,j}\mathbb{E}^{0,i}\Big[\int_0^t\int_{\mathbb{R}} f(k,x,j)\mathbb{1}_{\{J_s=k\}}\,F_{k,j}(\mathrm{d}x)\,\mathrm{d}s\Big]\Big) \\ &= \sum_{k=1}^n \int_0^t \mathbb{P}^{0,i}(J_s = k)\,\mathrm{d}s\,\Big(\int_{\mathbb{R}\setminus\{0\}} f(k,x,k)\,\Pi_k(\mathrm{d}x) \\ &\quad + \sum_{j\neq k} q_{k,j}\int_{\mathbb{R}} f(k,x,j)\,F_{k,j}(\mathrm{d}x)\Big).\end{aligned} \tag{4.2}$$

Since A is simply the uniform motion, we will also refer to just $\boldsymbol{\Pi}$ as the Lévy system for the remainder of this chapter. As remarked in [110], this can be generalized to the following identity for any predictable process $(Z_t)_{t\ge 0}$ and $g \in \mathcal{B}_+(\Theta\times\mathbb{R}\times\mathbb{R}\times\Theta)$:

$$\begin{aligned}&\mathbb{E}^{0,i}\Big[\sum_{s\le t} Z_s g(J_{s-}, \xi_{s-}, \xi_s, J_s)\mathbb{1}_{\{\Delta\xi_s\neq 0 \text{ or } J_{s-}\neq J_s\}}\Big] \\ &= \sum_{k=1}^n \Big(\mathbb{E}^{0,i}\Big[\int_0^t \mathrm{d}s\, Z_s \mathbb{1}_{\{J_s=k\}} \int_{\mathbb{R}\setminus\{0\}} \Pi_k(\mathrm{d}x)\, g(k, \xi_s, \xi_s + x, k)\Big] \\ &\quad + \sum_{j\neq k} q_{k,j}\mathbb{E}^{0,i}\Big[\int_0^t \mathrm{d}s\, Z_s\mathbb{1}_{\{J_s=k\}}\int_{\mathbb{R}} F_{k,j}(\mathrm{d}x)\, g(k,\xi_s,\xi_s+x,j)\Big]\Big).\end{aligned} \tag{4.3}$$

Let us now dive into fluctuation theory of MAPs, which in the form suited to our needs was developed in [66]. An essential tool for our upcoming analysis of the overshoots is the *ascending ladder MAP* $(H_t^+, J_t^+)_{t\ge 0}$, which is defined as follows (see the appendix of [66] for more details). Let $(\mathsf{L}_t^{(i)})_{t\ge 0}$ be a version of local time at the point $(0,i)$ for the strong Markov process $(\bar\xi_t - \xi_t, J_t)_{t\ge 0}$, where $\bar\xi_t := \sup_{s\le t}\xi_s$. Define then $\mathsf{L}_t := \sum_{i=1}^n \mathsf{L}_t^{(i)}$, which is a continuous additive functional of $(\bar\xi_t - \xi_t, J_t)_{t\ge 0}$, increasing almost surely on the set of times when $\boldsymbol{\xi}$ attains a new maximum.

With this at hand we define the ladder height process $(\boldsymbol{H}^+, \boldsymbol{J}^+)$ by the time change

$$(H_t^+, J_t^+) = \begin{cases} (\xi_{\mathsf{L}_t^{-1}}, J_{\mathsf{L}_t^{-1}}), & 0 \le t < \mathsf{L}_\infty, \\ \vartheta = (\infty, \varpi), & t \ge \mathsf{L}_\infty, \end{cases}$$

where $\mathsf{L}_t^{-1} := \inf\{s \geq 0 : \mathsf{L}_s > t\}$ is the right-continuous inverse of L. It can be shown that $(\boldsymbol{H}^+, \boldsymbol{J}^+)$ is a Markov additive subordinator with lifetime L_∞, i.e. a Markov additive process such that the ordinator $\boldsymbol{H}^+$ has increasing paths almost surely before killing. Moreover, $(\mathsf{L}_t^{-1})_{0\leq t<\infty}$ almost surely equals the ordered set of times, when $\boldsymbol{\xi}$ reaches a maximum and hence the closure of the range of $\boldsymbol{H}^+$ up to its lifetime is identical to that of the supremum process $\bar{\xi}$ almost surely. Denote by $\boldsymbol{H}^{+,(i)}$ the Lévy subordinators appearing in the decomposition of $(\boldsymbol{H}^+, \boldsymbol{J}^+)$ in the spirit of Proposition 4.1. The respective drifts and Lévy measures are denoted by d_i^+ and Π_i^+, the intensity matrix of $\boldsymbol{J}^+$ by $\mathbf{Q}^+ = (q_{i,j}^+)_{i,j\in\Theta}$ and the killing rates of $\boldsymbol{H}^{+,(i)}$ by $\dagger_i^+$, i.e., when $\dagger_i^+ > 0$, the lifetime ζ_i^+ of $\boldsymbol{H}^{+,(i)}$ is exponentially distributed with mean $1/\dagger_i^+$ and otherwise, for $\dagger_i^+ = 0$, $\zeta_i^+ = \infty$ almost surely. Note that the MAP subordinator $(\boldsymbol{H}^+, \boldsymbol{J}^+)$ is then uniquely characterized by its Laplace exponent, given as follows:

$$\boldsymbol{\Phi}^+(\theta) := \mathrm{diag}\big(\Phi_1^+(\theta), \ldots, \Phi_n^+(\theta)\big) - \boldsymbol{Q}^+ \odot \boldsymbol{G}^+(\theta), \quad \theta \geq 0, \tag{4.4}$$

where Φ_i^+ is the Laplace exponent of $\boldsymbol{H}^{+,(i)}$ and $\boldsymbol{G}^+(\theta) = (G_{i,j}^+(\theta))_{i,j\in\Theta} = (\mathbb{E}[\exp(-\theta\Delta_{i,j}^+)])_{i,j\in\Theta}$. It then holds that

$$\mathbb{E}^{0,i}\big[\exp(-\theta H_t^+);\ J_t^+ = j\big] = \big(\mathrm{e}^{-\boldsymbol{\Phi}^+(\theta)t}\big)_{i,j}, \quad t \geq 0, \theta \geq 0, i, j \in \Theta.$$

Let us also denote the family of potential measures of $(\boldsymbol{H}^+, \boldsymbol{J}^+)$ by $(U_{i,j}^+)_{i,j\in\Theta}$.

In analogy to the case for Lévy processes we also need the ascending ladder height process of the *dual* of the MAP $(\boldsymbol{\xi}, \boldsymbol{J})$, i.e. a MAP which has the same law as the time reversed MAP $(\boldsymbol{\xi}, \boldsymbol{J})$. As remarked in [66] the construction of the dual MAP is slightly more elaborate compared to the Lévy case, where the dual process is simply the negative of the original Lévy process, because we have to take care of time reversion of the ordinator $\boldsymbol{J}$. Suppose that $\boldsymbol{J}$ is irreducible–and hence ergodic thanks to its finite state space—and denote its invariant distribution by $\boldsymbol{\pi} = (\pi(i))_{i\in\Theta}$. Moreover, let

$$\widehat{q}_{i,j} = \frac{\pi(j)}{\pi(i)} q_{j,i}, \quad i, j \in \Theta,$$

which are the intensities of the time reversed modulating Markov chain $\boldsymbol{J}$ and let $\widehat{\boldsymbol{Q}} = (\widehat{q}_{i,j})_{i,j\in\Theta}$. Now let $(\widehat{\mathbb{P}}^{x,i})_{(x,i)\in\mathbb{R}\times\Theta}$ be a family of probability measures such that $(\boldsymbol{\xi}, \boldsymbol{J})$ has characteristic matrix exponent given by

$$\widehat{\boldsymbol{\Psi}}(\theta) = \big(\widehat{\mathbb{E}}^{0,i}\big[\exp(\mathrm{i}\theta\xi_1);\ J_1 = j\big]\big)_{i,j\in\Theta} = \mathrm{diag}(\psi_1(-\theta), \ldots, \psi_n(-\theta)) + \widehat{\boldsymbol{Q}} \odot \boldsymbol{G}(-\theta)^\top, \quad \theta \in \mathbb{R}.$$

Then indeed, under $\mathbb{P}^{0,\boldsymbol{\pi}} := \sum_{i=1}^n \pi(i)\mathbb{P}^{0,i}$, it holds that the time reversed process $(\xi_{(t-s)-} - \xi_t, J_{(t-s)-})_{0\leq s\leq t}$ is equal in law to $(\xi_s, J_s)_{s\leq t}$ under $\widehat{\mathbb{P}}^{0,\boldsymbol{\pi}}$, see Lemma 21 in [66]. Let $\boldsymbol{\Delta}_{\boldsymbol{\pi}} := \mathrm{diag}(\boldsymbol{\pi})$ and denote the matrix Laplace exponent of the ascending ladder height process of the dual process of $(\boldsymbol{\xi}, \boldsymbol{J})$, $(\boldsymbol{H}^-, \boldsymbol{J}^-)$ by $\boldsymbol{\Phi}^-$ and also the objects belonging to its Lévy system in the obvious way.[1] The key result for fluctuation theory of MAPs is the (spatial) Wiener–Hopf factorization given in Theorem 26 of [66], which states that up to pre-multiplication by a positive diagonal matrix corresponding to the scaling of local time at the supremum,

$$-\boldsymbol{\Psi}(\theta) = \boldsymbol{\Delta}_{\boldsymbol{\pi}}^{-1}\boldsymbol{\Phi}^-(\mathrm{i}\theta)^\top\boldsymbol{\Delta}_{\boldsymbol{\pi}}\boldsymbol{\Phi}^+(-\mathrm{i}\theta) = \boldsymbol{\Delta}_{\boldsymbol{\pi}}^{-1}\boldsymbol{\Psi}^-(-\theta)^\top\boldsymbol{\Delta}_{\boldsymbol{\pi}}\boldsymbol{\Psi}^+(\theta), \quad \theta \in \mathbb{R}, \tag{4.5}$$

[1] A word of caution at this point: $\boldsymbol{\Phi}^-$ is *not* the matrix exponent of the dual of the ascending ladder height MAP $(\boldsymbol{H}^+, \boldsymbol{J}^+)$. To not confuse the reader we will therefore withhold the temptation to denote the ascending ladder height process of the dual of $(\boldsymbol{\xi}, \boldsymbol{J})$ by $(\widehat{H}^+, \widehat{J}^+)$.

and thus gives a factorization of the characteristic matrix exponent $\boldsymbol{\Psi}$ of $(\boldsymbol{\xi}, \boldsymbol{J})$ in terms of the characteristic exponents $\boldsymbol{\Psi}^+$ and $\boldsymbol{\Psi}^-$ of the ascending ladder height processes of $(\boldsymbol{\xi}, \boldsymbol{J})$ and its dual, respectively. This identity is the key for understanding the interplay between the parent MAP $\boldsymbol{\xi}$ and the ladder height processes, which we will further explore in Section 4.4.

4.3 Stability analysis of overshoots of MAPs

In this section, we assume that the lifetime ζ of $(\boldsymbol{\xi}, \boldsymbol{J})$ is equal to ∞ on all of Ω. For $t \geq 0$ define the ordinator‘s $\boldsymbol{\xi}$ first hitting time T_t of the set (t, ∞) by

$$T_t := \inf\{s \geq 0 : \xi_s > t\}.$$

Note that by right-continuous paths of the process and right-continuity of the filtration $(\mathcal{F}_t)_{t\geq 0}$ underlying $(\boldsymbol{\xi}, \boldsymbol{J})$ this is a stopping time for the MAP. Set also

$$\bar{\xi}_\infty := \sup_{0\leq t<\infty} \xi_t.$$

We now define the process $(\mathcal{O}_t, \mathcal{J}_t)_{t\geq 0}$ by

$$(\mathcal{O}_t, \mathcal{J}_t) = \begin{cases} (\xi_{T_t} - t, J_{T_t}), & \text{if } t < \bar{\xi}_\infty, \\ \vartheta, & \text{if } t \geq \bar{\xi}_\infty, \end{cases} \qquad t \geq 0,$$

i.e. if the level t is smaller than the supremum of the process over its entire lifetime, then $\mathcal{O}_t$ corresponds to the overshoot of $\boldsymbol{\xi}$ over t and $\mathcal{J}_t$ is equal to the state of the modulator at first passage of t, whereas for $t \geq \bar{\xi}_\infty$ the process is sent to the cemetery state ϑ. An essential observation for our analysis is that $(\mathcal{O}_t, \mathcal{J}_t)_{t\geq 0}$ is indistinguishable with respect to the family of probability measures $(\mathbb{P}^{x,i})_{(x,i)\in(\mathbb{R}_+\times\Theta)_\vartheta}$ from the process $(\mathcal{O}^+_t, \mathcal{J}^+_t)_{t\geq 0}$ corresponding to the ascending ladder MAP $(\boldsymbol{H}^+, \boldsymbol{J}^+)$, and hence is given by

$$(\mathcal{O}^+_t, \mathcal{J}^+_t) = \begin{cases} (H^+_{T^+_t} - t, J^+_{T^+_t}), & \text{if } t < \overline{H}^+_\infty, \\ \vartheta, & \text{if } t \geq \overline{H}^+_\infty, \end{cases} \qquad t \geq 0,$$

where $(T^+_t)_{t\geq 0}$ is the first passage process of $\boldsymbol{H}^+$, which by increasing paths of $\boldsymbol{H}^+$ is equal to its right-continuous inverse. Indistinguishability of the processes follows immediately from the fact that on $[0, \mathrm{L}_\infty)$, the range of the increasing process $(\mathrm{L}^{-1}_t)_{t\geq 0}$ almost surely equals the set of times when $\boldsymbol{\xi}$ reaches a maximum. Using this relationship, (4.3) and arguing as in the classical proof for the law of the undershoot/overshoot distribution for Lévy processes (see [109, Theorem 5.6]), we obtain the following formula for the marginal distribution of the overshoot process

$$\begin{aligned} \mathbb{P}^{x,i}(\mathcal{O}_t \in \mathrm{d}y, \mathcal{J}_t = j) &= \mathbb{P}^{0,i}(\mathcal{O}^+_{t-x} \in \mathrm{d}y, \mathcal{J}^+_t = j) \\ &= \int_{[0,t-x)} \Pi^+_j(u + \mathrm{d}y)\, U^+_{i,j}(t - x - \mathrm{d}u) \\ &\quad + \sum_{k\neq j} q^+_{k,j} \int_{[0,t-x)} F^+_{k,j}(u + \mathrm{d}y)\, U^+_{i,k}(t - x - \mathrm{d}u), \qquad i, j \in \Theta, x \in [0, t), y \geq 0, \end{aligned} \tag{4.6}$$

and

$$\mathbb{E}^{x,i}[f(\mathcal{O}_t,\mathcal{J}_t)] = f(x-t,i), \quad x \in [t,\infty), i \in \Theta, y \geq 0, \tag{4.7}$$

provided that $\mathbb{P}^{0,i}(T_0^+ = 0) = 1$. Assumption ($\mathscr{A}0$) introduced below will ensure this property. Equation (4.7) describes the characteristic behavior of the overshoot process away from 0 in the sense that if $\mathcal{O}_t(\omega) = y > 0$ we have $\mathcal{O}_s(\omega) = y - (s-t)$ for $s \in [t, t+y]$, i.e. the origin is approached at unit speed. This characteristic path structure of the overshoot process is visualized in Figure 4.1 for the case of a compound Poisson subordinator $\boldsymbol{\sigma}$ with positive drift, and is the reason why for such Lévy subordinators the overshoot process is also known as sawtooth process, cf. Chapter II.3 in [32]. We will therefore also refer to it as the sawtooth structure for MAP overshoots.

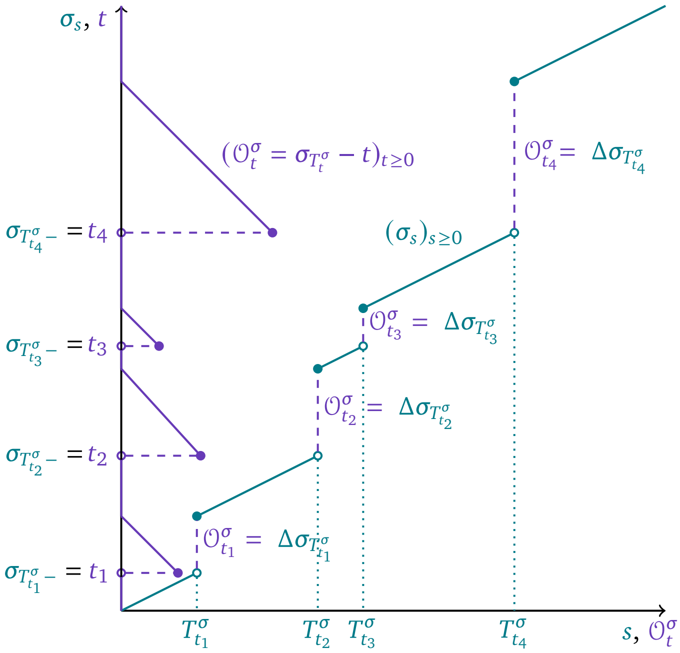

Figure 4.1: Path of a compound Poisson subordinator with drift, $\boldsymbol{\sigma}$, and associated overshoot process $\boldsymbol{\mathcal{O}}^\sigma$

Let $\mathcal{G}_t := \mathcal{F}_{T_t}$ for $t \geq 0$ and define the filtration $\mathbb{G} := (\mathcal{G}_t)_{t\geq 0}$. The following technical results hold.

Lemma 4.2. *$\mathbb{G$ is right-continuous.*

Proof. First note that $\mathcal{F}_{T_t} = \mathcal{F}_{T_t+}$, with

$$\mathcal{F}_{T_t+} := \left\{\Lambda \in \mathcal{F} : \Lambda \cap \{T_t < s\} \in \mathcal{F}_s \text{ for all } s \geq 0\right\}$$

since the latter can be shown to be equal to

$$\left\{\Lambda \in \mathcal{F} : \Lambda \cap \{T_t \leq s\} \in \mathcal{F}_{s+} \text{ for all } s \geq 0\right\}$$

which in turn equals $\mathcal{F}_{T_t}$ thanks to right-continuity of $\mathbb{F}$. Letting $\Lambda \in \mathcal{G}_{t+} = \bigcap_{n\in\mathbb{N}} \mathcal{F}_{T_{t+1/n}}$ we obtain by right-continuity of $t \mapsto T_t$ that for any $s \geq 0$

$$\Lambda \cap \{T_t < s\} = \bigcup_{n\in\mathbb{N}} \Lambda \cap \left\{T_{t+\frac{1}{n}} < s\right\} \in \mathcal{G}_s,$$

since any set in the right-hand union belongs to $\mathcal{G}_s$ thanks to $\mathcal{F}_{T_{t+1/n}} = \mathcal{F}_{T_{(t+1/n)+}}$. It follows that $\Lambda \in \mathcal{F}_{T_t+} = \mathcal{F}_{T_t} = \mathcal{G}_t$, which proves right-continuity of $\mathbb{G}$. ∎

COROLLARY 4.3. *For any $0 \leq s \leq \infty$ the running supremum $\bar{\xi}_s$ is a stopping time with respect to $\mathbb{G}$. In particular, the lifetime $\bar{\xi}_\infty$ of $(\mathcal{O}_t, \mathcal{J}_t)_{t\geq 0}$ is a $\mathbb{G}$-stopping time.*

Proof. Let $s \in [0, \infty]$. For any $t \geq 0$

$$\{\bar{\xi}_s < t\} = \{T_t > s\} \in \mathcal{F}_{T_t} = \mathcal{G}_t,$$

which implies $\{\bar{\xi}_s \leq t\} \in \mathcal{G}_{t+}$ and since $\mathcal{G}_{t+} = \mathcal{G}_t$ by Lemma 4.2 we conclude $\{\bar{\xi}_s \leq t\} \in \mathcal{G}_t$. ∎

We now show that under a technical assumption, the overshoot process given by the quintuple $(\Omega, \mathcal{F}, \mathbb{G}, (\mathcal{O}_t, \mathcal{J}_t)_{t\geq 0}, (\mathbb{P}^{x,i})_{(x,i)\in(\mathbb{R}_+\times\Theta)_\vartheta})$ determines a Feller process and therefore also a Borel right process. The technical assumption under which we will be working throughout the rest of the chapter without further mention, is the following.

($\mathscr{A}$0) The MAP $(\boldsymbol{\xi}, \boldsymbol{J})$ is upward regular, i.e. for any $i \in \Theta$ it holds that $\mathbb{P}^{0,i}(T_0 = 0) = 1$.

By definition, $(\boldsymbol{\xi}, \boldsymbol{J})$ is upward regular if, independently of the starting point of the modulator $\boldsymbol{J}$, $\boldsymbol{\xi}$ started from 0 immediately hits the upper half line. By the path decomposition given in Proposition 4.1, this is the case if and only if the underlying Lévy processes $\boldsymbol{\xi}^{(i)}$ are regular upward for any $i \in \Theta$. Upward regularity for Lévy processes is completely understood, see the full characterization given in Theorem 6.5 of [109], and hence upward regularity of the MAP can be characterized by properties of its underlying Lévy processes. Moreover, by the general theory on local times of Markov processes, see e.g. Chapter 4 in Bertoin [25] or the landmark paper Blumenthal and Getoor [30], it follows that upward regularity implies that for each $i \in \Theta$, the local time $\mathsf{L}^{(i)}$ of $(\bar{\xi} - \xi, J)$ at $(0, i)$ is almost surely continuous and hence $\mathsf{L} = \sum_{i=1}^{n} \mathsf{L}^{(i)}$ is almost surely continuous as well. Hence, the right-continuous inverse $(\mathsf{L}_t^{-1})_{t\geq 0}$, corresponding to the set of times when a new maximum of $\boldsymbol{\xi}$ is reached, is strictly increasing on $[0, \mathsf{L}_\infty)$ almost surely and it follows that $\boldsymbol{H}^+$ is strictly increasing up to its lifetime. This property is essential for $(\mathfrak{O}, \mathfrak{J})$ being a Feller process, as the proof of the following proposition shows.

PROPOSITION 4.4. $(\mathfrak{O}, \mathfrak{J})$ *is a càdlàg Feller process with lifetime $\bar{\xi}_\infty$.*

Proof. Càdlàg paths of the process are a direct consequence of càdlàg paths of $(\boldsymbol{\xi}, \boldsymbol{J})$ and the fact that $t \mapsto T_t$ is right-continuous on $[0, \infty)$ and increasing on $[0, \bar{\xi}_\infty)$. Let now $f \in \mathcal{B}_b((\mathbb{R}_+ \times \Theta)_\vartheta)$ and $(x, i) \in (\mathbb{R}_+ \times \Theta)_\vartheta$. Recalling that $\bar{\xi}_\infty$ is a $\mathbb{G}$-stopping time and using $T_{t+s} = T_t + T_{t+s} \circ \theta_{T_t}$, on $\{T_t < \infty\}$, where $(\theta_t)_{t\geq 0}$ are the transition opertors of $(\boldsymbol{\xi}, \boldsymbol{J})$, it follows that $\mathbb{P}^{x,i}$-a.s.

$$\begin{aligned}
\mathbb{E}^{x,i}[f(\mathcal{O}_{t+s}, \mathcal{J}_{t+s})|\mathcal{G}_t] &= \mathbb{E}^{x,i}\left[f\left(\xi_{T_{t+s}} - (t+s), J_{T_{t+s}}\right) \circ \theta_{T_t}|\mathcal{F}_{T_t}\right]\mathbb{1}_{\{t<\bar{\xi}_\infty\}} + f(\vartheta)\mathbb{1}_{\{t\geq\bar{\xi}_\infty\}} \\
&= \mathbb{E}^{\xi_{T_t}, J_{T_t}}\left[f\left(\xi_{T_{t+s}} - (t+s), J_{T_{t+s}}\right)\right]\mathbb{1}_{\{t<\bar{\xi}_\infty\}} + f(\vartheta)\mathbb{1}_{\{t\geq\bar{\xi}_\infty\}} \\
&= \mathbb{E}^{\xi_{T_t}-t, J_{T_t}}\left[f\left(\xi_{T_s} - s, J_{T_s}\right)\right]\mathbb{1}_{\{t<\bar{\xi}_\infty\}} + f(\vartheta)\mathbb{1}_{\{t\geq\bar{\xi}_\infty\}}
\end{aligned}$$

$$= \mathbb{E}^{\mathcal{O}_t, \mathcal{J}_t}\big[f\big(\mathcal{O}_s, \mathcal{J}_s\big)\big].$$

Here, we used the strong Markov property of $(\boldsymbol{\xi}, \boldsymbol{J})$ for the second and spatial homogeneity of $\boldsymbol{\xi}$ for the third equality. This proves the Markov property of $(\mathbf{O}, \mathbf{J})$. Moreover, for $x > 0$ and $i \in \Theta$ we have $\mathbb{P}^{x,i}(T_0 = 0) = 1$ and by upward regularity of $\boldsymbol{\xi}$ we also have $\mathbb{P}^{0,i}(T_0 = 0) = 1$. Thus, $\mathbb{P}^{x,i}(\mathcal{O}_0, \mathcal{J}_0) = (x, i)$ for any $(x, i) \in (\mathbb{R}_+ \times \Theta)_\vartheta$, i.e. the process is a normal Markov process and its lifetime is given by $\bar{\xi}_\infty$ by construction. Let $(\mathcal{P}_t)_{t \geq 0}$ be its sub-Markov transition semigroup, i.e.

$$\mathcal{P}_t f(x, i) = \mathbb{E}^{x,i}[f(\mathcal{O}_t, \mathcal{J}_t);\ t < \bar{\xi}_\infty], \quad (x, i) \in (\mathbb{R}_+ \times \Theta)_\vartheta, f \in \mathcal{B}_b((\mathbb{R}_+ \times \Theta)_\vartheta).$$

Let us check the Feller property. Let $f \in \mathcal{C}_0(\mathbb{R}_+ \times \Theta)$. Since Θ is finite and recalling our convention that $f(\vartheta) = 0$, it suffices to show for fixed $i \in \Theta$ that $x \mapsto \mathcal{P}_t f(x, i) = \mathbb{E}^{x,i}[f(\mathcal{O}_t, \mathcal{J}_t)]$ is continuous to prove that $(x, i) \mapsto \mathbb{E}^{x,i}[f(\mathcal{O}_t, \mathcal{J}_t)]$ is continuous. If $x > t$ this is obvious. For $x \leq t$ let first $y \uparrow x$. By right-continuity of $t \mapsto (\mathcal{O}_t, \mathcal{J}_t)$, continuity and boundedness of f, dominated convergence and conditional spatial homogeneity of $(\boldsymbol{\xi}, \boldsymbol{J})$, it follows that

$$\lim_{y \uparrow x} \mathbb{E}^{y,i}[f(\mathcal{O}_t, \mathcal{J}_t)] = \lim_{y \uparrow x} \mathbb{E}^{0,i}[f(\mathcal{O}_{t-y}, \mathcal{J}_{t-y})] = \mathbb{E}^{0,i}[f(\mathcal{O}_{t-x}, \mathcal{J}_{t-x})] = \mathbb{E}^{x,i}[f(\mathcal{O}_t, \mathcal{J}_t)],$$

showing left-continuity of $x \mapsto \mathbb{E}^{x,i}[f(\mathcal{O}_t, \mathcal{J}_t)]$. To show right-continuity, note that for $y \downarrow x$ it holds that T^+_{t-y} increases to $\inf\{s \geq 0 : H^+_s \geq t - x\}$ on $\{T^+_{t-x} < \infty\}$ and since $\boldsymbol{H}^+$ is strictly increasing up to its lifetime by upward regularity of $\boldsymbol{\xi}$, it follows that the latter hitting time is almost surely equal to T^+_{t-x}. Since $(\boldsymbol{H}^+, \boldsymbol{J}^+)$ as a Feller process is quasi-left-continuous, it therefore follows that on $\{T^+_{t-x} < \infty\}$,

$$\lim_{y \downarrow x} \left(H^+_{T^+_{t-y}}, J^+_{T^+_{t-y}}\right) = \left(H^+_{T^+_{t-x}}, J^+_{T^+_{t-x}}\right), \quad \mathbb{P}^{0,i}\text{-a.s.}$$

By indistinguishability of $(\mathbf{O}^+, \mathbf{J}^+)$ and $(\mathbf{O}, \mathbf{J})$ we therefore obtain

$$\lim_{y \downarrow x} \mathbb{E}^{y,i}[f(\mathcal{O}_t, \mathcal{J}_t)] = \lim_{y \downarrow x} \mathbb{E}^{0,i}[f(\mathcal{O}^+_{t-y}, \mathcal{J}^+_{t-y})] = \mathbb{E}^{0,i}[f(\mathcal{O}^+_{t-x}, \mathcal{J}^+_{t-x})] = \mathbb{E}^{x,i}[f(\mathcal{O}_t, \mathcal{J}_t)],$$

proving also right-continuity of $(x, i) \mapsto \mathcal{P}_t f(x, i)$. Since moreover Θ is compact and for fixed $i \in \Theta$,

$$\lim_{x \to \infty} \mathcal{P}_t f(x, i) = \lim_{x \to \infty} f(x - t, i) = 0$$

thanks to $f \in \mathcal{C}_0(\mathbb{R}_+ \times \Theta)$, we conclude that $\mathcal{P}_t \mathcal{C}_0(\mathbb{R}_+ \times \Theta) \subset \mathcal{C}_0(\mathbb{R}_+ \times \Theta)$. Finally, for fixed $(x, i) \in \mathbb{R}_+ \times \Theta$ (again applying to upward regularity in case $x = 0$) it follows from $T_t \to 0$ a.s. as $t \downarrow 0$ and dominated convergence, that $\mathcal{P}_t f(x, i) \to \mathcal{P}_0 f(x, i) = f(x, i)$. This is enough to show that $(\mathcal{P}_t)_{t \geq 0}$ is a Feller semigroup, as discussed in Appendix 2.

It remains to check right-continuity and completeness of $\mathbb{G}$. Right-continuity was shown in Lemma 4.2. Moreover, it can be easily seen that the $\mathbb{P}^{x,i}$-augmentation of $\mathcal{F}_{T_t}$ is equal to $\mathcal{F}_{T_t}$ itself, since $\mathbb{F}$ is $\mathbb{P}^{x,i}$-augmented already, see also p.36 of [31]. This finishes the proof. ■

Having established the Markovian nature of the overshoot process, we now proceed by investigating its stability properties and long-time behavior. We must therefore restrict to the case, when the overshoot process is almost surely unkilled, which is the case if and only if $\sup_{0 \leq s < \infty} \xi_s = \infty$, $\mathbb{P}^{0,i}$-a.s. for all $i \in \Theta$. As for Lévy processes, there is a dichotomy concerning the long-time behavior of the ordinator $\boldsymbol{\xi}$, namely that exactly one of the following cases can occur:

(a) for any $(x, i) \in \mathbb{R} \times \Theta$, $\limsup_{t\to\infty} \xi_t = \infty$, $\mathbb{P}^{x,i}$-almost surely, and in this case either $\liminf_{t\to\infty} \xi_t = -\infty$ or $\lim_{t\to\infty} \xi_t = \infty$, $\mathbb{P}^{x,i}$-a.s.;

(b) for any $(x, i) \in \mathbb{R} \times \Theta$, $\lim_{t\to\infty} \xi_t = -\infty$, $\mathbb{P}^{x,i}$-almost surely.

When $\boldsymbol{J}$ is irreducible and the MAP's ordinator possesses an exponential moment, which of these cases occurs for a given MAP is determined by a Perron–Frobenius type eigenvalue of the MAP's Laplace exponent, see Asmussen [15, Proposition XII.2.10]. We will therefore henceforth work under the following additional assumption, which guarantees that $(\mathcal{O}, \mathfrak{J})$ is an unkilled Borel right Markov process and therefore gives us access to the theory of stability for Markov processes from Chapter 2.

($\mathscr{A}1$) For any $(x, i) \in (\mathbb{R} \times \Theta)$ it $\mathbb{P}^{x,i}$-almost surely holds $\limsup_{t\to\infty} \xi_t = \infty$.

Let us give the following definition.

DEFINITION 4.5. Let $\boldsymbol{A} = (a_{i,j})_{i,j=1,\dots,n} \in \mathbb{R}^{n\times n}$ be a matrix with $a_{i,j} \geq 0$ for any $i \neq j$. We say that $\boldsymbol{A}$ is *irreducible*, if for any $i \neq j$ there exists $(a_{i_k,i_{k+1}})_{k=0,\dots,m-1}$ for some $m \in \mathbb{N}$ with $i_0 = i$, $i_m = j$ such that $\prod_{k=0}^{m-1} a_{i_k,i_{k+1}} > 0$. An irreducible matrix $\widetilde{\boldsymbol{A}} = (\widetilde{a}_{i,j})_{i,j=1,\dots,n}$ such that $\operatorname{diag}(\boldsymbol{A}) = \operatorname{diag}(\widetilde{\boldsymbol{A}})$ and $\widetilde{a}_{i,j} \in \{a_{i,j}, 0\}$ for any $i \neq j$ is said to be a *minimal irreducible version* of an irreducible matrix $\boldsymbol{A}$, if any matrix obtained from $\widetilde{\boldsymbol{A}}$ by setting some off-diagonal element to 0 is not irreducible anymore.

If we visualize a matrix $\boldsymbol{A}$ as in the definition above as a directed graph with vertices $V = \{1, \dots, n\}$ representing the on-diagonal elements of $\boldsymbol{A}$ and edges $E = \{(i, j) : a_{i,j} > 0\}$ representing the non-zero off-diagonal elements of $\boldsymbol{A}$, irreducibility of $\boldsymbol{A}$ is equivalent to connectedness of the graph of $\boldsymbol{A}$. The graph of a minimal irreducible version $\widetilde{\boldsymbol{A}}$ of an irreducible matrix $\boldsymbol{A}$ is therefore a minimal connected subgraph of the graph of $\boldsymbol{A}$ with $\widetilde{V} = V$ and $\widetilde{E} \subset E$. Also note that a continuous time Markov chain is irreducible if and only if its Q-matrix is irreducible.

As a minimal requirement for stability we need to ensure irreducibility of the Markov process $(\mathcal{O}, \mathfrak{J})$. We therefore introduce the following assumption.

($\mathscr{A}2$) The modulator $\boldsymbol{J}^+$ of the ascending ladder MAP is irreducible, i.e., $\boldsymbol{Q}^+$ is an irreducible matrix.

For general MAPs irreducibility of $\boldsymbol{J}$ does not necessarily entail irreducibility of $\boldsymbol{J}^+$, with the latter property implying that $\boldsymbol{\xi}$ can reach a maximum in any phase of $\boldsymbol{J}$. E.g., if one of the Lévy components $\boldsymbol{\xi}^{(i)}$ is a negative subordinator and $\Delta_{j,i} < 0$ for any $j \in \Theta$, $\boldsymbol{J}^+$ is not irreducible since $\boldsymbol{\xi}$ never reaches a new maximum when its phase is i. However, the following result shows that irreducibility of $\boldsymbol{J}^+$ is given for a wide range of MAPs with irreducible modulator $\boldsymbol{J}$. To give one particular example covered by Proposition 4.6 below, suppose that for any $j \in \Theta$ the Lévy component $\boldsymbol{\xi}^{(j)}$ is neither a negative subordinator nor spectrally negative with bounded variation, or, when this fails for some $j \in \Theta$ this is compensated for by some unbounded transitional jump of $\boldsymbol{\xi}$ when $\boldsymbol{J}$ switches to j. Then, $\boldsymbol{J}^+$ is irreducible whenever $\boldsymbol{J}$ is irreducible and Assumption ($\mathscr{A}1$) is in place. We emphasize that upward regularity ($\mathscr{A}0$) is not needed for the statement of Proposition 4.6. Recall that for any measure ν on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, the support $\operatorname{supp}(\nu)$ is defined as the set of points $x \in \mathbb{R}$ such that for any open neighborhood U_x of x it holds $\nu(U_x) > 0$.

PROPOSITION 4.6. *Suppose that $\boldsymbol{J}$ is irreducible and ($\mathscr{A}1$) holds.*

(i) Introduce the following conditions for $j \in \Theta$:

($\mathscr{H}(j)$) $\boldsymbol{\xi}^{(j)}$ is of unbounded variation or $\operatorname{supp}(\Pi_j) \cap (0,\infty) \neq \varnothing$*;*

($\mathscr{I}(j)$) there exists $k \neq j$ such that $\operatorname{supp}(q_{k,j}F_{k,j})$ *is unbounded from above.*

Let $\Lambda_1 := \{j \in \Theta :$ ($\mathscr{H}(j)$) or ($\mathscr{I}(j)$) holds$\}$ and

$$\Lambda_2 := \{j \in \Theta \setminus \Lambda_1 : \exists k \in \Lambda_1 \textit{ s.t. } \operatorname{supp}(q_{k,j}F_{k,j}) \cap (0,\infty) \neq \varnothing\}.$$

Then, $\boldsymbol{J}^+$ is irreducible if $\Lambda_1 \cup \Lambda_2 = \Theta$.

(ii) Let $\widetilde{\boldsymbol{Q}}$ be a minimal irreducible version of $\boldsymbol{Q}$. If

$$\begin{aligned}&\{(i,j) \in \Theta^2 \setminus \{(i,i) : i \in \Theta\} : \widetilde{q}_{i,j} > 0\}\\ &\quad \subset \{(i,j) \in \Theta^2 \setminus \{(i,i) : i \in \Theta\} : (\mathscr{H}(j)) \textit{ holds or } \operatorname{supp}(q_{i,j}F_{i,j}) \cap (0,\infty) \neq \varnothing\},\end{aligned}$$

then $\boldsymbol{J}^+$ is irreducible.

Proof. (i) Fix $i,j \in \Theta$ with $i \neq j$. We have to show that $\mathbb{P}^{0,i}(\tau^+(j) < \infty) > 0$, where $\tau^+(j) := \inf\{t \geq 0 : J_t^+ = j\}$. Recall that $(\sigma_n)_{n\in\mathbb{N}}$ denote the jump times of the modulating chain $\boldsymbol{J}$. Let $n \in \mathbb{N}$ such that $\mathbb{P}^{0,i}(J_{\sigma_n} = j) > 0$, which exists by irreducibility of $\boldsymbol{J}$. Let $G_t := \sup\{0 \leq s < t : \xi_s = \bar{\xi}_s\}$ be the last time before $t > 0$ at which $\boldsymbol{\xi}$ attains its supremum. By construction of local time at the supremum L, the range of $(\mathsf{L}_t^{-1})_{t\geq 0}$ almost surely equals the set of times, when $\boldsymbol{\xi}$ reaches a maximum. Thus, we have

$$\begin{aligned}\mathbb{P}^{0,i}(\tau^+(j) < \infty) &\geq \mathbb{P}^{0,i}(G_{\sigma_{n+1}} \geq \sigma_n, J_{\sigma_n} = j)\\ &\geq \max\left\{\mathbb{P}^{0,i}(\xi_{\sigma_n} \geq \bar{\xi}_{\sigma_n -}, J_{\sigma_n} = j), \mathbb{P}^{0,i}(G_{\sigma_{n+1}} > \sigma_n, J_{\sigma_n} = j)\right\}.\end{aligned} \tag{4.8}$$

Suppose first that ($\mathscr{H}(j)$) holds. By the path decomposition of $(\boldsymbol{\xi}, \boldsymbol{J})$ from Proposition 4.1, we obtain

$$\begin{aligned}&\mathbb{P}^{0,i}(G_{\sigma_{n+1}} > \sigma_n, J_{\sigma_n} = j)\\ &= \mathbb{P}^{0,i}\big(\{J_{\sigma_n} = j\} \cap \{\exists t \in (0, \sigma_{n+1} - \sigma_n) : \xi_{t+\sigma_n} - \xi_{\sigma_n} \geq \bar{\xi}_{\sigma_n} - \xi_{\sigma_n}\}\big)\\ &= \mathbb{E}^{0,i}\big[\mathbb{P}^{0,j}(\exists t \in (0,\sigma_1) : \xi_t \geq x)|_{x=\bar{\xi}_{\sigma_n} - \xi_{\sigma_n}};\, J_{\sigma_n} = j\big]\\ &\geq \mathbb{E}^{0,i}\big[\mathbb{P}^{0,j}(\xi_{\sigma_1/2} \geq x)|_{x=\bar{\xi}_{\sigma_n} - \xi_{\sigma_n}};\, J_{\sigma_n} = j\big]\\ &= \mathbb{E}^{0,i}\bigg[\bigg(\int_0^\infty -2q_{j,j}\mathrm{e}^{2q_{j,j}t}\mathbb{P}(\xi_t^{(j)} \geq x)\,\mathrm{d}t\bigg)\bigg|_{x=\bar{\xi}_{\sigma_n} - \xi_{\sigma_n}};\, J_{\sigma_n} = j\bigg]\\ &> 0.\end{aligned} \tag{4.9}$$

To argue that the last inequality holds, note that since ($\mathscr{H}(j)$) was assumed, Theorem 24.7 in [147] yields that for any $t > 0$, $\operatorname{supp}(\mathbb{P}(\xi_t^{(j)} \in \cdot))$ is not bounded from above. Thus, $\mathbb{P}(\xi_t^{(j)} \geq x) > 0$ for any $x \in \mathbb{R}$ and hence

$$\bigg(\int_0^\infty -2q_{j,j}\mathrm{e}^{2q_{j,j}t}\mathbb{P}(\xi_t^{(j)} \geq x)\,\mathrm{d}t\bigg)\bigg|_{x=\bar{\xi}_{\sigma_1} - \xi_{\sigma_1}} > 0, \quad \mathbb{P}^{0,i}\text{-a.s..}$$

Combining this with $\mathbb{P}^{0,i}(J_{\sigma_n} = j) > 0$ by our choice of $n \in \mathbb{N}$, the inequality follows.

Suppose now that $(\mathcal{J}(j))$ holds, i.e., $\operatorname{supp}(F_{k,j})$ is unbounded from above for some $k \neq j$ s.t. $q_{k,j} > 0$. Let $m \in \mathbb{N}$ such that $\mathbb{P}^{0,i}(J_{\sigma_{m-1}} = k, J_{\sigma_m} = j) > 0$, which exists by irreducibility of $\boldsymbol{J}$ and $q_{k,j} > 0$. Then, again by Proposition 4.1,

$$\mathbb{P}^{0,i}(\xi_{\sigma_m} \geq \bar{\xi}_{\sigma_m-}, J_{\sigma_m} = j) = \sum_{k\neq j} \mathbb{E}^{0,i}\big[\mathbb{P}(\Delta_{k,j} \geq x)\big]\big|_{x=\bar{\xi}_{\sigma_m-}-\xi_{\sigma_m-}};\ J_{\sigma_{m-1}} = k, J_{\sigma_m} = j\big] > 0,$$

where the inequality follows from

$$\mathbb{P}(\Delta_{k,j} \geq x)|_{x=\bar{\xi}_{\sigma_m-}-\xi_{\sigma_m-}} > 0, \quad \mathbb{P}^{0,i}\text{-a.s.},$$

thanks to assumed unboundedness of the support of $F_{k,j}$. We therefore conclude with (4.8) that $\mathbb{P}^{0,i}(\tau^+(j) < \infty) > 0$ for $j \in \Lambda_1$. Suppose now that $j \in \Lambda_2$, i.e., there exists $k \in \Lambda_1$ s.t. $\operatorname{supp}(q_{k,j}F_{k,j}) \cap (0,\infty) \neq \emptyset$. Then, by Lemma 4.32, $q^+_{k,j} > 0$ and since $k \in \Lambda_1$, it follows from above that $\mathbb{P}^{0,i}(\tau^+(k) < \infty) > 0$. Combining these observations yields again $\mathbb{P}^{0,i}(\tau^+(j) < \infty) > 0$. Thus, the assumption $\Lambda_1 \cup \Lambda_2 = \Theta$ implies $\mathbb{P}^{0,i}(\tau^+(j) < \infty) > 0$ for any $j \neq i$, as desired.

(ii) Let $(i,j) \in \Theta^2$ with $i \neq j$ s.t. $\widetilde{q}_{i,j} > 0$. Suppose first that $(\mathcal{H}(j))$ holds. Then, $q^+_{i,j} > 0$ holds if we can show that $\mathbb{P}^{0,i}(G_{\sigma_2} > \sigma_1, J_{\sigma_1} = j) > 0$. This is an immediate consequence of (4.9) with $n = 1$ since $q_{i,j} = \widetilde{q}_{i,j} > 0$ implies $\mathbb{P}^{0,i}(J_{\sigma_1} = j) = -q_{i,j}/q_{i,i} > 0$. Suppose now that $\operatorname{supp}(q_{i,j}F_{i,j}) \cap (0,\infty) \neq \emptyset$. Then, again by Lemma 4.32, $q^+_{i,j} > 0$ as well. Thus, the assumption yields

$$\{(i,j) \in \Theta^2 \setminus \{(i,i) : i \in \Theta\} : \widetilde{q}_{i,j} > 0\} \subset \{(i,j) \in \Theta^2 \setminus \{(i,i) : i \in \Theta\} : q^+_{i,j} > 0\},$$

and irreducibility of $\boldsymbol{Q}^+$ follows from irreducibility of $\boldsymbol{Q}$.

■

Assume for the rest of this section that $(\mathcal{A}2)$ is satisfied and denote by $\boldsymbol{\pi}^+ = (\pi^+(1), \ldots, \pi^+(n))$ the invariant distribution of $\boldsymbol{J}^+$. Moreover, for a measure η on $\mathbb{R}_+$, denote by $\bar{\eta}(x) = \eta((x,\infty))$, $x \geq 0$, the tail measure of η. Our main goal is to understand the asymptotic behavior of overshoots. As a natural extension of the well-known limiting distributional behavior of overshoots of Lévy processes, cf. [24], it is shown in Theorem 28 of [66] that under assumptions $(\mathcal{A}1)$ and $(\mathcal{A}2)$ the overshoot process converges weakly to the limiting distribution

$$\mu(\mathrm{d}y, \{i\}) := \frac{1}{\mathbb{E}^{0,\boldsymbol{\pi}^+}[H_1^+]}\Big(\pi^+(i)d_i^+\delta_0(\mathrm{d}y) + \mathbb{1}_{(0,\infty)}(y)\Big(\pi^+(i)\overline{\Pi}_i^+(y) + \sum_{j\neq i}\pi^+(j)q^+_{j,i}\bar{F}^+_{j,i}(y)\Big)\,\mathrm{d}y\Big),$$

$(y,i) \in \mathbb{R}_+ \times \Theta$, if and only if $\mathbb{E}^{0,\boldsymbol{\pi}^+}[H_1^+] < \infty$.[2] The Feller property of the overshoot process guarantees that in this case μ is also an invariant measure. We will show that deleting the scaling factor $\mathbb{E}^{0,\boldsymbol{\pi}^+}[H_1^+]^{-1}$ yields the essentially unique invariant measure of the overshoot process and

[2]Here we made a correction to [66], since in the authors' statement the limiting distribution of the parents modulator $\boldsymbol{J}$, $\boldsymbol{\pi}$, appears instead of $\boldsymbol{\pi}^+$. As argued before, irreducibility of $\boldsymbol{J}$ does not necessarily imply irreducibility of $\boldsymbol{J}^+$ and even when $\boldsymbol{J}^+$ is irreducible, $\boldsymbol{\pi}$ and $\boldsymbol{\pi}^+$ are not the same, see [110, Proposition 2.19]. Our analysis will show however that stationarity of the ascending ladder height's modulator and its stationary distribution are decisive for tight overshoots.

hence a stationary distribution coinciding with μ exists iff overshoots are tight. Moreover, we will dig deeper into the mode of convergence, establishing conditions ensuring convergence in the total variation norm and exponential or polynomial speed of convergence, which also gives new results for the special case of Lévy process overshoots.

An analytical tool of central importance to us is the resolvent of the overshoot process. Let $(\mathcal{P}_t)_{t\geq 0}$ be the transition function of $(\mathfrak{O},\mathfrak{J})$ defined by $\mathcal{P}_t f(x,i) = \mathbb{E}^{x,i}[f(\mathcal{O}_t,\mathcal{J}_t)]$ for any $f \in \mathcal{B}_b(\mathbb{R}_+\times\Theta)\cup\mathcal{B}_+(\mathbb{R}_+\times\Theta)$ and $(\mathcal{U}_\lambda)_{\lambda>0}$ be the associated resolvent given by

$$\mathcal{U}_\lambda f(x,i) = \int_0^\infty \mathrm{e}^{-\lambda t}\mathcal{P}_t f(x,i)\,\mathrm{d}t,$$

for any $\lambda > 0$. Our proof for the explicit formula of the resolvent is close in spirit to the proof for the overshoot process of a Lévy subordinator in Blumenthal [32], which in turn is a special case of a general result by Itō for Markov processes possessing a local time at a specific point of the state space, see [94, Theorem 2.5.5]. The detailed proof is quite long and can be found in Appendix 4.A.

Theorem 4.7. *For any $f \in \mathcal{B}_+(\mathbb{R}_+\times\Theta)\cup\mathcal{B}_b(\mathbb{R}_+\times\Theta)$ and $x\in\mathbb{R}_+$ it holds that*

$$(\mathcal{U}_\lambda f(x,i))^\top_{i=1,\dots,n} = (Q_\lambda f(x,i))^\top_{i=1,\dots n} + \mathrm{e}^{-\lambda x}\boldsymbol{\Phi}^+(\lambda)^{-1}\cdot\boldsymbol{\psi}(f,\lambda), \tag{4.10}$$

where

$$\boldsymbol{\psi}(f,\lambda) = \left(d_i^+ f(0,i) + \int_0^\infty Q_\lambda f(x,i)\,\Pi_i^+(\mathrm{d}x) + \sum_{j\neq i} q^+_{i,j}\mathbb{E}[Q_\lambda f(\Delta^+_{i,j},j)]\right)^\top_{i=1,\dots,n}$$

and

$$Q_\lambda f(x,i) = \int_0^x \mathrm{e}^{-\lambda t} f(x-t,i)\,\mathrm{d}t, \quad (x,i)\in\mathbb{R}_+\times\Theta.$$

The resolvent formula has far reaching consequences for understanding the behavior of the MAP at first passage. A first neat observation is the strong Feller property of the resolvent operator, which implies that $(\mathfrak{O},\mathfrak{J})$ is a T-process.

Corollary 4.8. *For any $\lambda > 0$ the resolvent $\mathcal{U}_\lambda$ has the strong Feller property. In particular the overshoot process $(\mathfrak{O},\mathfrak{J})$ is a T-process.*

Proof. Let $\lambda > 0$ and let $f\in\mathcal{B}_b(\mathbb{R}_+\times\Theta)$. Since we can write $Q_\lambda f(x,i) = \mathrm{e}^{-\lambda x}\int_0^x \mathrm{e}^{\lambda t} f(t,i)\,\mathrm{d}t$, it follows that $(x,i)\mapsto Q_\lambda f(x,i)$ is continuous and hence $(x,i)\mapsto\mathcal{U}_\lambda f(x,i)$ is clearly continuous. Moreover, $\mathcal{U}_\lambda f$ is bounded and thus, $\mathcal{U}_\lambda\mathcal{B}_b(\mathbb{R}_+\times\Theta)\subset\mathcal{C}_b(\mathbb{R}_+\times\Theta)$ follows, i.e. $\mathcal{U}_\lambda$ has the strong Feller property. Hence, the resolvent kernel $\mathcal{R}_\lambda := \lambda\mathcal{U}_\lambda$ is a continuous component for itself, implying that $(\mathfrak{O},\mathfrak{J})$ is a T-process. ∎

We will also use the resolvent formula combined with Proposition 2.8 to determine an invariant measure for the overshoot process. To show its essential uniqueness, we need to establish Harris recurrence first, which is taken care of in the following proposition.

Proposition 4.9. *The overshoot process $(\mathfrak{O},\mathfrak{J})$ is Harris recurrent.*

Proof. Let $j \in \Theta$ be arbitrarily chosen and let $\nu := \delta_0 \otimes \delta_j$. Fix $(x, i) \in \mathbb{R}_+ \times \Theta$ and let $B \in \mathcal{B}(\mathbb{R}_+ \times \Theta)$ such that $\nu(B) > 0$, i.e. $\{0\} \times \{j\} \in B$. Since $\boldsymbol{J}^+$ is irreducible and $t \mapsto T_t^+$ is continuous and increases to ∞ as $t \to \infty$, it follows that $\mathbb{P}^{x,i}(\mathrm{t}^+(j) < \infty) > 0$, where $\mathrm{t}^+(j) := \inf\{t > 0 : \mathcal{J}_t^+ = j\}$ is the first hitting time of $\{j\}$ of $\mathfrak{J}^+$. Let $T_\Lambda = \inf\{t \geq 0 : (\mathcal{O}_t, \mathcal{J}_t) \in \Lambda\}$ be the first hitting time of a set $\Lambda \in \mathbb{R}_+ \times \Theta$ by $(\mathfrak{O}, \mathfrak{J})$ and denote by T_Λ^+ the first hitting time of $(\mathfrak{O}^+, \mathfrak{J}^+)$. By the sawtooth structure of $\mathfrak{O}^+$ we have $T_{\{0\}\times\{j\}}^+ = x$, $\mathbb{P}^{x,j}$-a.s.. Since $\mathrm{t}^+(j) \leq T_{\{0\}\times\{j\}}^+$ it therefore follows by the strong Markov property of $(\mathfrak{O}^+, \mathfrak{J}^+)$ that

$$\begin{aligned}
\mathbb{P}^{x,i}(T_B < \infty) \geq \mathbb{P}^{x,i}\big(T_{\{0\}\times\{j\}} < \infty\big) &= \mathbb{P}^{x,i}\big(T_{\{0\}\times\{j\}}^+ < \infty\big) \\
&= \mathbb{E}^{x,i}\Big[\mathbb{P}^{\mathcal{O}_{\mathrm{t}^+(j)}^+, \mathcal{J}_{\mathrm{t}^+(j)}^+}\big(T_{\{0\}\times\{j\}}^+ < \infty\big) \mathbb{1}_{\{\mathrm{t}^+(j) < \infty\}}\Big] \\
&= \mathbb{P}^{x,i}(\mathrm{t}^+(j) < \infty) > 0,
\end{aligned}$$

where we used for the last equality that $\mathcal{J}_{\mathrm{t}^+(j)}^+ = j$ and $\mathcal{O}_{\mathrm{t}^+(j)}^+ < \infty$ almost surely. It now follows from Proposition 2.1 in [130] that $(\mathfrak{O}, \mathfrak{J})$ is irreducible with irreducibility measure

$$\mathcal{R}_1^\nu(\mathrm{d}y) := \int_{\mathbb{R}_+ \times \Theta} \mathcal{R}_1(x, \mathrm{d}y)\, \nu(\mathrm{d}x) = \mathcal{R}_1((0, j), \mathrm{d}y), \quad y \in \mathbb{R}_+ \times \Theta.$$

Moreover, $(\mathfrak{O}, \mathfrak{J})$ is a T-process by Corollary 4.8. Hence, if we can argue that the process is non-evanescent, i.e. that there exists a compact set K such that $(\mathfrak{O}, \mathfrak{J})$ returns to K at arbitrarily large times, it will follow from Theorem 3.2 in [130] that $(\mathfrak{O}, \mathfrak{J})$ is Harris recurrent. But non-evanescence is a direct consequence of the sawtooth structure of the overshoot process, since for the compact set $K := \{0\} \times \Theta$ we have for any $(x, i) \in \mathbb{R}_+ \times \Theta$ and $t > 0$

$$\mathbb{P}^{x,i}(\inf\{s \geq t : (\mathcal{O}_s, \mathcal{J}_s) \in \{0\} \times \Theta\} < \infty) = \mathbb{E}^{x,i}[\mathbb{P}^{\mathcal{O}_t, \mathcal{J}_t}(T_{\{0\}\times\Theta} < \infty)] = 1,$$

where we used that $T_{\{0\}\times\Theta} = y$, $\mathbb{P}^{y,j}$-a.s. for any $(y, j) \in \mathbb{R}_+ \times \Theta$ and $\mathcal{O}_t < \infty$ almost surely. Hence, $(\mathfrak{O}, \mathfrak{J})$ is non-evanescent and the assertion follows. ■

As a consequence of irreducibility implied by Harris recurrence and $(\mathfrak{O}, \mathfrak{J})$ being a T-process, we obtain that every compact set is petite, which will be useful for our proof of exponential convergence of the overshoot process later on.

COROLLARY 4.10. *Every compact set is petite for the overshoot process.*

Proof. This is an immediate consequence of Theorem 5.1 in [164] since $(\mathfrak{O}, \mathfrak{J})$ is a Harris recurrent T-process under the given assumptions and Harris recurrence implies irreducibility. ■

Let us now determine the essential unique invariant measure of $(\mathfrak{O}, \mathfrak{J})$ and also derive a necessary and sufficient condition for the existence of a unique stationary distribution, which is the same condition needed for weak convergence of overshoots.

THEOREM 4.11. *The overshoot process* $(\mathfrak{O}, \mathfrak{J})$ *has an essentially unique invariant measure given by*

$$\chi(\mathrm{d}y, \{i\}) = \pi^+(i) d_i^+ \delta_0(\mathrm{d}y) + \mathbb{1}_{(0,\infty)}(y)\Big(\pi^+(i)\overline{\Pi}_i^+(y) + \sum_{j \neq i} \pi^+(j) q_{j,i}^+ \overline{F}_{j,i}^+(y)\Big)\,\mathrm{d}y, \quad (y, i) \in \mathbb{R}_+ \times \Theta. \tag{4.11}$$

In particular, a stationary distribution for $(\mathfrak{O}, \mathfrak{J})$ *exists if and only if*

$$\mathbb{E}^{0,\boldsymbol{\pi}^+}[H_1^+] := \sum_{i=1}^n \pi^+(i)\mathbb{E}^{0,i}[H_1^+] < \infty.$$

Proof. Define $\boldsymbol{\alpha}(\lambda) := \boldsymbol{\pi}^+ \cdot \boldsymbol{\Phi}^+(\lambda)$ and

$$\alpha_\lambda := \sum_{i=1}^n \pi^+(i)\Phi_i^+(\lambda)\,\delta_{\{0\}\times\{i\}},$$

then $\lim_{\lambda\downarrow 0} \operatorname{diag}(\Phi_1^+(\lambda), \dots, \Phi_n^+(\lambda)) = \mathbb{0}_{n\times n}$ implies that $\alpha_\lambda(\mathbb{R}_+ \times \Theta) \to 0$ as $\lambda \downarrow 0$. Moreover, since $\boldsymbol{\pi}^+$ is the stationary distribution of $\boldsymbol{J}^+$ and $\boldsymbol{G}^+(0) = \mathbb{I}_n$, we have

$$\lim_{\lambda\downarrow 0} \boldsymbol{\pi}^+ \cdot (\boldsymbol{Q}^+ \odot \boldsymbol{G}^+(\lambda)) = \boldsymbol{\pi}^+ \cdot \boldsymbol{Q}^+ = \mathbb{0}_{1\times n}.$$

Recall from Chapter 2 the notation $\alpha_\lambda \mathcal{U}_\lambda(\mathrm{d}x) := \int_{\mathbb{R}_+\times\Theta} \mathcal{U}_\lambda(y, \mathrm{d}x)\,\alpha_\lambda(\mathrm{d}y)$. Plugging into the resolvent formula from Theorem 4.7 yields for any $f \in \mathcal{B}_b(\mathbb{R}_+ \times \Theta) \cap \mathcal{B}_+(\mathbb{R}_+ \times \Theta)$ that

$$\begin{aligned}
\lim_{\lambda\downarrow 0} \alpha_\lambda \mathcal{U}_\lambda(f) &= \lim_{\lambda\downarrow 0} \sum_{i=1}^n \pi_i^+ \Phi_i^+(\lambda)\mathcal{U}_\lambda f(0, i) \\
&= \lim_{\lambda\downarrow 0} \boldsymbol{\alpha}(\lambda) \cdot (\mathcal{U}_\lambda f(0, i))_{i=1,\dots,n}^\top \\
&= \lim_{\lambda\downarrow 0} \boldsymbol{\pi}^+ \cdot \left(d_i^+ f(0, i) + \int_0^\infty Q_\lambda f(y, i)\,\Pi_i^+(\mathrm{d}y) + \sum_{j\neq i} q_{i,j}^+ Q_\lambda f(y, j)\,F_{i,j}^+(\mathrm{d}y) \right)_{i=1,\dots n}.
\end{aligned} \tag{4.12}$$

By monotone convergence and an integration by parts it follows that for any measure ν on $\mathbb{R}_+$

$$\begin{aligned}
\lim_{\lambda\downarrow 0} \int_0^\infty Q_\lambda f(y, i)\,\nu(\mathrm{d}y) &= \int_0^\infty \int_0^y f(y - t, i)\,\mathrm{d}t\,\nu(\mathrm{d}y) \\
&= \int_0^\infty \int_0^y f(t, i)\,\mathrm{d}t\,\nu(\mathrm{d}y) \\
&= \int_0^\infty \bar{\nu}(y) f(y, i)\,\mathrm{d}y,
\end{aligned}$$

where $\bar{\nu}(y) := \nu(y, \infty)$. Thus, we obtain from (4.12) that

$$\begin{aligned}
\lim_{\lambda\downarrow 0} \alpha_\lambda \mathcal{U}_\lambda(f) &= \boldsymbol{\pi}^+ \cdot \left(d_i^+ f(0, i) + \int_0^\infty f(y, i)\overline{\Pi}_i^+(y)\,\mathrm{d}y + \sum_{j\neq i} q_{i,j}^+ f(y, j)\bar{F}_{i,j}^+(y)\,\mathrm{d}y \right)_{i=1,\dots,n}^\top \\
&= \sum_{i=1}^n \pi^+(i)\left(d_i^+ f(0, i) + \int_0^\infty f(y, i)\overline{\Pi}_i^+(y)\,\mathrm{d}y + \sum_{j\neq i} q_{i,j}^+ \int_0^\infty f(y, i)\bar{F}_{i,j}^+(y)\,\mathrm{d}y \right) \\
&= \sum_{i=1}^n \left(\pi^+(i)\left(d_i^+ f(0, i) + \int_0^\infty f(y, i)\overline{\Pi}_i^+(y)\,\mathrm{d}y \right) + \sum_{j\neq i} \pi^+(j) q_{j,i}^+ \int_0^\infty f(y, i)\bar{F}_{j,i}^+(y)\,\mathrm{d}y \right)
\end{aligned}$$

$$= \int_{\mathbb{R}_+\times\Theta} f(y,z)\,\chi(\mathrm{d}y\times\mathrm{d}z),$$

where for the second to last equality we used that

$$\begin{aligned}\sum_{i=1}^{n}\pi^+(i)\sum_{j\neq i}q^+_{i,j}\int_0^\infty f(y,j)\bar F^+_{i,j}\,\mathrm{d}y &= \sum_{j=1}^{n}\sum_{i\neq j}q^+_{i,j}\pi^+(i)\int_0^\infty f(y,j)\bar F^+_{i,j}(y)\,\mathrm{d}y\\ &= \sum_{i=1}^{n}\sum_{j\neq i}q^+_{j,i}\pi^+(j)\int_0^\infty f(y,i)\bar F^+_{j,i}(y)\,\mathrm{d}y.\end{aligned}$$

From Proposition 2.8 it now follows that χ is indeed an invariant measure for $(\mathfrak{O},\mathfrak{J})$. By irreducibility of $\boldsymbol{J}^+$, $(\mathfrak{O},\mathfrak{J})$ is a Harris recurrent Feller process according to Propositions 4.4 and 4.9 and hence Theorem 2.5 in [17] yields that χ is essentially unique.

Finally, using the Laplace exponent of $(\boldsymbol{H}^+,\boldsymbol{J}^+)$ we obtain

$$\begin{aligned}\left(\mathbb{E}^{0,i}[H_1^+\mathbb{1}_{\{J_1^+=j\}}]\right)_{i,j=1,\dots,n} &= \frac{\partial}{\partial\lambda}\boldsymbol{\Phi}^+(\lambda)\big|_{\lambda=0}\\ &= \operatorname{diag}\left(\left(\mathbb{E}[H_1^{+,(i)}]\right)\right)_{i\in\Theta} + \boldsymbol{Q}^+\odot\left(\mathbb{E}[\Delta^+_{i,j}]\right)_{i,j=1,\dots,n}\\ &= \operatorname{diag}\left(\left(d_i^+ + \int_0^\infty \overline{\Pi}_i^+(x)\,\mathrm{d}x\right)\right)_{i\in\Theta} + \boldsymbol{Q}^+\odot\left(\int_0^\infty \bar F^+_{i,j}(x)\,\mathrm{d}x\right)_{i,j=1,\dots,n},\end{aligned}$$

and hence

$$\mathbb{E}^{0,i}\left[H_1^+\right] = d_i^+ + \int_0^\infty \overline{\Pi}_i^+(x)\,\mathrm{d}x + \sum_{j\neq i}q^+_{i,j}\int_0^\infty \bar F^+_{i,j}(x)\,\mathrm{d}x,\quad i\in\Theta,$$

which shows that

$$\chi(\mathbb{R}_+\times\Theta) = \mathbb{E}^{0,\boldsymbol{\pi}^+}\left[H_1^+\right].$$

Thus, χ can be normalized to an invariant distribution if and only if $\mathbb{E}^{0,\boldsymbol{\pi}^+}[H_1^+] < \infty$. ■

Remark 4.12. The finite mean condition for the ascending ladder height process is exactly the same condition, which is necessary and sufficient for stationary overshoots of MAPs in the sense of weak convergence. As shown in Theorem 35 of [66] as an extension of Theorem 8 in [70] for Lévy processes, this condition is equivalent to $\mathbb{E}^{0,i}[|\xi_1|] < \infty$ and either $\lim_{t\to\infty}\xi_t = \infty$, $\mathbb{P}^{0,i}$-a.s., or $\limsup_{t\to\infty}\xi_t = -\liminf_{t\to\infty}\xi_t = \infty$, $\mathbb{P}^{0,i}$-a.s., together with

$$\int_\kappa^\infty \frac{x\sum_{i=1}^n \boldsymbol{\Pi}(i,[x,\infty)\times\Theta)}{1+\int_0^x\int_y^\infty\sum_{i=1}^n\boldsymbol{\Pi}(i,(-\infty,-z]\times\Theta)\,\mathrm{d}z\,\mathrm{d}y}\,\mathrm{d}x < \infty, \tag{4.13}$$

for some $\kappa > 0$.

Our discussion of the interplay between Harris recurrence and invariant measures for Markov process in Appendix 2 now also yields that χ is a maximal Harris meaure.

COROLLARY 4.13. *The invariant measure χ given in* (4.11) *is a maximal Harris measure.*

Remark 4.14. This could have also been shown directly by an alternative proof of Proposition 4.9 based on Kaspi and Mandelbaum's characterization of Harris recurrence in terms of almost sure finiteness of first hitting times (2.3) and the characteristic property (4.2) of the Lévy system belonging to $(\boldsymbol{H}^+,\boldsymbol{J}^+)$.

Having established the existence of a unique invariant distribution, we now proceed to investigate ergodicity of overshoots. To this end, we need to find criteria ensuring the existence of an irreducible skeleton chain. One of these criteria will be a strictly positive creeping probability of the MAP and we lift a sufficient criterion for this to happen from the well-known Lévy process situation.

LEMMA 4.15. *Suppose that $d_i^+ > 0$ for some $i \in \Theta$. Then, for any $t > 0$ we have*

$$\mathbb{P}^{0,i}\big(\xi_{T_t} = t, J_{T_t} = i\big) > 0.$$

Proof. Let σ_1^+ be the first jump time of $\boldsymbol{J}^+$. If $q_{i,i}^+ = 0$, then under $\mathbb{P}^{0,i}$, $\boldsymbol{H}^+$ is a Lévy subordinator with positive drift and therefore has positive creeping probability by Theorem 5.9 in [109], implying the claim. Suppose now $-q_{i,i}^+ > 0$. Then, using the representation from Proposition 4.1 we have

$$\begin{aligned}
\mathbb{P}^{0,i}(\mathcal{O}_t = 0, \mathcal{J}_t = i) &= \mathbb{P}^{0,i}(\mathcal{O}_t^+ = 0, \mathcal{J}_t^+ = i) \\
&\geq \mathbb{P}^{0,i}\Big(H^{+,0,i}_{T_t^{+,0,i}} = t, T_t^{+,0,i} < \sigma_1^+\Big) \\
&= \int_0^\infty \mathbb{P}\Big(H^{+,(i)}_{T_t^{+,(i)}} = t, T_t^{+,(i)} < y\Big)\, \mathbb{P}^{0,i}(\sigma_1^+ \in \mathrm{d}y) \\
&= -q_{i,i}^+ \int_0^\infty \mathrm{e}^{q_{i,i}^+ y}\, \mathbb{P}\Big(H^{+,(i)}_{T_t^{+,(i)}} = t, T_t^{+,(i)} < y\Big)\, \mathrm{d}y,
\end{aligned}$$

where we used independence of $H^{+,0,i}$ and $\boldsymbol{J}^+$ for the third equality. Since again by Theorem 5.9 in [109], $d_i^+ > 0$ gives that $\mathbb{P}(H^{+,(i)}_{T_t^{+,(i)}} = t) > 0$ for all $t \geq 0$ and

$$\lim_{y\to\infty} \mathbb{P}\Big(H^{+,(i)}_{T_t^{+,(i)}} = t, T_t^{+,(i)} < y\Big) = \mathbb{P}\Big(H^{+,(i)}_{T_t^{+,(i)}} = t\Big),$$

it follows that there is $z > 0$ such that $\mathbb{P}\big(H^{+,(i)}_{T_t^{+,(i)}} = t, T_t^{+,(i)} < y\big) > 0$ for all $y \geq z$ and hence, from above it follows that

$$\mathbb{P}^{0,i}(\mathcal{O}_t = 0, \mathcal{J}_t = i) \geq -q_{i,i}^+ \int_z^\infty \mathrm{e}^{q_{i,i}^+ y}\, \mathbb{P}\Big(H^{+,(i)}_{T_t^{+,(i)}} = t, T_t^{+,(i)} < y\Big)\, \mathrm{d}y > 0.$$

∎

Remark 4.16. The irreducibility assumption ($\mathcal{A}$2) is not required for this statement.

Let us now state properties of the ascending ladder height process that imply existence of an irreducible skeleton of $(\mathfrak{O}, \mathfrak{J})$.

PROPOSITION 4.17. *If*

(i) $d_i^+ > 0$ for some $i \in \Theta$, then $(\mathfrak{O}, \mathfrak{J})$ is aperiodic and any Δ-skeleton is irreducible.

(ii) for some $j \in \Theta$ it holds $\lambda|_{(0,\infty)} \ll \Pi_j^+|_{(0,\infty)}$, then any Δ-skeleton $(\mathfrak{O}^\Delta, \mathfrak{J}^\Delta)$ is $\lambda_+ \otimes \delta_j$-irreducible.

(iii) *for some $j \in \Theta$ there exists an interval $(a,b) \subset \mathbb{R}_+$ such that $\lambda|_{(a,b)} \ll \Pi_j^+|_{(a,b)}$ and for any $i \in \Theta$ and $x > 0$ it holds that $U_{i,j}^+([0,x)) > 0$, then for any $\Delta \in (0,(a+b)/2)$, the Δ-skeleton $(\mathcal{O}^\Delta, \mathcal{J}^\Delta)$ is $\lambda_+(\cdot \cap (a,(a+b)/2)) \otimes \delta_j$-irreducible.*

(iv) *for some $(j,k) \in \Theta^2$ with $k \neq j$ it holds $\lambda|_{(0,\infty)} \ll F_{k,j}^+|_{(0,\infty)}$ and $q_{k,j}^+ > 0$, then any Δ-skeleton $(\mathcal{O}^\Delta, \mathcal{J}^\Delta)$ is $\lambda_+ \otimes \delta_j$-irreducible.*

(v) *for some $(j,k) \in \Theta^2$ with $k \neq j$ it holds $q_{k,j}^+ > 0$, there exists an interval $(a,b) \subset \mathbb{R}_+$ such that $\lambda|_{(a,b)} \ll F_{k,j}^+|_{(a,b)}$ and for any $i \in \Theta$ and $x > 0$ it holds that $U_{i,k}^+([0,x)) > 0$, then for any $\Delta \in (0,(a+b)/2)$, the Δ-skeleton $(\mathcal{O}^\Delta, \mathcal{J}^\Delta)$ is $\lambda_+(\cdot \cap (a,(a+b)/2)) \otimes \delta_j$-irreducible.*

Proof.

(i) The singleton set $C = \{0\} \times \{i\}$ is trivially small (just choose $\nu_a = P_t((0,i),\cdot)$ for $a = \delta_t$ and some $t > 0$.). Further, $C \in \mathcal{B}^+(\mathbb{R}_+ \times \Theta)$ since Corollary 4.13 tells us that the invariant measure χ is an irreducibility measure for $(\mathcal{O}, \mathcal{J})$ and thanks to $d_i^+ > 0$, we have $\chi(C) > 0$. Lemma 4.15 gives that

$$\mathbb{P}^{0,i}((\mathcal{O}_t, \mathcal{J}_t) \in C) = \mathbb{P}^{0,i}(\mathcal{O}_t = 0, \mathcal{J}_t = i) > 0$$

for all $t \geq 0$, which implies that $(\mathcal{O}, \mathcal{J})$ is aperiodic with defining singleton set $C = \{0\} \times \{i\}$, which by Lemma 2.9 also implies that any Δ-skeleton is irreducible.

(ii) Let $B = B_1 \times B_2 \in \mathcal{B}(\mathbb{R}_+ \times \Theta)$ such that $\lambda_+ \otimes \delta_j(B) > 0$. Without loss of generality we may assume that $0 \notin B_1$. Since J^+ is irreducible it holds $\mathbb{P}^{0,i}(J_t^+ = j) > 0$ for any $t > 0$ and $i \in \Theta$ and hence by monotone convergence,

$$\lim_{x \to \infty} U_{i,j}^+([0,x)) = \int_0^\infty \mathbb{P}^{0,i}(J_t^+ = j)\,\mathrm{d}t > 0,$$

which yields that there exists $\bar{x} > 0$ such that $U_{i,j}^+([0,x)) > 0$ for all $x \geq \bar{x}$ and $i \in \Theta$. For given $x \geq 0$ let $t > x + \bar{x}$. Then, by the overshoot formula and Fubini it follows that for any $i \in \Theta$ we have

$$\begin{aligned}\mathbb{P}^{x,i}(\mathcal{O}_t \in B_1, \mathcal{J}_t \in B_2) &\geq \int_{[0,t-x)} \int_{B_1} \Pi_j^+(y + \mathrm{d}u)\, U_{i,j}^+(t - x - \mathrm{d}y) \\ &= \int_{[0,t-x)} \Pi_j^+(B_1 + t - x - y)\, U_{i,j}^+(\mathrm{d}y).\end{aligned} \tag{4.14}$$

Since by translation invariance of the Lebesgue measure it holds $\lambda(B_1 + z) > 0$ for any $z \geq 0$ and $\lambda|_{(0,\infty)} \ll \Pi_j^+|_{(0,\infty)}$ by assumption, it follows that for any $y \in [0, t-x)$ we have $\Pi_j^+(B_1 + t - x - y) > 0$. By our choice of t it also holds that $U_{i,j}^+([0,t-x)) > 0$, thus (4.14) yields that $\mathbb{P}^{x,i}(\mathcal{O}_t \in B_1, \mathcal{J}_t \in B_2) > 0$. Hence, given $\Delta > 0$, choosing $n_x \in \mathbb{N}$ large enough such that $n_x\Delta > x + \bar{x}$, it follows that $\mathbb{P}^{x,i}((\mathcal{O}_{n_x\Delta}, \mathcal{J}_{n_x\Delta}) \in B) > 0$ for any $i \in \Theta$, which shows that any Δ-skeleton is $\lambda_+ \otimes \delta_j$-irreducible.

(iii) Choose $B = B_1 \times B_2 \in \mathcal{B}(\mathbb{R}_+ \times \Theta)$ such that $\lambda(\cdot \cap (a,(a+b)/2)) \otimes \delta_j(B) > 0$. Again we may assume that $0 \notin B_1$. Let $(x,i) \in \mathbb{R}_+ \times \Theta$ and $t \in (x, x + (b-a)/2)$. Since for any $z \geq 0$ it holds that

$$(B_1 + z) \cap (a,b) = (B_1 \cap (a - z, b - z)) + z$$

it follows for $z \in (0, (b-a)/2)$ by translation invariance of the Lebesgue measure that

$$\lambda((B_1 + z) \cap (a, b)) = \lambda(B_1 \cap (a - z, b - z)) \geq \lambda(B_1 \cap (a, (a + b)/2)) > 0.$$

By our choice of $t \in (x, x+(b-a)/2)$ it holds that $0 < t-x-y < (b-a)/2$ for all $y \in (0, t-x)$ and therefore $\lambda((B_1 + t - x - y) \cap (a, b)) > 0$, which by our assumption $\lambda|_{(a,b)} \ll \Pi_j^+|_{(a,b)}$ implies that $\Pi_j^+(B_1 + t - x - y) > 0$. Since $U_{i,j}^+([0, t-x)) > 0$ by assumption it now follows from (4.14) that $\mathbb{P}^{x,i}((\mathcal{O}_t, \mathcal{J}_t) \in B) > 0$. Hence, given $\Delta \in (0, (b-a)/2)$, if we choose $k \in \mathbb{N}$ such that $k\Delta \in (x, x + (b-a)/2)$ it follows that $\mathbb{P}^{x,i}((\mathcal{O}_{k\Delta}, \mathcal{J}_{k\Delta}) \in B) > 0$ and therefore $\sum_{k=1}^{\infty} \mathbb{P}^{x,i}((\mathcal{O}_{k\Delta}, \mathcal{J}_{k\Delta}) \in B) > 0$. Since $(x, i) \in \mathbb{R}_+ \times \Theta$ was chosen arbitrarily we conclude that the Δ-skeleton is irreducible with irreducibility measure $\lambda_+(\cdot \cap (a, (a+b)/2)) \otimes \delta_j$.

Parts (iv) and (v) can be demonstrated exactly as parts (ii) and (iii) when instead of (4.14) we use that for $B = B_1 \times B_2 \in \mathcal{B}(\mathbb{R}_+ \times \Theta)$ with $j \in B_2$, $(x, i) \in \mathbb{R}_+ \times \Theta$ and $t > x$ it holds

$$\mathbb{P}^{x,i}(\mathcal{O}_t \in B_1, \mathcal{J}_t \in B_2) \geq q_{k,j}^+ \int_{[0,t-x)} F_{k,j}^+(B_1 + t - x - y)\, U_{i,k}^+(\mathrm{d}y).$$

∎

Remark 4.18. The condition in part (iii) and (v) that $U_{i,j}^+([0, x)) > 0$ for all $i \neq j$ is non-redundant in general. If, e.g., $F_{i,j}^+([0, x)) = 0$ for some $i \neq j$, then $U_{i,j}^+([0, x)) = 0$.

These conditions in combination with Harris recurrence now allow us to determine when $(\mathfrak{O}, \mathfrak{J})$ is ergodic.

Theorem 4.19. *Suppose that $\mathbb{E}^{0,\pi^+}[H_1^+] < \infty$. Then, under any of the conditions of Proposition 4.17, it holds that $(\mathfrak{O}, \mathfrak{J})$ is ergodic, i.e.*

$$\forall (x, i) \in \mathbb{R}_+ \times \Theta : \quad \lim_{t\to\infty} \|\mathbb{P}^{x,i}((\mathcal{O}_t, \mathcal{J}_t) \in \cdot) - \mu\|_{\mathrm{TV}} = 0,$$

where for $(x, i) \in \mathbb{R}_+ \times \Theta$,

$$\mu(\mathrm{d}y, \{i\}) := \frac{1}{\mathbb{E}^{0,\pi^+}[H_1^+]} \Big(\pi^+(i) d_i^+ \delta_0(\mathrm{d}y) + \mathbb{1}_{(0,\infty)}(y) \Big(\pi^+(i) \overline{\Pi}_i^+(y) + \sum_{j \neq i} \pi^+(j) q_{j,i}^+ \overline{F}_{j,i}^+(y) \Big)\, \mathrm{d}y \Big), \tag{4.15}$$

is the stationary distribution of $(\mathfrak{O}, \mathfrak{J})$.

Proof. As a consequence of Proposition 4.4, Proposition 4.9 and Theorem 4.11, it follows that under any of the conditions of Proposition 4.17, $(\mathfrak{O}, \mathfrak{J})$ is a positive Harris recurrent Borel right Markov process with unique stationary distribution given in (4.15) such that some Δ-skeleton is irreducible. Thus, Theorem 6.1 in [130] yields the assertion. ∎

A direct implication of ergodicity is that a continuous time version of the von Neumann–Birkhoff ergodic theorem holds, see the discussion in [146].

Corollary 4.20. *Given the assumptions from Theorem 4.19, it holds for any $f \in L^p(\mathbb{R}_+ \times \Theta, \mu)$ and $(x, i) \in \mathbb{R}_+ \times \Theta$ that*

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T f(\mathcal{O}_t, \mathcal{J}_t)\, \mathrm{d}t = \mu(f), \quad \mathbb{P}^{x,i}\text{-a.s. and in } L^p(\mathbb{P}^\mu).$$

Once we have derived an analogue of Vigon's équations amicales inversés in Section 4.4, we will be able to express conditions on the Lévy system $\boldsymbol{\Pi}$ of $(\boldsymbol{\xi}, \boldsymbol{J})$ that guarantee one of the conditions on the Lévy system $\boldsymbol{\Pi}^+$ of $(\boldsymbol{H}^+, \boldsymbol{J}^+)$ required for ergodicity. For the moment we content ourselves with studying the drifts d_i^+ of the subordinators associated to the ascending ladder height process.

LEMMA 4.21. *If $\boldsymbol{J}$ is irreducible, then for any $i \in \Theta$ and an appropriate scaling of local time, the diffusion parameter b_i of $\boldsymbol{\xi}^{(i)}$ is given by*

$$b_i^2 = 2d_i^+ d_i^-.$$

Proof. Let $i \in \Theta$. Considering the diagonal of $\boldsymbol{\Psi}$, the spatial Wiener–Hopf factorization (4.5) yields for every $\theta \in \mathbb{R}$

$$\begin{aligned} &\mathrm{i}a_i\theta - \frac{b_i^2}{2}\theta^2 + \int_{\mathbb{R}} \left(\mathrm{e}^{\mathrm{i}\theta x} - 1 - \mathrm{i}\theta x \mathbb{1}_{[-1,1]}(x)\right) \Pi_i(\mathrm{d}x) + q_{i,i} \\ &\quad = \left(q_{i,i}^- - \dagger_i^- - \mathrm{i}d_i^-\theta + \int_0^\infty \left(\mathrm{e}^{-\mathrm{i}\theta x} - 1\right) \Pi_i^-(\mathrm{d}x)\right) \cdot \left(q_{i,i}^+ - \dagger_i^+ + \mathrm{i}d_i^+\theta + \int_0^\infty \left(\mathrm{e}^{\mathrm{i}\theta x} - 1\right) \Pi_i^+(\mathrm{d}x)\right) \\ &\qquad + \sum_{k \neq i} \frac{\pi(k)}{\pi(i)} q_{k,i}^- q_{k,i}^+ G_{k,i}^-(-\theta) G_{k,i}^+(\theta). \end{aligned}$$

Since

$$\lim_{|\theta|\to\infty} \frac{1}{\theta^2} \int_{\mathbb{R}} \left(\mathrm{e}^{\mathrm{i}\theta x} - 1 - \mathrm{i}\theta x \mathbb{1}_{[-1,1]}(x)\right) \Pi_i(\mathrm{d}x) = 0,$$

and

$$\lim_{|\theta|\to\infty} \frac{1}{|\theta|} \int_0^\infty \left(\mathrm{e}^{\mathrm{i}\theta x} - 1\right) \Pi_i^+(\mathrm{d}x) = 0, \qquad \lim_{|\theta|\to\infty} \frac{1}{|\theta|} \int_0^\infty \left(\mathrm{e}^{-\mathrm{i}\theta x} - 1\right) \Pi_i^-(\mathrm{d}x) = 0,$$

and moreover $|G_{k,i}^-(-\theta) G_{k,i}^+(\theta)| \leq 1$, comparing coefficients yields $b_i^2 = 2d_i^+ d_i^-$. ■

Thus, $b_i > 0$ if and only if $d_i^+ \wedge d_i^- > 0$ and therefore Theorem 4.19 shows that for any MAP with tight overshoots and some Lévy component $\boldsymbol{\xi}^{(i)}$ with non-zero diffusion component, convergence to the stationary overshoot distribution takes place in total variation.

As a next step we show that under appropriate moment conditions on the Lévy processes and transitional jumps underlying the ascending ladder height MAP, overshoots converge with polynomial rate and in case of existence of exponential moments even exponentially fast. Thus, the speed of convergence is reflected in the tail behavior of the jump measures associated to the Lévy system $\boldsymbol{\Pi}^+$, with light tails giving exponential decay and moderately heavy tails resulting in polynomial decay. For the proof we yet again make use of the resolvent formula (4.10) to find Lyapunov functions needed for the resolvent drift criteria (2.8) and (2.10).

THEOREM 4.22. *Suppose that one of the conditions of Proposition 4.17 is satisfied.*

(i) Suppose there exists $\lambda > 0$ such that the exponential λ-moment exists for all $\boldsymbol{H}^{+,(i)}$, $i \in \Theta$, and for all $\Delta_{i,j}^+$, $i \neq j$, such that $q_{i,j}^+ \neq 0$. Then, for the choice $V_\lambda(x,i) = \exp(\lambda x)$, $(x,i) \in \mathbb{R}_+ \times \Theta$, $(\mathcal{O}, \mathcal{J})$ is $\mathcal{R}_\lambda V_\lambda$-uniformly ergodic, i.e.

$$\sup_{|f| \leq \mathcal{R}_\lambda V_\lambda} \left|\mathbb{E}^{x,i}[f(\mathcal{O}_t, \mathcal{J}_t)] - \mu(f)\right| \leq C\mathcal{R}_\lambda V_\lambda(x,i)\mathrm{e}^{-\kappa t}, \quad (x,i) \in \mathbb{R}_+ \times \Theta,$$

for some universal constants $C, \kappa > 0$. Moreover for any $\delta \in (0,1)$, it holds that

$$\|\mathbb{P}^{x,i}((\mathcal{O}_t, \mathcal{J}_t) \in \cdot) - \mu\|_{\mathrm{TV}} \leq \mathfrak{C}(\delta)\mathcal{R}_\lambda V_\lambda(x,i)\mathrm{e}^{-t/(2+\delta)}, \quad (x,i) \in \mathbb{R}_+ \times \Theta, \tag{4.16}$$

for some constant $\mathfrak{C}(\delta) > 0$.

(ii) Suppose that for some $\lambda > 1$ the λ-moment exists for all $\boldsymbol{H}^{+,(i)}$, $i \in \Theta$, and for all $\Delta^+_{i,j}$, $i \neq j$, such that $q^+_{i,j} \neq 0$. Then, there exists $\widetilde{C} > 0$ such that

$$\|\mathbb{P}^{x,i}((\mathcal{O}_t, \mathcal{J}_t) \in \cdot) - \mu\|_{\mathrm{TV}} \leq \widetilde{C}\mathcal{R}_\lambda \widetilde{V}_\lambda(x,i)t^{1-\lambda}, \quad (x,i) \in \mathbb{R}_+ \times \Theta,$$

where $\widetilde{V}_\lambda(x,i) = \mathrm{e}^{\lambda x}\mathbb{1}_{[0,1)}(x) + x^\lambda \mathbb{1}_{[1,\infty)}(x)$.

Proof. (i) For a matrix $A \in \mathbb{R}^{n\times n}$, let $\|A\|_\infty := \max_{i=1,\dots,n}\sum_{j=1}^n |a_{ij}|$ be its matrix norm induced by the sup-norm. Let Q_λ be the operator from the statement of Theorem 4.7. Then,

$$\lambda Q_\lambda V_\lambda(x,i) = \lambda \int_0^x \mathrm{e}^{-\lambda t} V_\lambda(x-t,i)\,\mathrm{d}t = \frac{1}{2}\big(\mathrm{e}^{\lambda x} - \mathrm{e}^{-\lambda x}\big), \quad (x,i) \in \mathbb{R}_+ \times \Theta.$$

Since by Taylor expansion there exists $a > 0$ such that $\mathrm{e}^{\lambda x} - \mathrm{e}^{-\lambda x} \leq ax$ for any $x \in (0,1)$ and Π^+_i are Lévy subordinator measures, it follows that

$$\int_0^1 \lambda Q_\lambda V_\lambda(x,i)\,\Pi^+_i(\mathrm{d}x) < \infty.$$

Moreover, by assumption $\boldsymbol{H}^{+,(i)}$ has an exponential λ-moment, which according to Theorem 3.6 of [109] is equivalent to $\int_1^\infty \exp(\lambda x)\,\Pi^+_i(\mathrm{d}x) < \infty$, implying that

$$\int_1^\infty \lambda Q_\lambda V_\lambda(x,i)\,\Pi^+_i(\mathrm{d}x) < \infty$$

as well and thus

$$\int_0^\infty \lambda Q_\lambda V_\lambda(x,i)\,\Pi^+_i(\mathrm{d}x) < \infty$$

for all $i \in \Theta$. Since additionally $\mathbb{E}[\exp(\lambda\Delta^+_{i,j})] < \infty$ for any $i,j \in \Theta$ such that $i \neq j$ and $q^+_{i,j} > 0$, it follows that if we define

$$\begin{aligned} b &:= \lambda\|\boldsymbol{\Phi}^+(\lambda)^{-1}\|_\infty \sum_{i=1}^n \Bigg(d^+_i + \int_0^\infty Q_\lambda V_\lambda(x,i)\,\Pi^+_i(\mathrm{d}x) + \sum_{j\neq i} q^+_{i,j}\mathbb{E}[Q_\lambda V_\lambda(\Delta^+_{i,j}, j)]\Bigg) \\ &\leq \|\boldsymbol{\Phi}^+(\lambda)^{-1}\|_\infty \sum_{i=1}^n \Bigg(\lambda d^+_i + \frac{1}{2}\int_0^\infty \big(\mathrm{e}^{\lambda x} - \mathrm{e}^{-\lambda x}\big)\,\Pi^+_i(\mathrm{d}x) + \sum_{j\neq i}\frac{q^+_{i,j}}{2}\mathbb{E}\big[\exp\big(\lambda\Delta^+_{i,j}\big)\big]\Bigg), \end{aligned}$$

we have $b < \infty$. Using (4.10) it therefore follows for any $i \in \Theta$ that

$$\mathcal{R}_\lambda V_\lambda(x,i) = \lambda\mathcal{U}_\lambda V_\lambda(x,i) \leq \frac{1}{2}\big(\mathrm{e}^{\lambda x} - \mathrm{e}^{-\lambda x}\big) + b < \frac{1}{2}V_\lambda(x,i) + b, \quad (x,i) \in \mathbb{R}_+ \times \Theta, \tag{4.17}$$

which shows that (2.8) holds for $\beta_0 = 1/2$ and $b < \infty$ as above. Under the given assumptions, $(\mathcal{O}, \mathcal{J})$ is Harris recurrent and there exists an irreducible skeleton chain by

Proposition 4.9 and Proposition 4.17, hence $(\mathfrak{O}, \mathfrak{J})$ is irreducible and aperiodic. Moreover, V_λ is unbounded off petite sets since V_λ is increasing and continuous and hence for any $z > 0$, the set $\{(x,i) \in \mathbb{R}_+ \times \Theta : V_\lambda(x,i) \le z\}$ is compact and hence petite, according to Corollary 4.10. Thus, (2.8) being satisfied for our choice of V_λ, Theorem 5.2 in [74] implies that $(\mathfrak{O}, \mathfrak{J})$ is $\mathcal{R}_\lambda V_\lambda$-uniformly ergodic.

To establish the more explicit rate of convergence for the total variation norm in (4.16), note that (4.17) combined with (2.9) shows that for the petite set $C(\varepsilon) = \{V_\lambda \le 2b/\varepsilon\}$, $\varepsilon \in (0,1)$ and $\phi(z) = (1-\varepsilon)z/2$ we have

$$\mathcal{R}_\lambda V_\lambda(x,i) \le \frac{1+\varepsilon}{2} V_\lambda(x,i) + b\mathbb{1}_{C(\varepsilon)} = V_\lambda(x,i) - \phi \circ V_\lambda(x,i) + b\mathbb{1}_{C(\varepsilon)}, \quad (x,i) \in \mathbb{R}_+ \times \Theta,$$

and thus, the claim follows easily from (2.11).

(ii) Since $\widetilde{V}_\lambda(x,i) = V_\lambda(x,i)$ for $x \in [0,1)$, $i \in \Theta$, it follows from above that

$$\int_0^1 \lambda Q_\lambda \widetilde{V}_\lambda(x,i)\, \Pi_i^+(\mathrm{d}x) < \infty.$$

Moreover, for $x \ge 1$ we have $\lambda Q_\lambda \widetilde{V}_\lambda(x,i) \le x^\lambda$ and thus by our moment assumptions on $\boldsymbol{H}^{+,(i)}$ and $\Delta_{i,j}^+$

$$\int_1^\infty \lambda Q_\lambda \widetilde{V}_\lambda(x,i)\, \Pi_i^+(\mathrm{d}x) < \infty, \quad \mathbb{E}[\lambda Q_\lambda \widetilde{V}_\lambda(\Delta_{i,j}^+, j)] < \infty.$$

This shows that

$$\widetilde{b} := \lambda \|\boldsymbol{\Phi}^+(\lambda)^{-1}\|_\infty \sum_{i=1}^n \left(d_i^+ + \int_0^\infty Q_\lambda \widetilde{V}_\lambda(x,i)\, \Pi_i^+(\mathrm{d}x) + \sum_{j \ne i} q_{i,j}^+ \mathbb{E}[Q_\lambda \widetilde{V}_\lambda(\Delta_{i,j}^+, j)] \right) < \infty.$$

Observe now that integrating by parts twice yields that for $x \ge 1$ and $i \in \Theta$,

$$\lambda Q_\lambda \widetilde{V}_\lambda(x,i) \le \widetilde{V}_\lambda(x,i) - x^{\lambda-1} + \mathrm{e}^{-\lambda x}\left(\mathrm{e}^\lambda + \int_1^x (\lambda-1)\mathrm{e}^{\lambda t} t^{\lambda-2}\, \mathrm{d}t \right)$$

and for $x \in [0,1)$,

$$\lambda Q_\lambda \widetilde{V}_\lambda(x,i) \le \mathrm{e}^{\lambda x}.$$

Thus, for all $(x,i) \in \mathbb{R}_+ \times \Theta$, we have

$$\lambda Q_\lambda \widetilde{V}_\lambda(x,i) \le \widetilde{V}_\lambda(x,i) - (\widetilde{V}_\lambda(x,i))^{\frac{\lambda-1}{\lambda}} + \mathrm{e}^{-\lambda x}\left(\mathrm{e}^\lambda + \int_1^x (\lambda-1)\mathrm{e}^{\lambda t} t^{\lambda-2}\, \mathrm{d}t \right) + \mathrm{e}^{\lambda-1}\mathbb{1}_{[0,1]}(x)$$

and hence by the resolvent formula and the definiton of $\widetilde{b}$,

$$\mathcal{R}_\lambda \widetilde{V}_\lambda(x,i) \le \widetilde{V}_\lambda(x,i) - (\widetilde{V}_\lambda(x,i))^{\frac{\lambda-1}{\lambda}} + \mathrm{e}^{-\lambda x}\left(\widetilde{b} + \mathrm{e}^\lambda + \int_1^x (\lambda-1)\mathrm{e}^{\lambda t} t^{\lambda-2}\, \mathrm{d}t \right) + \mathrm{e}^{\lambda-1}\mathbb{1}_{[0,1]}(x). \tag{4.18}$$

Let $x^* > 1$ be large enough such that for all $x > x^*$

$$\psi_\lambda(x) := \mathrm{e}^{-\lambda x}\left(\widetilde{b} + \mathrm{e}^\lambda + \int_1^x (\lambda-1)\mathrm{e}^{\lambda t} t^{\lambda-2}\, \mathrm{d}t \right) \le \frac{1}{2} x^{\lambda-1}.$$

By the same arguments as in the previous part, the compact set $C := [0, x^*] \times \Theta$ is petite and it follows from (4.18) that

$$\mathcal{R}_\lambda \widetilde{V}_\lambda \leq \widetilde{V}_\lambda - \phi \circ \widetilde{V}_\lambda + \widetilde{c}\mathbb{1}_C, \tag{4.19}$$

where $\widetilde{c} := \mathrm{e}^{\lambda-1} + \max_{x\in[0,x^*]} \psi_\lambda(x) < \infty$ and $\phi(z) = \frac{1}{2} z^{1-1/\lambda}$, $z \geq 1$, is concave, differentiable and increasing. Hence, (2.10) is satisfied. The assertion now follows from (2.11) upon noting that

$$H_\phi(t) = \int_1^t (1/\phi(s))\,\mathrm{d}s = 2\lambda(t^{1/\lambda} - 1), \quad H_\phi^{-1}(t) = \Big(1 + \frac{t}{2\lambda}\Big)^\lambda,$$

and therefore the rate of convergence $\Xi(t)$ defined in Appendix 2 is given by

$$\Xi(t) = 1/(\phi \circ H_\phi^{-1})(t) = 2\Big(1 + \frac{t}{2\lambda}\Big)^{1-\lambda} \leq 2(2\lambda)^{\lambda-1} t^{1-\lambda}.$$

∎

Remark 4.23. Let us emphasize that the parameter λ in the exponential λ-moment assumption from part (i) is only reflected in a multiplicative constant and hence negligible for the speed of convergence. This is particularly nice for statistical considerations, where exact control over the speed of convergence is decisive for making a minimax approach over entire classes of MAPs feasible. Moreover, our analysis of the mixing behavior of self-similar Markov processes later on profits immensly from this exact rate in terms of expliciteness, since the Lamperti–Kiu transform turns the exponential rate into a polynomial one. On the other hand, if the MAP components only have some finite λ-moment, then the maximal size of λ is highly relevant for the polynomial speed of convergence.

Finally, we will establish polynomial β-mixing of stationary overshoots provided that the process converges at polynomial rate and make use of Masuda's criterion for exponential β-mixing given exponential ergodicity of a Markov process stated in (2.15) to establish exponential β-mixing of overshoots for any initial distribution with exponential moments. To this end, we need one more technical result, which is the natural generalization of a result well known for renewal functions of Lévy subordinators to the MAP situation.

LEMMA 4.24. *For any $x, y > 0$ and $i, j \in \Theta$ it holds*

$$U_{i,j}^+(x + y) - U_{i,j}^+(x) \leq U_{j,j}^+(y).$$

Proof. Let $(\theta_t^+)_{t\geq 0}$ be the family of transition operators for the Markov process (H^+, J^+). Then, with a change of variables and an application of the strong Markov property it follows

$$\begin{aligned} U_{i,j}^+(x+y) - U_{i,j}^+(x) &= \mathbb{E}^{0,i}\Bigg[\int_{T_x^+}^{T_{x+y}^+} \mathbb{1}_{\{J_t^+=j\}}\,\mathrm{d}t\Bigg] = \mathbb{E}^{0,i}\Bigg[\int_0^{T_{x+y}^+\circ\theta_{T_x^+}^+ + T_x^+} \mathbb{1}_{\{J_{t+T_x^+}^+=j\}}\,\mathrm{d}t\Bigg] \\ &= \mathbb{E}^{0,i}\Bigg[\int_0^{T_{x+y}^+} \mathbb{1}_{\{J_t^+=j\}}\,\mathrm{d}t \circ \theta_{T_x^+}^+\Bigg] = \mathbb{E}^{0,i}\Bigg[\mathbb{E}^{H_{T_x^+}^+, J_{T_x^+}^+}\Bigg[\int_0^{T_{x+y}^+} \mathbb{1}_{\{J_t^+=j\}}\,\mathrm{d}t\Bigg]\Bigg] \\ &\leq \mathbb{E}^{0,i}\Big[U_{J_{T_x^+}^+,j}^+(y)\Big] \leq U_{j,j}^+(y). \end{aligned}$$

Here, we used for the first inequality, that by spatial homogeneity of a MAP, $\{T^+_{x+y}, \mathbb{P}^{z,k}\} \stackrel{\mathrm{d}}{=} \{T^+_{x+y-z}, \mathbb{P}^{0,k}\}$ for any $z \geq 0$ and combinded this observation with the fact that $z \mapsto T^+_z$ is increasing and that $H^+_{T^+_x} \geq x$ by definition. The second inequality follows from $U^+_{i,j}(y) \leq U^+_{j,j}(y)$ thanks to increasing paths of $\boldsymbol{H}^+$. ∎

Recall the definition of the β-mixing coefficient from (2.14).

THEOREM 4.25. *Suppose that one of the conditions of Proposition 4.17 is satisfied.*

(i) Suppose that the exponential moment assumption from Theorem 4.22.(i) is satisfied and let η be a probability measure on $(\mathbb{R}_+ \times \Theta, \mathcal{B}(\mathbb{R}_+ \times \Theta))$ such that $\eta(\cdot, \Theta)$ has an exponential λ-moment. Then, $(\mathbf{O}, \mathbf{J})$ started in η is exponentially β-mixing with the β-mixing coefficient $\beta(\eta, \cdot)$ satisfying
$$\beta(\eta, t) \leq 2\varrho(\eta, \lambda, \delta)\mathrm{e}^{-t/(2+\delta)},$$
for
$$\varrho(\eta, \lambda) := \mathfrak{C}(\delta) \sup_{t\geq 0} \int_{\mathbb{R}_+\times\Theta} \mathcal{R}_\lambda V_\lambda(y, z)\, \mathbb{P}^\eta(\mathcal{O}_t \in \mathrm{d}y, \mathcal{J}_t \in \mathrm{d}z) < \infty,$$
for some constant $\mathfrak{C}(\delta) > 0$ and $V_\lambda(x, i) = \exp(\lambda x)$, $(x, i) \in \mathbb{R}_+ \times \Theta$.

(ii) Suppose that the λ-moment assumption from Theorem 4.22.(ii) is satisfied for some $\lambda > 2$. Then, $(\mathcal{O}_t, \mathcal{J}_t)_{t\geq 0}$ started in its invariant distribution is β-mixing with rate
$$\beta(\mu, t) \lesssim t^{2-\lambda}, \quad t \geq 0.$$

Proof. (i) By Theorem 4.22, for any $\delta \in (0, 1)$ there exists $\mathfrak{C}(\delta) > 0$ s.t.
$$\|\mathbb{P}^{x,i}(\mathcal{O}_t \in \cdot) - \mu\|_{\mathrm{TV}} \leq \mathfrak{C}(\delta)\mathcal{R}_\lambda V_\lambda(x, i)\mathrm{e}^{-t/(2+\delta)}, \quad (x, i) \in \mathbb{R}_+ \times \Theta.$$

Hence, the assertion will follow from Lemma 3.9 in Masuda [123] if we can establish that $\varrho(\eta, \lambda, \delta) < \infty$. To this end, observe that by the explicit form of the λ-resolvent $\mathcal{U}_\lambda$ of the overshoot process and with the constants a, b appearing in the proof of Theorem 4.22 we have
$$\begin{aligned}
&\int_{\mathbb{R}_+\times\Theta} \mathcal{R}_\lambda V_\lambda(y, u)\, \mathbb{P}^\eta(\mathcal{O}_t \in \mathrm{d}y, \mathcal{J}_t \in \mathrm{d}u)\\
&\quad\leq b + \sum_{i,j=1}^n \frac{1}{2} \int_{\mathbb{R}_+} \int_{\mathbb{R}_+} (\mathrm{e}^{\lambda y} - \mathrm{e}^{-\lambda y})\, \mathbb{P}^{x,i}(\mathcal{O}_t \in \mathrm{d}y, \mathcal{J}_t = j)\, \eta(\mathrm{d}x, \{i\})\\
&\quad\leq b + \sum_{i,j=1}^n \Bigg(\frac{1}{2} \int_{\mathbb{R}_+} \int_1^\infty \mathrm{e}^{\lambda y}\, \mathbb{P}^{x,i}(\mathcal{O}_t \in \mathrm{d}y, \mathcal{J}_t = j)\, \eta(\mathrm{d}x, \{i\})\\
&\qquad\qquad + \frac{a}{2} \int_{\mathbb{R}_+} \int_0^1 y\, \mathbb{P}^{x,i}(\mathcal{O}_t \in \mathrm{d}y, \mathcal{J}_t = j)\, \eta(\mathrm{d}x, \{i\})\Bigg)\\
&\quad\leq b + \sum_{i,j=1}^n \Bigg(\frac{1}{2} \int_{\mathbb{R}_+} \int_1^\infty \mathrm{e}^{\lambda y}\, \mathbb{P}^{x,i}(\mathcal{O}_t \in \mathrm{d}y, \mathcal{J}_t = j)\, \eta(\mathrm{d}x, \{i\}) + \frac{a}{2}\Bigg).
\end{aligned}$$

Hence, to prove the assertion it suffices to show that for any $(i,j) \in \Theta^2$,

$$\sup_{t\geq 0} \int_{\mathbb{R}_+} \int_1^\infty \mathrm{e}^{\lambda y}\, \mathbb{P}^{x,i}(\mathcal{O}_t \in \mathrm{d}y, \mathcal{J}_t = j)\, \eta(\mathrm{d}x, \{i\}) < \infty.$$

With the sawtooth structure, the overshoot formula and multiple uses of Fubini's theorem we obtain for fixed $t \geq 0$,

$$\begin{aligned}
&\int_{\mathbb{R}_+} \int_1^\infty \mathrm{e}^{\lambda y}\, \mathbb{P}^{x,i}(\mathcal{O}_t \in \mathrm{d}y, \mathcal{J}_t = j)\, \eta(\mathrm{d}x, \{i\}) \\
&\qquad \leq \int_0^t \int_1^\infty \mathrm{e}^{\lambda y}\, \mathbb{P}^{0,i}(\mathcal{O}_{t-x} \in \mathrm{d}y, \mathcal{J}_t = j)\, \eta(\mathrm{d}x, \{i\}) + \int_t^\infty \mathrm{e}^{\lambda(x-t)}\, \eta(\mathrm{d}x, \{i\}) \\
&\qquad \leq \int_0^t \eta(\mathrm{d}x, \{i\}) \Bigg(\int_0^{t-x} U^+_{i,j}(\mathrm{d}y) \int_1^\infty \Pi^+_i(t-x-y+\mathrm{d}u)\, \mathrm{e}^{\lambda u} \\
&\qquad\qquad\qquad + \sum_{k\neq j} q^+_{k,j} \int_0^{t-x} U^+_{i,k}(\mathrm{d}y) \int_1^\infty F^+_{k,j}(t-x-y+\mathrm{d}u)\, \mathrm{e}^{\lambda u} \Bigg) + \int_0^\infty \mathrm{e}^{\lambda x}\, \eta(\mathrm{d}x, \{i\}) \\
&\qquad = \int_0^t \eta(\mathrm{d}x, \{i\}) \Bigg(\int_0^{t-x} U^+_{i,j}(\mathrm{d}y) \int_1^\infty \Pi^+_i(\mathrm{d}u)\, \mathrm{e}^{\lambda(u+y+x-t)} \\
&\qquad\qquad\qquad + \sum_{k\neq j} q^+_{k,j} \int_0^{t-x} U^+_{i,k}(\mathrm{d}y) \int_1^\infty F^+_{k,j}(\mathrm{d}u)\, \mathrm{e}^{\lambda(u+y+x-t)} \Bigg) + \int_0^\infty \mathrm{e}^{\lambda x}\, \eta(\mathrm{d}x, \{i\}) \\
&\qquad = \int_1^\infty \mathrm{e}^{\lambda u}\, \Pi^+_i(\mathrm{d}u) \int_0^t \eta(\mathrm{d}x, \{i\})\, \mathrm{e}^{\lambda x} \int_0^{t-x} \mathrm{e}^{\lambda(y-t)}\, U^+_{i,j}(\mathrm{d}y) \\
&\qquad\quad + \sum_{k\neq j} q^+_{k,j} \int_1^\infty \mathrm{e}^{\lambda u}\, F^+_{k,j}(\mathrm{d}u) \int_0^t \eta(\mathrm{d}x, \{i\})\, \mathrm{e}^{\lambda x} \int_0^{t-x} \mathrm{e}^{\lambda(y-t)}\, U^+_{i,k}(\mathrm{d}y) \\
&\qquad\quad + \int_0^\infty \mathrm{e}^{\lambda x}\, \eta(\mathrm{d}x, \{i\})
\end{aligned}$$

From Lemma 4.24 we know that for $t > x$ and $i, j \in \Theta$

$$U^+_{i,j}((x,t]) = U^+_{i,j}(t) - U^+_{i,j}(x) \leq \sum_{k=1}^n U^+_{k,j}(t-x)$$

and thus

$$\begin{aligned}
\int_0^{t-x} \mathrm{e}^{\lambda(y-t)} U^+_{i,j}(\mathrm{d}y) &\leq \int_0^t \mathrm{e}^{\lambda(y-t)} U^+_{i,j}(\mathrm{d}y) = \lambda \int_0^t \int_{-\infty}^y \mathrm{e}^{\lambda(x-t)}\, \mathrm{d}x\, U^+_{i,j}(\mathrm{d}y) \\
&= \lambda \int_{-\infty}^t \mathrm{e}^{\lambda(x-t)} U^+_{i,j}((0 \vee x, t])\, \mathrm{d}x = \mathrm{e}^{-\lambda t} U^+_{i,j}(t) + \lambda \int_0^t \mathrm{e}^{\lambda(x-t)} U^+_{i,j}((x,t])\, \mathrm{d}x \\
&\leq \mathrm{e}^{-\lambda t} U^+_{j,j}(t) + \lambda \int_0^t \mathrm{e}^{\lambda(x-t)} U^+_{j,j}(t-x)\, \mathrm{d}x \\
&\leq \mathrm{e}^{-\lambda t} U^+_{j,j}(t) + \lambda \int_0^\infty \mathrm{e}^{-\lambda z} U^+_{j,j}(z)\, \mathrm{d}z.
\end{aligned}$$

Theorem 28 in [66] tells us that $U^+_{j,j}(z) \sim \mathbb{E}^{0,\pi^+}[H^+_1]^{-1} z\pi^+(j)$ as $z \to \infty$ and since $U^+_{j,j}(z)$ is moreover non-negative and increasing, we conclude that

$$\sup_{t\geq 0}\int_0^t \mathrm{e}^{\lambda(y-t)} U^+_{i,j}(\mathrm{d}y) < \infty.$$

Plugging in now yields

$$\begin{aligned}
&\sup_{t\geq 0}\int_{\mathbb{R}_+}\int_1^\infty \mathrm{e}^{\lambda y}\,\mathbb{P}^{x,i}(\mathcal{O}_t\in \mathrm{d}y, \mathcal{J}_t = j)\,\eta(\mathrm{d}x,\{i\})\\
&\qquad \leq \int_1^\infty \mathrm{e}^{\lambda u}\,\Pi^+_i(\mathrm{d}u)\int_0^\infty \mathrm{e}^{\lambda x}\,\eta(\mathrm{d}x,\{i\})\sup_{t\geq 0}\int_0^t \mathrm{e}^{\lambda(y-t)}\,U^+_{i,j}(\mathrm{d}y)\\
&\qquad\quad + \sum_{k\neq j} q^+_{k,j}\int_1^\infty \mathrm{e}^{\lambda u}\,F^+_{k,j}(\mathrm{d}u)\int_0^\infty \mathrm{e}^{\lambda x}\,\eta(\mathrm{d}x,\{i\})\sup_{t\geq 0}\int_0^t \mathrm{e}^{\lambda(y-t)}\,U^+_{i,k}(\mathrm{d}y)\\
&\qquad\quad + \int_0^\infty \mathrm{e}^{\lambda x}\,\eta(\mathrm{d}x,\Theta)\\
&\qquad < \infty,
\end{aligned}$$

where finiteness is a consequence of the above discussion and our assumptions that $H^{+,(i)}_1$, $\eta(\cdot,\Theta)$ and $\Delta^+_{i,j}$ for $i\neq j$ with $q^+_{i,j}\neq 0$ all have an exponential λ-moment. This finishes the proof.

(ii) By stationarity, it holds that

$$\beta(\mu,t) = \int_{\mathbb{R}_+\times\Theta} \|\mathcal{P}_t((x,z),\cdot)-\mu\|_{\mathrm{TV}}\,\mu(\mathrm{d}x\times \mathrm{d}z) = \sum_{i=1}^n\int_{\mathbb{R}_+}\|\mathcal{P}_t((x,i),\cdot)-\mu\|_{\mathrm{TV}}\,\mu(\mathrm{d}x,\{i\}).$$

Since the $(\lambda-1)$th moments of $H^{+,(i)}_1$ for all $i\in\Theta$ and $\Delta^+_{i,j}$ for all $i,j\in\Theta$ such that $q^+_{i,j}\neq 0$ exist, it follows from Theorem 4.22.(ii) that

$$\beta(\mu,t)\leq \widetilde{C}t^{2-\lambda}\sum_{i=1}^n\int_{\mathbb{R}_+}\mathcal{R}_{\lambda-1}\widetilde{V}_{\lambda-1}(x,i)\,\mu(\mathrm{d}x,\{i\}),$$

and hence, to prove the assertion it is enough to show that the integrals on the right-hand side are finite. From the drift inequality (4.19) established in the proof of Theorem 4.22.(ii) we obtain that for any $i\in\Theta$,

$$\int_{\mathbb{R}_+\times\Theta}\mathcal{R}_{\lambda-1}\widetilde{V}_{\lambda-1}(x,i)\,\mu(\mathrm{d}x,\{i\})\leq \int_0^1 \mathrm{e}^{(\lambda-1)x}\mu(\mathrm{d}x,i)+\widetilde{c}\mu(C)+\int_1^\infty x^{\lambda-1}\,\mu(\mathrm{d}x,\{i\}).$$

Since by our moment assumptions

$$\begin{aligned}
\int_1^\infty x^{\lambda-1}\,\mu(\mathrm{d}x,\{i\}) &= \frac{1}{\mathbb{E}^{0,\pi^+}[H^+_1]}\int_1^\infty x^{\lambda-1}\Big(\pi^+(i)\overline{\Pi}^+_i(x)+\sum_{j\neq i}\pi^+(j)q^+_{j,i}\overline{F}^+_{j,i}(x)\Big)\,\mathrm{d}x,\\
&\leq \frac{1}{\mathbb{E}^{0,\pi^+}[H^+_1]}\Big(\pi^+(i)\int_1^\infty x^\lambda\,\Pi^+_i(\mathrm{d}x)+\sum_{j\neq i}\pi^+(j)q^+_{j,i}\int_1^\infty x^\lambda\,F^+_{j,i}(\mathrm{d}x)\Big)\\
&< \infty,
\end{aligned}\tag{4.20}$$

the assertion follows.

■

Remark 4.26. As in (4.20), it is easily established that when the jump measures associated to $\mathbf{\Pi}^+$ have exponential decay, $\mu(\cdot, \Theta)$ possesses an exponential moment and hence for part (i), the stationary overshoot process is exponentially β-mixing as well.

As a direct corollary we obtain the exponential resp. polynomial β-mixing behavior of MAPs sampled at first hitting times provided that creeping is possible or the Lévy system has some minor regularity and moreover the respective moment conditions on the MAP are satisfied. Let

$$\mathcal{K}_t := \sigma\big((\xi_{T_s}, J_{T_s}), s \leq t\big), \quad \overline{\mathcal{K}}_t := \sigma\big((\xi_{T_s}, J_{T_s}), s \geq t\big), \quad t \geq 0$$

be the σ-algebras generated by the MAP sampled at first hitting times up to level t and from level t onwards, respectively.

Corollary 4.27. *Suppose that the assumptions of Theorem 4.22.(i) are satisfied and let η be a probability measure on $(\mathbb{R}_+ \times \Theta, \mathcal{B}(\mathbb{R}_+ \times \Theta))$ such that $\eta(\cdot, \Theta)$ has an exponential λ-moment. Then, for any $\delta \in (0,1)$,*

$$\sup_{t>0} \beta_{\mathbb{P}^\eta}\big(\mathcal{K}_t, \overline{\mathcal{K}}_{t+s}\big) \leq 2\varrho(\eta, \lambda, \delta)\mathrm{e}^{-s/(2+\delta)}, \quad s > 0,$$

where $\varrho(\eta, \lambda, \delta) > 0$ is the constant from Theorem 4.25. If instead the assumptions from 4.22.(ii) are satisfied with $\lambda > 2$, then

$$\sup_{t>0} \beta_{\mathbb{P}^\mu}\big(\mathcal{K}_t, \overline{\mathcal{K}}_{t+s}\big) \lesssim s^{2-\lambda}, \quad s > 0.$$

4.4 Équations amicales inversés for MAPs

With the help of the spatial Wiener–Hopf factorization for MAPs we can generalize Vigon's équation amicale inversé for Lévy processes to a characterization of the Lévy system of the ascending ladder height MAP in terms of the Lévy system of the parent MAP and the potential measures of the ascending ladder height process of the dual MAP. This is crucial for our results since this relation will allow to impose conditions on the parent MAP instead of the ascending ladder height MAP that imply the overshoot convergence results from the previous section. To this end, we first need to recall some concepts from distribution theory and introduce more notation.

Let $\mathcal{S}(\mathbb{R})$ be the Schwartz space of rapidly decreasing smooth functions on $\mathbb{R}$ and consider its dual space $\mathcal{S}'(\mathbb{R})$, the space of tempered distributions. For $\mu \in \mathcal{S}'(\mathbb{R})$ the k-th derivative $\mu^{(k)} \in \mathcal{S}'(\mathbb{R})$ is defined by

$$\big\langle \mu^{(k)}, \phi \big\rangle = (-1)^k \big\langle \mu, \phi^{(k)} \big\rangle, \quad \phi \in \mathcal{S}(\mathbb{R}), k \in \mathbb{N}.$$

If μ is induced by a function $\psi \in \mathcal{B}(\mathbb{R})$ via

$$\langle \mu, \phi \rangle = \int_{\mathbb{R}} \psi(x)\phi(x)\,\mathrm{d}x, \quad \phi \in \mathcal{S}(\mathbb{R}),$$

we just write $\mu = \psi$, provided that the above integrals are well defined. Similarly, if μ is a measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ such that $\int \phi\,\mathrm{d}\mu$ is well-defined for any $\phi \in \mathcal{S}(\mathbb{R})$, we identify the distribution induced by $\phi \mapsto \int \phi\,\mathrm{d}\mu$ with μ.

For a Lévy measure ν integrating $x \mapsto |x|$ on $[-1,1]$, let $\mathbb{I}\nu$ be the tempered distribution defined via

$$\left\langle \mathbb{I}\nu, \phi \right\rangle := \int_{\mathbb{R}} (\phi(x) - \phi(0))\, \nu(\mathrm{d}x), \quad \phi \in \mathcal{S}(\mathbb{R}),$$

and for a general Lévy measure ν let $\mathbb{I}^2\nu$ be the tempered distribution defined via

$$\left\langle \mathbb{I}^2\nu, \phi \right\rangle := \int_{\mathbb{R}} (\phi(x) - \phi(0) - \phi'(0)x\mathbb{1}_{[-1,1]}(x))\, \nu(\mathrm{d}x), \quad \phi \in \mathcal{S}(\mathbb{R}).$$

Recall that for a tempered distribution $\mu \in \mathcal{S}'(\mathbb{R})$ the Fourier transform $\mathcal{F}\mu$ is defined by

$$\langle \mathcal{F}\mu, \phi \rangle := \langle \mu, \mathcal{F}\phi \rangle = \left\langle \mu, \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}x\cdot} \phi(x)\, \mathrm{d}x \right\rangle, \quad \phi \in \mathcal{S}(\mathbb{R}),$$

and that the Fourier transform is a bijective, continuous mapping on $\mathcal{S}'(\mathbb{R})$. If δ is the Dirac delta distribution and letting $\psi_2(x) = x^2$, $x \in \mathbb{R}$, it is immediate that

$$\mathcal{F}\delta = \mathrm{id}, \quad \mathcal{F}\delta' = -\mathrm{i}\cdot \mathrm{id}, \quad \mathcal{F}\delta'' = -\psi_2.$$

Hence, for a Lévy subordinator with characteristic Fourier exponent κ , Lévy measure ν, drift $d \geq 0$ and killing rate $q \geq 0$ we obtain

$$\left\langle \mathcal{F}\big(- q\delta - d\delta' + \mathbb{I}\nu\big), \phi \right\rangle = \int_{\mathbb{R}} \Big(- q + \mathrm{i}d\theta + \int_{\mathbb{R}} \big(\mathrm{e}^{\mathrm{i}\theta x} - 1\big)\, \nu(\mathrm{d}x)\Big)\phi(\theta)\, \mathrm{d}\theta = \int_{\mathbb{R}} \kappa(\theta)\phi(\theta)\, \mathrm{d}\theta,$$

and therefore it holds that

$$\mathcal{F}^{-1}\kappa = -q\delta - d\delta' + \mathbb{I}\nu.$$

Thus, if $\mathcal{A}^*$ denotes the infinitesimal generator of the subordinator's dual, then

$$\mathcal{A}^* f = (\mathcal{F}^{-1}\kappa) * f, \quad f \in \mathcal{S}(\mathbb{R}).$$

Similarly, we get for the characteristic exponent Ψ of a Lévy process with generating triplet (a, σ^2, ν) and killing rate q that

$$\mathcal{F}^{-1}\Psi = -q\delta - a\delta' + \frac{1}{2}\sigma^2\delta'' + \mathbb{I}^2\nu.$$

We start with a simple lemma. Let

$$\sigma(A) := \sup\{\mathrm{Re}(\lambda) : \lambda \text{ eigenvalue of } A\},$$

be the spectral bound of a quadratic complex matrix A.

LEMMA 4.28. *For any (non-trivial) MAP with characteristic matrix exponent $\boldsymbol{\Psi}$ and $\theta \in \mathbb{R}$, it holds that $\sigma(\boldsymbol{\Psi}(\theta)) \leq 0$ and for any $\lambda > 0$, $\lambda\mathbb{I}_n - \boldsymbol{\Psi}(\theta)$ is invertible.*

Proof. Let $\lambda > 0$ be arbitrary and $\mathbf{e}_\lambda$ be an independent exponential time with mean $1/\lambda$ and define for $x \in \mathbb{R}$, $i, j \in \Theta$,

$$^{\lambda}U_{i,j}(\mathrm{d}x) = \mathbb{E}^{0,i}\Big[\int_0^{\mathbf{e}_\lambda} \mathbb{1}_{\{\xi_t \in \mathrm{d}x, J_t = j\}}\, \mathrm{d}t\Big] = \int_0^\infty \mathbb{P}^{0,i}(\xi_t \in \mathrm{d}x, J_t = j, t < \mathbf{e}_\lambda)\, \mathrm{d}t$$

$$= \int_0^\infty \mathrm{e}^{-\lambda t} \mathbb{P}^{0,i}(\xi_t \in \mathrm{d}x, J_t = j)\,\mathrm{d}t,$$

i.e. ${}^\lambda U_{i,j}$ is the (finite) occupation measure of the MAP started in $(0, i)$, while the modulator $\boldsymbol{J}$ is in state j, killed at an independent exponential time. Clearly,

$$\left\{\mathcal{F}^\lambda U_{i,j}\right\}(\theta) = \left(\int_0^\infty \mathrm{e}^{t(\boldsymbol{\Psi}(\theta)-\lambda \mathbb{I}_n)}\,\mathrm{d}t\right)_{i,j},$$

where for a matrix valued function $f\colon \mathbb{R} \to \mathbb{R}^{n\times n}$, such that $f_{i,j}$ is integrable, $\int_{\mathbb{R}} f(t)\,\mathrm{d}t := (\int_{\mathbb{R}} f_{i,j}(t)\,\mathrm{d}t)_{i,j=1,\dots,n}$. Hence, if we let ${}^\lambda U := ({}^\lambda U_{i,j})_{i,j\in\Theta}$, it follows that

$$\left\{\mathcal{F}\,{}^\lambda U\right\}(\theta) = \int_0^\infty \mathrm{e}^{t(\boldsymbol{\Psi}(\theta)-\lambda \mathbb{I}_n)}\,\mathrm{d}t.$$

Noting that

$$(\lambda \mathbb{I}_n - \boldsymbol{\Psi}(\theta)) \int_0^T \mathrm{e}^{t(\boldsymbol{\Psi}(\theta)-\lambda \mathbb{I}_n)}\,\mathrm{d}t = \mathbb{I}_n - \mathrm{e}^{T(\boldsymbol{\Psi}(\theta)-\lambda \mathbb{I}_n)}, \tag{4.21}$$

and that the left-hand side converges to

$$(\lambda \mathbb{I}_n - \boldsymbol{\Psi}(\theta)) \cdot \left\{\mathcal{F}\,{}^\lambda U\right\}(\theta),$$

as $T \to \infty$, it follows that the matrix exponential $\mathrm{e}^{T(\boldsymbol{\Psi}(\theta)-\lambda \mathbb{I}_n)}$ must converge as well as $T \to \infty$. E.g. from Theorem 4.12 of [21], this can only be the case if $\sigma(\boldsymbol{\Psi}(\theta) - \lambda \mathbb{I}_n) \le 0$. But since $\lambda > 0$ was chosen arbitrarily, it follows that for any $\lambda > 0$, actually $\sigma(\boldsymbol{\Psi}(\theta) - \lambda \mathbb{I}_n) < 0$, implying $\sigma(\boldsymbol{\Psi}(\theta)) \le 0$. Again by Theorem 4.12 of [21], this implies that

$$\lim_{T\to\infty} \mathrm{e}^{T(\boldsymbol{\Psi}(\theta)-\lambda \mathbb{I}_n)} = \mathbb{0}_{n\times n}.$$

Thus, (4.21) yields that $\lambda \mathbb{I}_n - \boldsymbol{\Psi}(\theta)$ is invertible with inverse

$$(\lambda \mathbb{I}_n - \boldsymbol{\Psi}(\theta))^{-1} = \left\{\mathcal{F}\,{}^\lambda U\right\}(\theta). \tag{4.22}$$

■

Remark 4.29. This result generalizes part of Theorem 1 in [96] in the sense that, if we let $\Upsilon(z) = (\mathbb{E}^{0,i}[\exp(z\xi_1); J_1 = j])_{i,j\in\Theta}$ for $z \in \mathbb{C}$ whenever it is defined, $z \mapsto \det(\Upsilon(z) - \lambda \mathbb{I}_n)$ has no zeros on the imaginary axis, without having to assume anything on the jump structure of $(\boldsymbol{\xi}, \boldsymbol{J})$ or irreducibility of $\boldsymbol{J}$.

Let us assume for the rest of this section that

($\mathcal{A}$3) the modulator $\boldsymbol{J}$ of the MAP $(\boldsymbol{\xi}, \boldsymbol{J})$ is irreducible, i.e. $\boldsymbol{Q}$ is an irreducible matrix.

THEOREM 4.30 (Équations amicales inversés for MAPs). *For an appropriate scaling of local time at the supremum it holds for any $i, j \in \Theta$, $i \neq j$ and $x > 0$ that*

$$\Pi_i^+(\mathrm{d}x) = \int_0^\infty \Pi_i(y + \mathrm{d}x)\,U_{i,i}^-(\mathrm{d}y) + \sum_{k\neq i} \frac{\pi(k)}{\pi(i)} q_{k,i} \int_0^\infty F_{k,i}(y + \mathrm{d}x)\,U_{k,i}^-(\mathrm{d}y), \tag{4.23}$$

$$q^+_{i,j}F^+_{i,j}(\mathrm{d}x) = \frac{\pi(j)}{\pi(i)}\int_0^\infty \Pi_j(y+\mathrm{d}x)\,U^-_{j,i}(\mathrm{d}y) + \sum_{k\neq j}\frac{\pi(k)}{\pi(i)}q_{k,j}\int_0^\infty F_{k,j}(y+\mathrm{d}x)\,U^-_{k,i}(\mathrm{d}y). \tag{4.24}$$

and

$$\Pi^-_i(\mathrm{d}x) = \int_0^\infty \Pi_i(-y-\mathrm{d}x)\,U^+_{i,i}(\mathrm{d}y) + \sum_{k\neq i} q_{i,k}\int_0^\infty F_{i,k}(-y-\mathrm{d}x)\,U^+_{k,i}(\mathrm{d}y), \tag{4.25}$$

$$q^-_{i,j}F^-_{i,j}(\mathrm{d}x) = \frac{\pi(j)}{\pi(i)}\Big(\int_0^\infty \Pi_j(-y-\mathrm{d}x)\,U^+_{j,i}(\mathrm{d}y) + \sum_{k\neq j} q_{j,k}\int_0^\infty F_{j,k}(-y-\mathrm{d}x)\,U^+_{k,i}(\mathrm{d}y)\Big). \tag{4.26}$$

Remark 4.31. If we let $\mathbf{\Pi}(\mathrm{d}x) := (\mathbf{\Pi}(i, \mathrm{d}x\times\{j\}))_{i,j=1,\dots,n}$, $\mathbf{\Pi}^+(\mathrm{d}x) := (\mathbf{\Pi}^+(i,\mathrm{d}x\times\{j\}))_{i,j=1,\dots,n}$ and $\boldsymbol{U}^+(\mathrm{d}x) := (U^+_{i,j}(\mathrm{d}x))_{i,j=1,\dots,n}$ (with the analogous definitions for the ascending ladder height process of the dual MAP), then we may compactly express the équations amicales inversés (up to premultiplication of some diagonal matrix corresponding to the scaling of local time at the supremum) for $x > 0$ as

$$\mathbf{\Pi}^+(\mathrm{d}x) = \int_0^\infty \mathbf{\Delta}_{\boldsymbol{\pi}}^{-1}\boldsymbol{U}^-(\mathrm{d}y)^\top\mathbf{\Delta}_{\boldsymbol{\pi}}\mathbf{\Pi}(y+\mathrm{d}x),$$
$$\mathbf{\Pi}^-(\mathrm{d}x) = \int_0^\infty \mathbf{\Delta}_{\boldsymbol{\pi}}^{-1}\big(\mathbf{\Pi}(-y-\mathrm{d}x)\,\boldsymbol{U}^+(\mathrm{d}y)\big)^\top\mathbf{\Delta}_{\boldsymbol{\pi}},$$

where $\int_0^\infty (g_{i,j}(y)\,\nu_{i,j}(\mathrm{d}y))_{i,j=1,\dots,n} := (\int_0^\infty g_{i,j}(y)\,\nu_{i,j}(\mathrm{d}y))_{i,j=1,\dots,n}$ for integrable functions $g_{i,j}$ on $(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+), \nu_{i,j})$.

Proof of Theorem 4.30. Analogously to Vigon's [172] idea, we use inverse Fourier transformations of the quantities involved in the spatial Wiener–Hopf factorization for MAPs to prove the desired equalitites. To this end, recall from (4.5) that for an appropriate scaling of local time at the supremum, it holds that

$$\mathbf{\Psi}(\theta) = -\mathbf{\Delta}_{\boldsymbol{\pi}}^{-1}\mathbf{\Psi}^-(-\theta)^\top\mathbf{\Delta}_{\boldsymbol{\pi}}\mathbf{\Psi}^+(\theta), \quad \theta\in\mathbb{R}. \tag{4.27}$$

Rearranging yields for any $\lambda > 0$,

$$\mathbf{\Psi}^+(\theta) = -\mathbf{\Delta}_{\boldsymbol{\pi}}^{-1}\Big(\big(\mathbf{\Psi}^-(-\theta) - \lambda\mathbb{I}_n\big)^{-1}\Big)^\top\mathbf{\Delta}_{\boldsymbol{\pi}}\big(\mathbf{\Psi}(\theta)+\lambda\mathbf{\Psi}^+(\theta)\big), \quad \theta\in\mathbb{R}, \tag{4.28}$$

where invertibility of $\mathbf{\Psi}^-(-\theta) - \lambda\mathbb{I}_n$ is shown in Lemma 4.28. By the form of the characteristic matrix exponent of a MAP it follows by taking inverse Fourier transformation of the distribution induced by the left-hand side that

$$\mathcal{F}^{-1}\mathbf{\Psi}^+_{i,j} = \mathbb{1}_{\{i=j\}}\big((q^+_{i,i} - \dagger^+_i)\delta - d^+_i\delta' + \Gamma\Pi^+_i\big) + \mathbb{1}_{\{i\neq j\}}q^+_{i,j}F^+_{i,j}.$$

Note that by (4.22)

$$\Big(\lambda\mathbb{I}_n - \mathbf{\Psi}^-(-\cdot)\Big)^{-1}_{i,j} = \mathcal{F}\,{}^\lambda\widetilde{U}^-_{i,j},$$

where for an independent exponentially distributed random variable $\mathbf{e}_\lambda$ with mean $1/\lambda$ we define

$${}^\lambda\widetilde{U}^-_{i,j}(\mathrm{d}x) := \widehat{\mathbb{E}}^{0,i}\Big[\int_0^{\mathbf{e}_\lambda}\mathbb{1}_{\{-H^+_t\in\mathrm{d}x, J^+_t=j\}}\,\mathrm{d}t\Big], \quad x\in\mathbb{R}.$$

With this observation, our previous discussion of inverse Fourier transforms of Lévy characteristic exponents and the property that if we regard two tempered distributions whose Fourier transforms are induced by some measurable functions, the Fourier transform of the convolution of those distributions becomes the tempered distribution induced by the product of the functions, it follows that the inverse Fourier transformation of the distribution induced by the right-hand side of (4.28) may be written as

$$
\begin{aligned}
-\mathcal{F}^{-1}&\Big(\boldsymbol{\Delta}_{\boldsymbol{\pi}}^{-1}\Big(\big(\boldsymbol{\Psi}^-(-\cdot) - \lambda\mathbb{I}_n\big)^{-1}\Big)^{\top}\boldsymbol{\Delta}_{\boldsymbol{\pi}}\big(\boldsymbol{\Psi} + \lambda\boldsymbol{\Psi}^+\big)\Big)_{i,j} \\
&= -\sum_{k=1}^{n}\frac{\pi(k)}{\pi(i)}\,\mathcal{F}^{-1}\Big(\big(\boldsymbol{\Psi} + \lambda\boldsymbol{\Psi}^+\big)_{k,j}\big(\boldsymbol{\Psi}^-(-\cdot) - \lambda\mathbb{I}_n\big)^{-1}_{k,i}\Big) \\
&= \frac{\pi(j)}{\pi(i)}\Big(\mathbb{I}^2\Pi_j + \lambda\mathbb{I}\Pi_j^+ - (a_j + \lambda d_j^+)\delta' + \tfrac{1}{2}\sigma^2\delta'' + (q_{j,j} + \lambda(q_{j,j}^+ - \dagger_j^+))\delta\Big) * {}^{\lambda}\widetilde{U}^-_{j,i} \\
&\quad + \sum_{k\neq j}\frac{\pi(k)}{\pi(i)}\big(q_{k,j}F_{k,j} + \lambda q_{k,j}^+F_{k,j}^+\big) * {}^{\lambda}\widetilde{U}^-_{k,i}.
\end{aligned}
$$

Observe that the restriction of $\mathbb{I}\Pi_j^+$ and $\mathbb{I}^2\Pi_j$ to the space $\mathcal{D}_+$ of smooth functions on $\mathbb{R}$ with compact support in $(0,\infty)$ is equal to the distributions induced by Π_j^+ and Π_j on this space, see also Propriété 3.9 in [172]. Restricting to $(0,\infty)$ and equating both sides therefore yields the equality of distributions on $\mathcal{D}'_+$,

$$
\mathbb{1}_{\{i=j\}}\Pi_i^+ + \mathbb{1}_{\{i\neq j\}}q_{i,j}^+F_{i,j}^+ = \frac{\pi(j)}{\pi(i)}\big(\Pi_j + \lambda\Pi_j^+\big) * {}^{\lambda}\widetilde{U}^-_{j,i} + \sum_{k\neq j}\frac{\pi(k)}{\pi(i)}\big(q_{k,j}F_{kj} + \lambda q_{k,j}^+F_{k,j}^+\big) * {}^{\lambda}\widetilde{U}^-_{k,i}. \quad (4.29)
$$

Here, we used that for a measure μ on $\mathbb{R}$ such that the distribution $\mu * {}^{\lambda}\widetilde{U}^-_{k,i}$, is well-defined it holds that

$$
\big(\mu * {}^{\lambda}\widetilde{U}^-_{k,i}\big)\big|_{(0,\infty)} = \big(\mu|_{(0,\infty)} * {}^{\lambda}\widetilde{U}^-_{k,i}\big)\big|_{(0,\infty)},
$$

since ${}^{\lambda}\widetilde{U}^-_{k,i}$ has support $\mathbb{R}_-$, see also Propriété 3.8 in [172]. Denote

$$
\widetilde{U}^-_{k,i}(\mathrm{d}x) := \widehat{\mathbb{E}}^{0,k}\Big[\int_0^\infty \mathbb{1}_{\{-H_t^+\in \mathrm{d}x, J_t^+=i\}}\,\mathrm{d}t\Big], \quad x\in\mathbb{R},
$$

and let $\phi \in \mathcal{B}_b((0,\infty))$ be non-negative with support $\operatorname{supp}(\phi) \subset (a,b)$, where $0 < a < b < \infty$. Utilizing the strong Markov property and conditional spatial homogeneity of $(\boldsymbol{H}^-, \boldsymbol{J}^-)$ we can calculate as follows:

$$
\begin{aligned}
\int_0^\infty \phi(z)\,\Pi_j^+ * \widetilde{U}^-_{j,i}(\mathrm{d}z) &= \int_{-\infty}^0\int_{(a,b)}\phi(z)\,\Pi_j^+(\mathrm{d}z - y)\,\widetilde{U}^-_{j,i}(\mathrm{d}y) \\
&\le \widehat{\mathbb{E}}^{0,j}\Big[\int_0^\infty\int_{(a+H_t^+,b+H_t^+)}\phi(z - H_t^+)\,\Pi_j^+(\mathrm{d}z)\mathbb{1}_{\{J_t^+=i\}}\,\mathrm{d}t\Big] \\
&\le \|\phi\|_\infty\int_{(a,\infty)}\widehat{\mathbb{E}}^{0,j}\Big[\int_{(T^+_{z-b},T^+_{z-a})}\mathbb{1}_{\{J_t^+=i\}}\,\mathrm{d}t\Big]\,\Pi_j^+(\mathrm{d}z) \\
&= \|\phi\|_\infty\int_{(a,\infty)}\widehat{\mathbb{E}}^{0,j}\Big[\widehat{\mathbb{E}}^{H^+_{T^+_{z-b}},J^+_{T^+_{z-b}}}\Big[\int_{(0,T^+_{z-a})}\mathbb{1}_{\{J_t^+=i\}}\,\mathrm{d}t\Big]\Big]\,\Pi_j^+(\mathrm{d}z)
\end{aligned}
$$

$$
\begin{aligned}
&\leq \|\phi\|_\infty \int_{(a,\infty)} \widehat{\mathbb{E}}^{0,j}\Big[\widehat{\mathbb{E}}^{0,J^+_{T^+_{z-b}}}\Big[\int_{(0,T^+_{b-a})} \mathbb{1}_{\{J^+_t=i\}}\,\mathrm{d}t\Big]\Big]\,\Pi^+_j(\mathrm{d}z)\\
&\leq \|\phi\|_\infty \Pi^+_j((a,\infty)) \sum_{k=1}^n U^-_{k,i}(b-a)\\
&< \infty,
\end{aligned}
$$

and similarly, $\int_0^\infty \phi(z)\,\Pi_j * \widetilde{U}^-_{j,i}(\mathrm{d}z) < \infty$. Thus, using monotone convergence for the corresponding integrals wrt. the positive and negative part of a function $\phi \in \mathcal{D}_+$, we have $\langle \Pi^+_j * {}^\lambda\widetilde{U}^-_{j,i}, \phi\rangle \to \langle \Pi^+_j * \widetilde{U}^-_{j,i}, \phi\rangle$ and $\langle \Pi_j * {}^\lambda\widetilde{U}^-_{j,i}, \phi\rangle \to \langle \Pi_j * \widetilde{U}^-_{j,i}, \phi\rangle$ as $\lambda \downarrow 0$. Consequently, $\lambda\Pi^+_j * {}^\lambda\widetilde{U}^-_{j,i} \to 0$ and $\Pi_j * {}^\lambda\widetilde{U}^-_{j,i} \to \Pi_j * \widetilde{U}^-_{j,i}$ on $\mathcal{D}'_+$ as $\lambda \downarrow 0$. Similarly, we obtain $F^+_{k,j} * {}^\lambda\widetilde{U}^-_{k,i} \to F^+_{k,j} * \widetilde{U}^-_{k,i}$ as $\lambda \downarrow 0$. Thus, letting $\lambda \downarrow 0$ in (4.29) implies that restricted to $\mathcal{D}_+$ we have

$$
\mathbb{1}_{\{i=j\}}\Pi^+_i + \mathbb{1}_{\{i\neq j\}} q^+_{i,j} F^+_{i,j} = \frac{\pi(j)}{\pi(i)}\,\Pi_j * \widetilde{U}^-_{j,i} + \sum_{k\neq j} \frac{\pi(k)}{\pi(i)} q_{k,j}\, F_{k,j} * \widetilde{U}^-_{k,i}.
$$

The relations (4.23) and (4.24) follow upon noting that by a monotone class argument σ-finite measures with support on $(0,\infty)$ are uniquely characterized by their action on $\mathcal{D}_+$ and observing that $U^-_{i,j}(\mathrm{d}y) = \widetilde{U}^-_{i,j}(-\mathrm{d}y)$ for $y \geq 0$. Relations (4.25) and (4.26) are proved similarly by taking inverse Fourier transforms on both sides of of

$$
\mathbf{\Psi}^- = -\mathbf{\Delta}_{\boldsymbol{\pi}}^{-1}\Big(\big(\mathbf{\Psi}^+(-\cdot) - \lambda\mathbb{I}_n\big)^{-1}\Big)^{\top}\big(\mathbf{\Psi}(-\cdot)^{\top} + \lambda\mathbf{\Delta}_{\boldsymbol{\pi}}\mathbf{\Psi}^-\mathbf{\Delta}_{\boldsymbol{\pi}}^{-1}\big)\mathbf{\Delta}_{\boldsymbol{\pi}}, \quad \lambda > 0,
$$

which is a rearranged version of (4.27). ∎

Without loss of generality, for the remainder of this section we fix a scaling of local time at the supremum such that (4.5) is satisfied and hence the formulas given in Theorem 4.30 hold without further multiplicative constants. As a first consequence of the équations amicales inversés, we obtain a characterization of $\boldsymbol{Q}^+$ in terms of the transitional jumps of $(\boldsymbol{\xi}, \boldsymbol{J})$, which we made use of in Proposition 4.6.

LEMMA 4.32. *Suppose that for $i, j \in \Theta$ with $i \neq j$, we have $\operatorname{supp}(q_{i,j}F_{i,j}) \cap (0,\infty) \neq \emptyset$. Then, $q^+_{i,j} > 0$.*

Proof. By assumption, there exists $\varepsilon > 0$ such that $q_{i,j}\overline{F}_{i,j}(z) > 0$ for all $z \in (0,\varepsilon)$. Note also that $U^-_{i,i}([0,\varepsilon)) > 0$ by increasing and right-continuous paths of $(\boldsymbol{H}^+, \boldsymbol{J}^+)$ under $\widehat{\mathbb{P}}^{0,i}$. Plugging $(0,\infty)$ into (4.24) therefore yields

$$
q^+_{i,j} \geq \int_0^\infty q_{i,j}\overline{F}_{i,j}(z)\,U^-_{i,i}(\mathrm{d}z) \geq \int_0^\varepsilon q_{i,j}\overline{F}_{i,j}(z)\,U^-_{i,i}(\mathrm{d}z) > 0.
$$

∎

Another simple consequence is the following.

LEMMA 4.33. *If for some $j \in \Theta$, $\boldsymbol{\xi}^{(j)}$ has infinite jump activity on $\mathbb{R}_+$, i.e. $\Pi_j(\mathbb{R}_+) = \infty$, then $U^-_{j,i}$ does not have an atom at 0 for all $i \neq j$.*

Proof. Suppose that there exists $i \neq j$ s.t. $U^-_{j,i}(\{0\}) = \alpha > 0$. Then, again plugging $(0, \infty)$ into (4.24), implies

$$q^+_{i,j} \geq \frac{\pi(j)}{\pi(i)} \alpha \Pi_j(\mathbb{R}_+) = \infty,$$

which is impossible. ■

We can also use the équations amicales inversés to express our assumptions from Section 4.3 on the ascending ladder height process $(\boldsymbol{H}^+, \boldsymbol{J}^+)$ needed for ergodicity. That is, we can verify the conditions on the smoothness of the Lévy system required in Proposition 4.17 and the moment assumptions on the underlying Lévy processes and the transitional jumps required in Theorem 4.22 for exponential or polynomial ergodicity of overshoots, in terms of related conditions on the parent MAP $(\boldsymbol{\xi}, \boldsymbol{J})$.

Lemma 4.34. *(i) If there exists $i \in \Theta$ and $0 \leq a < b \leq \infty$ such that $\boldsymbol{\lambda}|_{(a,b)} \ll \Pi_i|_{(a,b)}$, then also $\boldsymbol{\lambda}|_{(a,b)} \ll \Pi^+_i|_{(a,b)}$.*

(ii) If there exists $i, j \in \Theta$ with $i \neq j$ and $0 \leq a < b \leq \infty$ such that $\boldsymbol{\lambda}|_{(a,b)} \ll q_{i,j} F_{i,j}|_{(a,b)}$, then also $\boldsymbol{\lambda}|_{(a,b)} \ll q^+_{i,j} F^+_{i,j}|_{(a,b)}$.

(iii) For fixed $i \in \Theta$, $\mathbb{E}[\exp(\lambda {H_1}^{+,(i)})] < \infty$ if

$$\int_1^\infty \mathrm{e}^{\lambda x}\, \Pi_i(\mathrm{d}x) + \sum_{k \neq i} q_{k,i} \int_1^\infty \mathrm{e}^{\lambda x}\, F_{k,i}(\mathrm{d}x) < \infty. \tag{4.30}$$

(iv) For fixed $i, j \in \Theta$ such that $q^+_{j,i} \neq 0$ and $\lambda > 0$, $\mathbb{E}[\exp(\lambda \Delta^+_{j,i})] < \infty$ if (4.30) holds.

(v) If $\lim_{t \to \infty} \xi_t = \infty$ a.s., then for $\lambda > 0$ and $i \in \Theta$, $\mathbb{E}[(H_1^{+,(i)})^\lambda] < \infty$ if

$$\int_1^\infty x^\lambda\, \Pi_i(\mathrm{d}x) + \sum_{k \neq i} q_{k,i} \int_1^\infty x^\lambda\, F_{k,i}(\mathrm{d}x) < \infty,$$

and for $i, j \in \Theta$ such that $q^+_{i,j} \neq 0$, $\mathbb{E}[(\Delta^+_{i,j})^\lambda] < \infty$ if

$$\int_1^\infty x^\lambda\, \Pi_j(\mathrm{d}x) + \sum_{k \neq i} q_{k,j} \int_1^\infty x^\lambda\, F_{k,j}(\mathrm{d}x) < \infty.$$

Proof.

(i) Let $B \subset (a, b)$ be a Borel set s.t. $\boldsymbol{\lambda}(B) > 0$. We may assume that $\sup B < b$ and hence $B + z \subset (a, b)$ for all $z \in (0, b - \sup B)$. By translation invariance of the Lebesgue measure, we have $\boldsymbol{\lambda}(B + z) > 0$ and therefore by assumption $\Pi_i(B + z) > 0$ for all $z \in (0, b - \sup B)$. From (4.23) it follows

$$\Pi^+_i(B) \geq \int_0^\infty U^-_{i,i}(\mathrm{d}z)\, \Pi_i(B + z) \geq \int_0^{b - \sup B} U^-_{i,i}(\mathrm{d}z)\, \Pi_i(B + z)$$

and since $U^-_{i,i}([0, b - \sup B)) > 0$ by increasing and right-continuous paths of $\boldsymbol{H}^-$, it follows $\Pi^+_i(B) > 0$, implying $\boldsymbol{\lambda}|_{(a,b)} \ll \Pi^+_i|_{(a,b)}$.

(ii) This is immediate from (4.24) in Theorem 4.30 and the same arguments as in part (i).

(iii) Since $\boldsymbol{J}$ is irreducible, it follows from the proof of the Wiener–Hopf factorization in Theorem 26 of [66] that $\boldsymbol{\Phi}^-$ is invertible and hence, for any $i, j \in \Theta$ we have

$$\int_0^\infty \mathrm{e}^{-\lambda y}\, U^-_{i,j}(\mathrm{d}y) = \big(\boldsymbol{\Phi}^-(\lambda)^{-1}\big)_{i,j}.$$

Thus, with Fubini and (4.23)

$$\begin{aligned}
&\int_1^\infty \mathrm{e}^{\lambda x}\, \Pi^+_i(\mathrm{d}x)\\
&= \int_0^\infty \int_1^\infty \mathrm{e}^{\lambda x}\, \Pi_i(y+\mathrm{d}x)\, U^-_{i,i}(\mathrm{d}y) + \sum_{k\neq i} \frac{\pi(k)}{\pi(i)} q_{k,i} \int_0^\infty \int_1^\infty \mathrm{e}^{\lambda x}\, F_{k,i}(y+\mathrm{d}x)\, U^-_{k,i}(\mathrm{d}y)\\
&= \int_0^\infty \int_{1+y}^\infty \mathrm{e}^{\lambda x}\, \Pi_i(\mathrm{d}x)\, \mathrm{e}^{-\lambda y} U^-_{i,i}(\mathrm{d}y) + \sum_{k\neq i} \frac{\pi(k)}{\pi(i)} q_{k,i} \int_0^\infty \int_{1+y}^\infty \mathrm{e}^{\lambda x}\, F_{k,i}(\mathrm{d}x)\, \mathrm{e}^{-\lambda y} U^-_{k,i}(\mathrm{d}y),\\
&\leq \int_1^\infty \mathrm{e}^{\lambda x}\, \Pi_i(\mathrm{d}x) \int_0^\infty \mathrm{e}^{-\lambda y} U^-_{i,i}(\mathrm{d}y) + \sum_{k\neq i} \frac{\pi(k)}{\pi(i)} q_{k,i} \int_1^\infty \mathrm{e}^{\lambda x}\, F_{k,i}(\mathrm{d}x) \int_0^\infty \mathrm{e}^{-\lambda y} U^-_{k,i}(\mathrm{d}y)\\
&= \int_1^\infty \mathrm{e}^{\lambda x}\, \Pi_i(\mathrm{d}x) \big(\boldsymbol{\Phi}^-(\lambda)^{-1}\big)_{i,i} + \sum_{k\neq i} \frac{\pi(k)}{\pi(i)} q_{k,i} \int_1^\infty \mathrm{e}^{\lambda x} F_{k,i}(\mathrm{d}x) \big(\boldsymbol{\Phi}^-(\lambda)^{-1}\big)_{k,i},
\end{aligned}$$

which is finite given the assumption.

(iv) Analogously to (iii).

(v) Under the assumption $\lim_{t\to\infty} \xi_t = \infty$ a.s., the ascending ladder height process $(\boldsymbol{H}^-, \boldsymbol{J}^-)$ of the dual of $(\boldsymbol{\xi}, \boldsymbol{J})$ is killed a.s. and hence for any $i, j \in \Theta$, $U^-_{i,j}$ is a finite measure. Thus, again by (4.23), (4.24) and a change of variables,

$$\int_1^\infty x^\lambda \Pi^+_i(\mathrm{d}x) \leq U^-_{i,i}(\mathbb{R}_+) \int_1^\infty x^\lambda\, \Pi_i(\mathrm{d}x) + \sum_{k\neq i} \frac{\pi(k)}{\pi(i)} q_{k,i} U^-_{k,i}(\mathbb{R}_+) \int_1^\infty x^\lambda F_{k,i}(\mathrm{d}x) < \infty$$

and

$$\begin{aligned}
q^+_{i,j} \int_1^\infty x^\lambda F^+_{i,j}(\mathrm{d}x) &\leq U^-_{j,i}(\mathbb{R}_+) \frac{\pi(j)}{\pi(i)} \int_1^\infty x^\lambda\, \Pi_j(\mathrm{d}x) + \sum_{k\neq j} \frac{\pi(k)}{\pi(i)} q_{k,j} U^-_{k,i}(\mathbb{R}_+) \int_1^\infty x^\lambda F_{k,j}(\mathrm{d}x)\\
&< \infty.
\end{aligned}$$

■

Remark 4.35. (i) Conditions (4.30) are sufficient but not necessary conditions for exponential moments of the components of the Lévy system $\boldsymbol{\Pi}^+$, since $U^-_{k,i}$ is trivial for some $k \neq i$ whenever $\boldsymbol{J}^+$ is not irreducible under $(\widehat{\mathbb{P}}^{0,i})_{i\in\Theta}$. It is however true that if $\mathbb{E}[\exp(\lambda H_1^{+,(i)})] < \infty$, we must necessarily have $\int_1^\infty \mathrm{e}^{\lambda x}\, \Pi_i(\mathrm{d}x) < \infty$ and if $\mathbb{E}[\exp(\lambda \Delta^+_{i,j})] < \infty$, it must hold $\int_1^\infty \mathrm{e}^{\lambda x}\, F_{i,j}(\mathrm{d}x) < \infty$, since the on-diagonal potential measures $U^-_{i,i}$ are non-trivial.

(ii) We restrict to the case $\lim_{t\to\infty}\xi_t = \infty$ a.s. in (v). The oscillatory case $\limsup_{t\to\infty}\xi_t = -\liminf_{t\to\infty}\xi_t = \infty$ a.s. is more difficult to handle since in this case $(\boldsymbol{H}^+, \boldsymbol{J}^+)$ is unkilled under the dual measures $\widehat{\mathbb{P}}^{0,i}$ and we have no control over $\boldsymbol{U}^-$ solely in terms of the characteristics of $(\boldsymbol{\xi}, \boldsymbol{J})$. In [66] the authors establish the necessary and sufficient integral criterion given in (4.13) for finiteness of the first moment of H_1^+ in the oscillatory regime by taking a detour via random walk theory, building on the strategy for the related problem for Lévy processes in [70]. Taking into account Theorem 1 of [48], such an ansatz, even though out of scope of this chapter, is a possible strategy to tackle the problem at hand in our case as well.

4.5 Application to real self-similar Markov processes

In this section we show how to apply our results on the exponential mixing behavior of Markov additive processes sampled at first hitting times to the class of α-self-similar Markov processes and in particular strictly α-stable Lévy processes. Even in the case of α-stable processes the application is non-trivial since such Lévy processes do not satisfy the fundamental assumption of finite mean of the associated ascending ladder height Lévy process, since in fact the ascending ladder height process is an α-stable subordinator with $\alpha \in (0, 1)$ and thus does not have a finite first moment. Because of non-ergodicity of the associated overshoots, we can therefore not expect a strong mixing behavior of the stable process sampled at first hitting times. However, making use of the *Lamperti–Kiu transform* for real self-similar Markov processes, we can give bounds on the β-mixing coefficient of the σ-algebras generated by the past and the future of α-self-similar process sampled at first hitting times given appropriate properties of the associated MAP. By considering the Lamperti-stable MAP and its explicit characterization found in [45], we are thus able to bound the β-mixing coefficient of the above σ-algebras for transient α-stable processes. To this end, let us first recall the precise definitions of real α-self-similar Markov processes and α-stable Lévy processes and give a brief overview on the Lamperti–Kiu transform and its implications.

We say that a real-valued Feller process $(\Omega, \mathcal{G}, (\mathcal{G}_t)_{t\geq 0}, (Z_t)_{t\geq 0}, (\mathbf{P}^x)_{x\in\mathbb{R}})$ is an α-self-similar Markov process, if it satisfies the *scaling property* that for any $c > 0$,

$$\{\boldsymbol{Z}, \mathbf{P}^{cx}\} \overset{\mathrm{d}}{=} \left\{\left(cZ_{c^{-\alpha}t}\right)_{t\geq 0}, \mathbf{P}^x\right\}, \quad x \in \mathbb{R}. \tag{4.31}$$

An (unkilled) Lévy process $\boldsymbol{X} = (X_t)_{t\geq 0}$ with associated family of probability measures $(\mathbf{P}^x)_{x\in\mathbb{R}}$ is a strictly α-stable Lévy process (or simply stable Lévy process for short if there is no room for confusion) for $\alpha \in (0, 2]$ if it satisfies (4.31). The case $\alpha = 2$ boils down to Brownian motion, which we exclude from here-on. Since Lévy processes are Feller, α-stable Lévy processes are therefore particular representatives of α-self-similar Markov processes.

Taking the perspective commonly encountered in the literature to parametrize the stable process through its index of self-similarity α and the positivity parameter $\rho := \mathbf{P}^0(X_t \geq 0)$, the Lévy measure Π of $\boldsymbol{X}$ is absolutely continuous with density π satisfying

$$\pi(x) = c_+ x^{-(\alpha+1)} \mathbb{1}_{(0,\infty)}(x) + c_- |x|^{-(\alpha+1)} \mathbb{1}_{(-\infty,0)}(x), \quad x \in \mathbb{R},$$

where

$$c_+ = \Gamma(\alpha+1)\frac{\sin(\pi\alpha\rho)}{\pi}, \quad \text{and} \quad c_- = \Gamma(\alpha+1)\frac{\sin(\pi\alpha(1-\rho))}{\pi}.$$

The Lévy–Khintchine exponent Ψ is given by

$$\Psi(\theta) = c|\theta|^{\alpha}\big(1 - \mathrm{i}\beta \tan \tfrac{\pi\alpha}{2}\operatorname{sgn}(\theta)\big), \quad \theta \in \mathbb{R},$$

where $\beta = (c_+ - c_-)/(c_+ + c_-)$ and our specific parametrization forces $c = \cos(\pi\alpha(\rho - 1/2))$. For all of the above statements we refer to Kyprianou [108].

We now come to the one-to-one correspondence between self-similar Markov processes on $\mathbb{R}$ and Markov additive processes on $\mathbb{R} \times \{-1, 1\}$ expressed through the Lamperti–Kiu transform, which is investigated in [104] and [45] for the real valued setting, and, more generally for arbitrary state spaces, in [5]. If we let $\boldsymbol{Z}$ be an α-self-similar Markov process on $\mathbb{R}$ absorbed at 0 with lifetime $\tau_0 = \inf\{t > 0 : X_t = 0\}$ and define $\mathbb{P}^{x,i} = \mathbf{P}^{i\mathrm{e}^x}$ for $(x, i) \in \mathbb{R} \times \{-1, 1\}$ and $\mathbb{P}^{-\infty,\varpi} = \mathbf{P}^0$, then the process $(\boldsymbol{\xi}, \boldsymbol{J})$ defined by

$$\begin{cases} \xi_t = \log|Z_{\tau(t)}| \text{ and } J_t = \operatorname{sgn}(Z_{\tau(t)}), & \text{if } t < \int_0^{\tau_0} |Z_s|^{-\alpha}\, \mathrm{d}s, \\ (\xi_t, J_t) = \vartheta =: (-\infty, \varpi), & \text{if } t \geq \int_0^{\tau_0} |Z_s|^{-\alpha}\, \mathrm{d}s, \end{cases}$$

where $t \mapsto \tau(t)$ is the time change given by the right-continuous inverse

$$\tau(t) := \inf\{s \geq 0 : \int_0^s |Z_u|^{-\alpha}\, \mathrm{d}u > t\},$$

of the continuous additive functional $(A_t)_{t\geq 0}$ of $\boldsymbol{Z}$, given by

$$A_t := \int_0^{t\wedge\tau_0} |Z_s|^{-\alpha}\, \mathrm{d}s, \quad t \geq 0,$$

and ϖ is some isolated state, then $((\boldsymbol{\xi}, \boldsymbol{J}), (\mathbb{P}^x)_{x\in(\mathbb{R}\times\{-1,1\})_\vartheta})$ is a MAP on $\mathbb{R} \times \{-1, 1\}$ with lifetime $\zeta = \int_0^{\tau_0} |Z_s|^{-\alpha}\, \mathrm{d}s$ and underlying filtration $(\mathcal{F}_t = \mathcal{G}_{\tau(t)})_{t\geq 0}$. Moreover, we have the following trichotomy characterizing the long-time behavior of the MAPs ordinator in terms of the hitting properties of $\boldsymbol{Z}$ at 0 (one can indeed verify that self-similarity of $\boldsymbol{Z}$ guarantees that these are the only possible cases):

(a) if $\mathbf{P}^x(\tau_0 < \infty) = 0$ for any $x \neq 0$, then $\lim_{t\to\infty} \xi_t = \infty$ almost surely;

(b) if $\mathbf{P}^x(\tau_0 < \infty, Z_{\tau_0-} = 0) = 1$ for any $x \neq 0$, then $\lim_{t\to\infty} \xi_t = -\infty$ almost surely;

(c) if $\mathbf{P}^x(\tau_0 < \infty, Z_{\tau_0-} \neq 0) = 1$ for any $x \neq 0$, then the MAP is almost surely killed and its lifetime ζ is exponentially distributed with a rate not depending on its initial distribution.

Conversely, for a given MAP $(\boldsymbol{\xi}, \boldsymbol{J})$ with lifetime ζ,

$$Z_t = J_{\sigma(t)} \mathrm{e}^{\xi_{\sigma(t)}} \mathbb{1}_{\left\{t < \int_0^{\zeta} \mathrm{e}^{\alpha\xi_s}\, \mathrm{d}s\right\}}, \quad t \geq 0,$$

where

$$\sigma(t) = \inf\left\{s \geq 0 : \int_0^s \mathrm{e}^{\alpha\xi_u}\, \mathrm{d}u > t\right\}, \quad t \geq 0,$$

defines an α-self-similar Markov process absorbed in 0 with lifetime $\tau_0 = \int_0^{\zeta} \mathrm{e}^{\alpha\xi_s}\, \mathrm{d}s$. This is however not the direction we are interested in and we refer the reader to the relevant literature cited above for details. Note also that in case of $\boldsymbol{Z}$ being strictly positive almost surely up to its lifetime, the Lamperti–Kiu transform boils down to the Lamperti transform for positive self-similar Markov processes and the associated MAP can be projected onto a killed Lévy process.

With the help of the Lamperti–Kiu transform we obtain the following result on the β-mixing coefficient of the σ-algebras generated by α-self-similar Markov processes sampled at past and future hitting times. While the Lamperti-stable MAP is exponentially β-mixing under the given assumptions, the β-mixing coefficient for the α-self similar Markov process sampled at first hitting times shows non-homogeneous decay with almost square root rate as a result of the logarithm present in the Lamperti–Kiu transform.

PROPOSITION 4.36. *Suppose that $\boldsymbol{Z}$ is α-self-similar such that $\mathbf{P}^x(\tau_0 < \infty) = 0$ for all $x \neq 0$ and moreover its associated Lamperti–Kiu MAP satisfies the assumptions from Theorem 4.22.(i). If η is some distribution on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ without atom at 0 such that*

$$\int_{\mathbb{R}} |x|^\lambda \,\eta(\mathrm{d}x) < \infty,$$

for some $\lambda > 0$, then for any $\delta \in (0,1)$ there exists a constant $C(\lambda, \eta, \delta) > 0$ such that for any $t \geq 1$ we have

$$\beta_{\mathbf{P}^\eta}(\mathcal{N}_t, \overline{\mathcal{N}}_{t+s}) \leq C(\lambda, \eta)\Big(\frac{t+s}{t}\Big)^{-1/(2+\delta)}, \quad s > 0,$$

where we denoted

$$\mathcal{N}_t = \sigma\big(Z_{T_s^Z}, s \leq t\big), \quad \overline{\mathcal{N}}_t = \sigma\big(Z_{T_s^Z}, s \geq t\big).$$

Proof. First, observe that $\boldsymbol{Z}$ not hitting 0 almost surely when started away from the origin implies that the time change $(\tau(t))_{t\geq 0}$ is strictly increasing and continuous almost surely. Thus, the overshoot process of $(\log|Z_t|, \operatorname{sgn}(Z_t))_{t\geq 0}$ is indistinguishable from the overshoot process of the associated Lamperti-MAP $(\boldsymbol{\xi}, \boldsymbol{J})$. Moreover, the mapping

$$\phi\colon \mathbb{R} \times \{-1, 1\} \longrightarrow \mathbb{R} \setminus \{0\}, \quad (x, i) \mapsto i\mathrm{e}^x,$$

is a homeomorphism and $Z_t = \phi(\log|Z_t|, \operatorname{sgn}(Z_t))$ for all $t \geq 0$ on the set $\Lambda = \{\omega \in \Omega : Z_t(\omega) \neq 0 \text{ for all } t \geq 0\}$, which is of $\mathbf{P}^\mu$-measure 1 for any distribution μ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ not having an atom at 0. It follows for any $t \geq 1$ with the notation from Corollary 4.27 that there exists some $\mathbf{P}^\mu$-nullset N_t^μ such that

$$\mathcal{N}_t \vee N_t^\mu = \big((\xi_{T_s}, J_{T_s}), s \leq \log t\big) \vee N_t^\mu = \mathcal{K}_{\log(t)} \vee N_t^\mu,$$

and

$$\overline{\mathcal{N}}_t \vee N_t^\mu = \big((\xi_{T_s}, J_{T_s}), s \geq \log t\big) \vee N_t^\mu = \overline{\mathcal{K}}_{\log(t)} \vee N_t^\mu,$$

where for two σ-algebras $\mathcal{A}, \mathcal{B}$ we write $\mathcal{A} \vee \mathcal{B} = \sigma(\mathcal{A} \cup \mathcal{B})$. Here we used the definition of the Lamperti–Kiu transform and the fact that for any $t \geq 1$ we have $T_t^{|Z|} = T_{\log t}$. Since moreover $\mathbf{P}^\eta = \mathbb{P}^{\eta\circ\phi}$ and by assumption

$$\int_{\mathbb{R}\times\{-1,1\}} \mathrm{e}^{\lambda x}\, \eta\circ\phi(\mathrm{d}x, \mathrm{d}i) = \int_{\mathbb{R}\setminus\{0\}} \mathrm{e}^{\lambda \log|z|}\, \eta(\mathrm{d}z) = \int_{\mathbb{R}\setminus\{0\}} |z|^\lambda\, \eta(\mathrm{d}z) < \infty,$$

it follows from Corollary 4.27 and the assumptions on the Lamperti-MAP $(\boldsymbol{\xi}, \boldsymbol{J})$ that for any $\delta \in (0,1)$ there exists $C(\lambda, \eta, \delta) > 0$ such that for any $t \geq 1$ and $s > 0$

$$\beta_{\mathbf{P}^\eta}(\mathcal{N}_t, \overline{\mathcal{N}}_{t+s}) = \beta_{\mathbb{P}^{\eta\circ\phi}}(\mathcal{K}_{\log t}, \overline{\mathcal{K}}_{\log(t+s)}) \leq C(\lambda, \eta, \delta)\mathrm{e}^{-(\log(t+s)-\log t)/(2+\delta)}$$

$$= C(\lambda, \eta, \delta)\left(\frac{t+s}{t}\right)^{-1/(2+\delta)},$$

as claimed. Note here that the nullsets N_t^η and N_{t+s}^η from above have no influence on the β-mixing coefficient by its definition. ∎

Consider now a scalar α-stable process $(X_t^0)_{t\geq 0}$ absorbed upon hitting of the origin, i.e. for $\tau_0 = \inf\{s \geq 0 : X_s = 0\}$,

$$X_t^0 = X_t \mathbb{1}_{[0,\tau_0)}(t), \quad t \geq 0.$$

We show that X^0 satisfies the assumptions from Proposition 4.36 that yield β-mixing of overshoots of the corresponding MAP $(\boldsymbol{\xi}, \boldsymbol{J})$ obtained through the Lamperti–Kiu transform, which we henceforth will refer to as the *Lamperti-stable MAP*.

Since the assumptions are couched in form of the ascending ladder height process $(\boldsymbol{H}^+, \boldsymbol{J}^+)$, one direct approach would be to make use of the *deep factorization* of $\boldsymbol{X}^0$ given in [108], where the MAP exponents $\boldsymbol{\Phi}^+$ and $\boldsymbol{\Phi}^-$ of the ascending ladder height processes of $(\boldsymbol{\xi}, \boldsymbol{J})$ and its dual were explicitly calculated. However, for the sake of exposition, we go another route by making use of the results based on Vigon's équations amicales inversés from Section 4.4 to infer the needed properties of $(\boldsymbol{H}^+, \boldsymbol{J}^+)$ from those of $(\boldsymbol{\xi}, \boldsymbol{J})$. The characteristics of the latter were calculated in Theorem 10 and Corollary 11 of Chaumont et al. [45], giving $\sigma_{\pm 1} = 0$, i.e. the underlying Lévy processes have no Gaussian component,

$$\Pi_{\pm 1}(\mathrm{d}x) = \mathrm{e}^x \pi(\pm(\mathrm{e}^x - 1))\,\mathrm{d}x, \quad x \in \mathbb{R},$$
$$F_{\pm 1,\mp 1}(\mathrm{d}x) = \frac{\alpha \mathrm{e}^x}{(1+\mathrm{e}^x)^{\alpha+1}}\,\mathrm{d}x, \quad x \in \mathbb{R},$$

and

$$q_{\pm 1,\mp 1} = \frac{c_\mp}{\alpha}.$$

If we assume that $\boldsymbol{X}$ does not have one-sided jumps, then $c_\pm > 0$ and hence $\boldsymbol{J}$ is irreducible. Since $\Pi_{\pm 1}$ has a strictly positive Lebesgue density on $(0,\infty)$ it follows by Lemma 4.34 that $\boldsymbol{\lambda}|_{(0,\infty)} \ll \Pi_{\pm 1}^+|_{(0,\infty)}$ as well. Further, we have for $\lambda > 0$

$$\int_1^\infty \mathrm{e}^{\lambda x}\,\Pi_1(\mathrm{d}x) = c_+ \int_1^\infty \mathrm{e}^{(\lambda+1)x}(\mathrm{e}^x - 1)^{-(\alpha+1)}\,\mathrm{d}x,$$

and hence

$$\int_1^\infty \mathrm{e}^{\lambda x}\,\Pi_1(\mathrm{d}x) < \infty \iff \lambda \in (0,\alpha).$$

Similarly, we obtain

$$\int_1^\infty \mathrm{e}^{\lambda x}\,\Pi_{-1}(\mathrm{d}x) < \infty \iff \lambda \in (0,\alpha),$$

and hence $\mathbb{E}[\exp(\lambda \xi_1^{(\pm 1)})] < \infty$ iff $\lambda \in (0,\alpha)$. Moreover,

$$\int_{\mathbb{R}} \mathrm{e}^{\lambda x}\,F_{\pm 1}(\mathrm{d}x) = \alpha \int_{\mathbb{R}} \mathrm{e}^{(\lambda+1)x}(1+\mathrm{e}^x)^{-(\alpha+1)}\,\mathrm{d}x < \infty \iff \lambda \in (0,\alpha).$$

Again by Lemma 4.34 we conclude that $H_1^{+,(\pm 1)}$ and $\Delta^+_{\pm 1,\mp 1}$ all possess an exponential λ-moment whenever $\lambda \in (0,\alpha)$.

Recall now that $\boldsymbol{X}$ does not hit 0 if and only if $\alpha \in (0,1)$ and hence the ordinator $\boldsymbol{\xi}$ of the Lamperti-stable MAP satisfies $\limsup_{t\to\infty} \xi_t = \infty$ almost surely if and only if $\alpha \in (0,1)$. Since our asymptotic approach on overshoots of MAPs requires this property, we will restrict to this case and can therefore identify $\boldsymbol{X} = \boldsymbol{X}^0$ almost surely. All that remains to show for exponential β-mixing of the Lamperti-stable MAP is now upward regularity and irreducibility of $\boldsymbol{J}^+$. Irreducibility of $\boldsymbol{J}^+$ is a direct consequence of Proposition 4.6 since $\boldsymbol{J}$ is irreducible and the support of $\Pi_\pm$ is unbounded. To verify upward regularity, we observe that by Theorem 1 in Kuznetsov and Pardo [107], $\boldsymbol{\xi}^{(1)}$ killed at an independent exponential time with rate c_-/α belongs to the class of *hypergeometric Lévy processes* with parameters $(1-\alpha(1-\rho), \alpha\rho, (1-\alpha)(1-\rho), \alpha(1-\rho))$. The ascending ladder height process $\boldsymbol{H}$ of such a hypergeometric Lévy process is a β-subordinator with parameters $(\alpha(1-\rho), \alpha(1-\rho), 1-\alpha\rho)$, whose Lévy measure is given by

$$\Pi_H(\mathrm{d}x) = \frac{1-\alpha\rho}{\Gamma(\alpha\rho)}(1-\mathrm{e}^{-x})^{\alpha\rho-2}\mathrm{e}^{-(1+\alpha(1-2\rho))x}\,\mathrm{d}x, \quad x > 0.$$

Clearly, $\Pi_H((0,1)) = \infty$ and hence $\boldsymbol{H}$ is not compound Poisson, which shows that the associated hypergeometric Lévy process is upward regular. Since killing has no influence on upward regularity, this now shows that $\boldsymbol{\xi}^{(1)}$ is indeed upward regular. Upward regularity of $\boldsymbol{\xi}^{(-1)}$ can be argued analogously once we observe that $\boldsymbol{\xi}^{(-1)}$ killed at rate c_+/α is the hypergeometric process obtained from killing the dual process $\widehat{\boldsymbol{X}}$ of $\boldsymbol{X}$ upon entering $(-\infty, 0]$. Hence, with the ergodic analysis of overshoots from Section 4.3 and Proposition 4.36, we have proved the following.

Proposition 4.37. *Let $\alpha \in (0,1)$ and $\boldsymbol{X}$ be strictly α-stable. Then the overshoot process of the Lamperti-stable MAP associated to $\boldsymbol{X}$ is $\mathcal{R}_\lambda V_\lambda$-uniformly ergodic and for any starting distribution μ such that $\mu(\cdot, \{-1,1\})$ has an exponential λ-moment for some $\lambda \in (0,\alpha)$, the overshoot process is exponentially β-mixing. Moreover, for any distribution η on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ without atom at 0 such that for some $\lambda \in (0,\alpha)$,*

$$\int_{\mathbb{R}} |x|^\lambda \,\eta(\mathrm{d}x) < \infty,$$

there exists a constant $C(\lambda,\eta,\delta) > 0$ for any $\delta \in (0,1)$ such that for any $t \geq 1$ we have

$$\beta_{\mathbf{P}^\eta}(\mathcal{N}_t, \overline{\mathcal{N}}_{t+s}) \leq C(\lambda,\eta,\delta)\Big(\frac{t+s}{t}\Big)^{-1/(2+\delta)}, \quad s > 0,$$

where we denoted

$$\mathcal{N}_t = \sigma\big(X_{T_s^X}, s \leq t\big), \quad \overline{\mathcal{N}}_t = \sigma\big(X_{T_s^X}, s \geq t\big).$$

Corollary 4.38. *Let $\alpha \in (0,1)$ and $\boldsymbol{X}$ be strictly α-stable. Then, for any $x > 0$ and $\delta \in (0,1)$ there exists a constant $C(x,\delta)$ sich that for any $t \geq 1$,*

$$\beta_{\mathbf{P}^0}(\mathcal{N}_t, \overline{\mathcal{N}}_{t+s}) \leq C(x,\delta)\Big(\frac{t+x+s}{t+x}\Big)^{-1/(2+\delta)}, \quad s > 0.$$

Proof. Fix $x > 0$ and choose some $\lambda \in (0,\alpha)$. By spatial homogeneity of $\boldsymbol{X}$ we have

$$\Big\{\Big(X_{T_{t+x}^X}, t \geq 0\Big), \mathbf{P}^x\Big\} \stackrel{\mathrm{d}}{=} \Big\{\Big(X_{T_t^X} + x, t \geq 0\Big), \mathbf{P}^0\Big\}$$

and therefore, using Proposition 4.37

$$\beta_{\mathbf{P}^0}(\mathcal{N}_t, \overline{\mathcal{N}}_{t+s}) = \beta_{\mathbf{P}^x}(\mathcal{N}_{t+x}, \overline{\mathcal{N}}_{t+x+s}) \leq C(x,\delta)\Big(\frac{t+x+s}{t+x}\Big)^{-1/(2+\delta)},$$

where $C(x) := C(\lambda, \delta_x, \delta)$. ∎

4.A Proof of the resolvent formula

Proof of Theorem 4.7. Note first that by assumed irreducibility of $\boldsymbol{J}^+$, it follows as a consequence of the Perron–Frobenius theorem that $\boldsymbol{\Phi}^+(\lambda)$ is invertible for any $\lambda > 0$, see Corollary 2.4 in Stephenson [157] or Remark 2.2 in Ivanovs et al. [96], and hence the statement of the theorem makes formally sense. Fix $(x,i) \in \mathbb{R}_+ \times \Theta$. Let $\tau_0 := \inf\{t \geq 0 : \mathcal{O}_t = 0\}$, which is clearly finite and a stopping time for $(\mathcal{O}_t, \mathcal{J}_t)$ since the process is Feller by Proposition 4.4. By the sawtooth structure of $(\boldsymbol{\mathcal{O}}, \boldsymbol{\mathcal{J}})$, see also Figure 4.1 for an illustration, we have $\tau_0 = x$ and $(\mathcal{O}_t, \mathcal{J}_t) = (x-t, i)$ for $t \in [0,x]$, $\mathbb{P}^{x,i}$-a.s.. Together with the strong Markov property of $(\boldsymbol{\mathcal{O}}, \boldsymbol{\mathcal{J}})$, we therefore obtain for $f \in \mathcal{B}_b(\mathbb{R}_+ \times \Theta)$

$$\mathcal{U}_\lambda f(x,i) = Q_\lambda f(x,i) + \mathrm{e}^{-\lambda x}\mathcal{U}_\lambda f(0,i).$$

Hence, we only need to calculate $\mathcal{U}_\lambda f(0,i)$.

We start with the case that the Lévy measures Π_i^+, $i \in \Theta$ are finite and then proceed by an approximation argument to the general case. Our assumption of upward regularity of $(\boldsymbol{\xi}, \boldsymbol{J})$ then forces $d_i^+ > 0$ for all $i \in \Theta$, that is the processes $\boldsymbol{H}^{+,(i)}$ are compound Poisson processes with drift. Denote for $i \in \Theta$ by $Y^{(i)}$ random variables independent of $(\boldsymbol{\xi}, \boldsymbol{J})$ corresponding to the jumps of $\boldsymbol{H}^{+,(i)}$, whose distribution is given by $\Pi_i^+(\mathrm{d}x)/\Pi_i^+(\mathbb{R}_+)$. Moreover, denote by $\sigma := \inf\{t \geq 0 : J_t^+ \neq J_0^+\}$ the first jump time of $\boldsymbol{J}^+$ and by $\tau = \inf\{t \geq 0 : \Delta H_t^{+,0,J_0^+} > 0\}$ the first jump time of the Lévy process driving the ascending ladder height process before the first phase transition. Then, from Proposition 4.1 and indistinguishability of $(\boldsymbol{\mathcal{O}}^+, \boldsymbol{\mathcal{J}}^+)$ and $(\boldsymbol{\mathcal{O}}, \boldsymbol{\mathcal{J}})$ we can infer that under $\mathbb{P}^{0,i}$, it holds that $T := \inf\{t \geq 0 : \Delta(\mathcal{O}_t, \mathcal{J}_t) \neq 0\} = H^{+,0,i}_{(\sigma\wedge\tau)-} = d_i^+(\sigma \wedge \tau)$ almost surely (consult again Figure 4.1 for an illustration).

$$\mathcal{U}_\lambda f(0,i) = \mathbb{E}^{0,i}\Big[\int_0^T + \int_T^{T+\tau_0\circ\theta_T} + \int_{T+\tau_0\circ\theta_T}^\infty \mathrm{e}^{-\lambda t} f(\mathcal{O}_t, \mathcal{J}_t)\,\mathrm{d}t\Big] =: I_1 + I_2 + I_3, \tag{4.32}$$

where $(\theta_t)_{t\geq 0}$ denotes the transition operator of $(\boldsymbol{\mathcal{O}}, \boldsymbol{\mathcal{J}})$. Since under $\mathbb{P}^{0,i}$, $\tau \overset{\mathrm{d}}{=} \mathrm{Exp}(\Pi_i^+(\mathbb{R}_+))$ is independent of $\sigma \overset{\mathrm{d}}{=} \mathrm{Exp}(-q_{i,i}^+)$ by Proposition 4.1, it follows that $T \overset{\mathrm{d}}{=} \mathrm{Exp}((\Pi_i^+(\mathbb{R}_+) - q_{i,i}^+)/d_i^+)$ and hence

$$\begin{aligned} I_1 = \mathbb{E}^{0,i}\Big[\int_0^T \mathrm{e}^{-\lambda t} f(0,i)\,\mathrm{d}t\Big] = f(0,i)\frac{1}{\lambda}\big(1 - \mathbb{E}^{0,i}[\mathrm{e}^{-\lambda T}]\big) &= f(0,i)\frac{1}{\lambda}\Big(1 - \frac{\Pi_i^+(\mathbb{R}_+) - q_{i,i}^+}{d_i^+\lambda + \Pi_i^+(\mathbb{R}_+) - q_{i,i}^+}\Big) \\ &= f(0,i)\frac{d_i^+}{d_i^+\lambda + \Pi_i^+(\mathbb{R}_+) - q_{i,i}^+}. \end{aligned}$$

For the second integral, we use that $\mathbb{P}^{0,i}(J_\sigma^+ = j) = -q_{i,j}^+/q_{i,i}^+$, independence of σ, J_σ^+ and $Y^{(i)}$ in combination with Proposition 4.1 and the strong Markov property to obtain

$$\begin{aligned} I_2 &= \mathbb{E}^{0,i}\Big[\mathrm{e}^{-\lambda T}\mathbb{E}^{0,i}\Big[\int_0^{\tau_0} \mathrm{e}^{-\lambda t} f(\mathcal{O}_t, \mathcal{J}_t)\,\mathrm{d}t \circ \theta_T \Big| \mathcal{G}_T\Big]\Big] \\ &= \mathbb{E}^{0,i}\Big[\mathrm{e}^{-\lambda T}\mathbb{E}^{\mathcal{O}_T,\mathcal{J}_T}\Big[\int_0^{\tau_0} \mathrm{e}^{-\lambda t} f(\mathcal{O}_t, \mathcal{J}_t)\,\mathrm{d}t\Big]\Big] \\ &= \mathbb{E}^{0,i}\big[\mathrm{e}^{-\lambda d_i^+\tau} Q_\lambda(Y^{(i)}, i)\,;\, \tau < \sigma\big] + \mathbb{E}^{0,i}\big[\mathrm{e}^{-\lambda d_i^+\sigma} Q_\lambda\big(\Delta^{+,1}_{i,J_\sigma^+}, J_\sigma^+\big)\,;\, \sigma < \tau\big] \\ &= \mathbb{E}^{0,i}[\mathrm{e}^{-\lambda d_i^+\tau}\,;\, \tau < \sigma]\,\mathbb{E}^{0,i}[Q_\lambda(Y^{(i)}, i)] + \mathbb{E}^{0,i}\big[\mathrm{e}^{-\lambda d_i^+\sigma}\,;\, \sigma < \tau\big]\,\mathbb{E}^{0,i}\big[Q_\lambda\big(\Delta^{+,1}_{i,J_\sigma^+}, J_\sigma^+\big)\big] \end{aligned}$$

$$
\begin{aligned}
&= \frac{\Pi_i^+(\mathbb{R}_+)}{\lambda d_i^+ + \Pi_i^+(\mathbb{R}_+) - q_{i,i}^+} \int_0^\infty Q_\lambda f(y,i)\, \Pi_i^+(\mathrm{d}y)/\Pi_i^+(\mathbb{R}_+) \\
&\quad + \frac{-q_{i,i}^+}{\lambda d_i^+ + \Pi_i^+(\mathbb{R}_+) - q_{i,i}^+} \sum_{j\neq i} \frac{q_{i,j}^+}{-q_{i,i}^+} \int_0^\infty Q_\lambda f(y,j)\, F_{i,j}^+(\mathrm{d}y) \\
&= \frac{1}{\lambda d_i^+ + \Pi_i^+(\mathbb{R}_+) - q_{i,i}^+} \Big(\int_0^\infty Q_\lambda f(y,i)\, \Pi_i^+(\mathrm{d}y) + \sum_{j\neq i} q_{i,j}^+ \int_0^\infty Q_\lambda f(y,j)\, F_{i,j}^+(\mathrm{d}y) \Big)
\end{aligned}
$$

With the same arguments as above we also obtain

$$
\begin{aligned}
I_3 &= \mathbb{E}^{0,i}\Big[\mathrm{e}^{-\lambda T} \mathbb{E}^{0,i}\Big[\int_{\tau_0}^\infty \mathrm{e}^{-\lambda t} f(\mathcal{O}_t, \mathcal{J}_t)\, \mathrm{d}t \Big| \mathcal{G}_T \Big]\Big] \\
&= \mathbb{E}^{0,i}\Big[\mathrm{e}^{-\lambda T} \mathbb{E}^{\mathcal{O}_T, \mathcal{J}_T}\Big[\int_{\tau_0}^\infty \mathrm{e}^{-\lambda t} f(\mathcal{O}_t, \mathcal{J}_t)\, \mathrm{d}t \Big]\Big] \\
&= \mathbb{E}^{0,i}[\mathrm{e}^{-\lambda d_i^+ \tau}; \tau < \sigma]\, \mathbb{E}^{0,i}\Big[\mathbb{E}^{y,i}\Big[\mathrm{e}^{-\lambda \tau_0} \mathbb{E}^{y,i}\Big[\int_0^\infty \mathrm{e}^{-\lambda t} f(\mathcal{O}_t, \mathcal{J}_t)\, \mathrm{d}t \circ \theta_{\tau_0} \Big| \mathcal{G}_{\tau_0} \Big]\Big]\Big|_{y=Y^{(i)}} \Big] \\
&\quad + \mathbb{E}^{0,i}[\mathrm{e}^{-\lambda d_i^+ \sigma}; \sigma < \tau] \sum_{j\neq i} \frac{q_{i,j}^+}{-q_{i,i}^+} \mathbb{E}^{0,i}\Big[\mathbb{E}^{y,j}\Big[\mathrm{e}^{-\lambda \tau_0} \mathbb{E}^{y,j}\Big[\int_0^\infty \mathrm{e}^{-\lambda t} f(\mathcal{O}_t, \mathcal{J}_t)\, \mathrm{d}t \circ \theta_{\tau_0} \Big| \mathcal{G}_{\tau_0} \Big]\Big]\Big|_{y=\Delta_{i,j}^{+,1}} \Big] \\
&= \frac{\Pi_i^+(\mathbb{R}_+)}{\lambda d_i^+ + \Pi_i^+(\mathbb{R}_+) - q_{i,i}^+} \mathcal{U}_\lambda f(0,i) \mathbb{E}^{0,i}\big[\mathrm{e}^{-\lambda Y^{(i)}}\big] \\
&\quad + \frac{1}{\lambda d_i^+ + \Pi_i^+(\mathbb{R}_+) - q_{i,i}^+} \sum_{j\neq i} q_{i,j}^+ \mathcal{U}_\lambda f(0,j) \mathbb{E}^{0,i}\big[\mathrm{e}^{-\lambda \Delta_{i,j}^{+,1}}\big] \\
&= \frac{1}{\lambda d_i^+ + \Pi_i^+(\mathbb{R}_+) - q_{i,i}^+} \Big(\mathcal{U}_\lambda f(0,i) \int_0^\infty \mathrm{e}^{-\lambda y}\, \Pi_i^+(\mathrm{d}y) + \sum_{j\neq i} q_{i,j}^+ \mathcal{U}_\lambda f(0,j) \int_0^\infty \mathrm{e}^{-\lambda y}\, F_{i,j}^+(\mathrm{d}y) \Big).
\end{aligned}
$$

Plugging into (4.32), using $G_{ij}^+(\lambda) = \int_0^\infty \exp(-\lambda y)\, F_{ij}^+(\mathrm{d}y)$ and rearranging now yields

$$
\begin{aligned}
&\mathcal{U}_\lambda f(0,i) \Big(d_i^+ \lambda + \int_0^\infty (1 - \mathrm{e}^{-\lambda y})\, \Pi_i^+(\mathrm{d}y) - q_{i,i}^+ \Big) - \sum_{j\neq i} q_{i,j}^+ G_{ij}^+(\lambda) \mathcal{U}_\lambda f(0,j) \\
&\qquad = d_i^+ f(0,i) + \int_0^\infty Q_\lambda f(y,i)\, \Pi_i^+(\mathrm{d}y) + \sum_{j\neq i} q_{i,j}^+ \int_0^\infty Q_\lambda f(y,i)\, F_{i,j}^+(\mathrm{d}y).
\end{aligned}
$$

By (4.4) the left hand side is equal to

$$
\mathcal{U}_\lambda f(0,i)(\Phi_i^+(\lambda) - q_{i,i}^+) - \sum_{j\neq i} q_{i,j}^+ G_{ij}^+(\lambda) \mathcal{U}_\lambda f(0,j) = \big(\boldsymbol{\Phi}^+(\lambda) \cdot (\mathcal{U}_\lambda f(0,j))_{j=1,\dots,n}^\top \big)_i
$$

and hence we conclude that

$$
(\mathcal{U}_\lambda f(0,i))_{i=1,\dots,n}^\top = \boldsymbol{\Phi}^+(\lambda)^{-1} \cdot \Big(d_i^+ f(0,i) + \int_0^\infty Q_\lambda f(x,i)\, \Pi_i^+(\mathrm{d}x) + \sum_{j\neq i} q_{i,j}^+ \mathbb{E}[Q_\lambda f(\Delta_{i,j}^+, j)] \Big)_{i=1,\dots,n}^\top, \tag{4.33}
$$

which proves the assertion in case that $(\boldsymbol{H}^+, \boldsymbol{J}^+)$ is a compound Poisson Markov additive subordinator. For the general case, suppose that $(\boldsymbol{\xi}, \boldsymbol{J})$ is an upward regular MAP and let for $\varepsilon > 0$,

$({}^{\varepsilon}\boldsymbol{H}^+, \boldsymbol{J}^+)$ be the ascending ladder height process corresponding to the ordinator constructed from the Lévy subordinators ${}^{\varepsilon}H^{+,(i)}$ defined by

$$ {}^{\varepsilon}H_t^{+,(i)} := (d_i^+ + \varepsilon)t + \sum_{s \le t} \Delta H_s^{+,(i)} \mathbb{1}_{(\varepsilon,\infty)}(\Delta H_s^{+,(i)}), \quad t \ge 0, $$

i.e. ${}^{\varepsilon}\boldsymbol{H}^{+,(i)}$ is obtained from $\boldsymbol{H}^{+,(i)}$ by deleting jumps smaller than ε and adding an additional drift ε. This ensures that ${}^{\varepsilon}\boldsymbol{H}^{+,(i)}$ is a compound Poisson subordinator with drift $d_i^+ + \varepsilon$ and Lévy measure $\Pi_i^{+,\varepsilon} = \Pi_i^+(\cdot \cap (\varepsilon, \infty))$ and hence we may apply (4.33) for the λ resolvent of the overshoot process

$$ ({}^{\varepsilon}\mathcal{O}_t^+, {}^{\varepsilon}\mathcal{J}_t^+)_{t \ge 0} := ({}^{\varepsilon}H^+_{T_t^{+,\varepsilon}} - t, J^+_{T_t^{+,\varepsilon}})_{t \ge 0}, $$

where $T_t^{+,\varepsilon} := \inf\{s \ge 0 : {}^{\varepsilon}H_s^+ > t\}$, $t \ge 0$. We first observe that for any $t > 0$ we obtain from Proposition 4.1

$$ \sup_{s \le t} |{}^{\varepsilon}H_s^+ - H_s^+| \le \varepsilon t + \sum_{s \le t} \Delta H_s^+ \mathbb{1}_{\{\Delta H_s^+ < \varepsilon\}}, $$

and since $\sum_{s \le t} \Delta H_s^+$ converges we obtain by dominated convergence that almost surely

$$ \sup_{s \le t} |{}^{\varepsilon}H_s^+ - H_s^+| \longrightarrow 0, \quad \text{as } \varepsilon \downarrow 0 $$

i.e. ${}^{\varepsilon}H^+$ converges to $\boldsymbol{H}^+$ uniformly on compact sets almost surely as $\varepsilon \downarrow 0$. Let Ξ be the set of $\mathbb{P}$-measure 1 on which ${}^{\varepsilon}\boldsymbol{H}^+$ and $(\boldsymbol{H}^+, \boldsymbol{J}^+)$ have càdlàg paths and on which the above convergence holds. Let $\omega \in \Xi$. Then ${}^{\varepsilon}H_\cdot^+(\omega), H_\cdot^+(\omega) \in \mathcal{D}(\mathbb{R}_+)$, the space of càdlàg functions mapping from $\mathbb{R}_+$ to $\mathbb{R}_+$, which we endow with Skorokhods J_1-topology. Since ${}^{\varepsilon}H_\cdot^+(\omega)$ converges uniformly on compact time sets to $H_\cdot^+(\omega)$, Proposition VI.1.17 in [98] tells us that ${}^{\varepsilon}H_\cdot^+(\omega)$ also converges with respect to the metric inducing the Skorokhod topology to $H_\cdot^+(\omega)$. For $t \ge 0$ let

$$ S_t \colon \mathcal{D}(\mathbb{R}_+) \longrightarrow [0, \infty], \quad \alpha \mapsto \inf\{s \ge 0 : |\alpha(s)| \ge t \text{ or } |\alpha(s-)| \ge t\}. $$

Since ${}^{\varepsilon}H_\cdot^+(\omega)$ and $H_\cdot^+(\omega)$ are strictly increasing it follows that $T_t^{+,\varepsilon}(\omega) = S_t({}^{\varepsilon}H_\cdot^+(\omega))$ and $T_t^+(\omega) = S_t(H_\cdot^+(\omega))$. Moreover, the set $\{t > 0 : S_t(H_\cdot^+(\omega)) \ne S_{t+}(H_\cdot^+(\omega))\}$ is empty by strictly increasing paths of $H_\cdot^+(\omega)$. Hence, we obtain from Proposition 2.11 and the proof of part c) of Proposition VI.2.12 in [98] that

$$ T_t^{+,\varepsilon}(\omega) = S_t({}^{\varepsilon}H_\cdot^+(\omega)) \longrightarrow S_t(H_\cdot^+(\omega)) = T_t^+(\omega), \quad \text{as } \varepsilon \downarrow 0, \tag{4.34} $$

and that for $t \notin \Lambda(\omega) = \{t > 0 : \Delta H^+_{T_t^+}(\omega) > 0 \text{ and } H^+_{T_t^+-}(\omega) = t\}$ we have

$$ {}^{\varepsilon}H^+_{T_t^{+,\varepsilon}}(\omega) \longrightarrow H^+_{T_t^+}(\omega), \quad \text{as } \varepsilon \downarrow 0. \tag{4.35} $$

But from the sawtooth structure of the paths of $\mathcal{O}$ it is easy to see that $\Lambda(\omega) = \{t > 0 : \Delta\mathcal{O}_t^+(\omega) > 0\}$, which is countable (alternatively, see Lemma VI.2.10.(d) in [98] for the same conclusion), hence non-convergence of ${}^{\varepsilon}\mathcal{O}_t^+(\omega)$ to $\mathcal{O}_t^+(\omega)$ only takes place on a set of Lebesgue measure 0. Furthermore, from (4.34) it follows that ${}^{\varepsilon}\mathcal{J}_t^+(\omega)$ converges to $\mathcal{J}_t^+(\omega)$ as $\varepsilon \downarrow 0$ except possibly on the set

$$ \Lambda'(\omega) := \{t > 0 : J^+_{T_t^+}(\omega) \ne J^+_{T_t^+-}(\omega)\} $$

$$= \{t > 0 : \Delta J^+_{T^+_t}(\omega) \neq 0, \Delta H^+_{T^+_t}(\omega) > 0\} \cup \{t > 0 : \Delta J^+_{T^+_t}(\omega) \neq 0, \Delta H^+_{T^+_t}(\omega) = 0\}$$
$$=: \Lambda'_1(\omega) \cup \Lambda'_2(\omega).$$

For $t \in \Lambda'_1(\omega)$ we have that in case $H^+_{T^+_t -}(\omega) < t \leq H^+_{T^+_t}(\omega)$ it holds that $T^+_s(\omega) = T^+_t(\omega)$ for $s \in [H_{T^+_t -}(\omega), t]$. Right-continuity of $s \mapsto T^+_s(\omega)$ and $s \mapsto J^+_s(\omega)$ therefore imply that for such t we also have ${}^{\varepsilon}\mathcal{J}^+_t(\omega) \to \mathcal{J}^+_t(\omega)$ as $\varepsilon \downarrow 0$. Further, since $t \mapsto H^+_t(\omega)$ is continuous in $T^+_t(\omega)$ if $\Delta H^+_{T^+_t}(\omega) = 0$, it follows from strictly increasing paths that for $s, t \in \Lambda'_2(\omega)$ we have $T^+_s(\omega) \neq T^+_t(\omega)$. Hence, $t \mapsto T^+_t(\omega)$ is injective on $\Lambda'_2(\omega)$. Since

$$T^+_{\cdot}(\omega)(\Lambda'_2(\omega)) = \{t > 0 : \Delta J^+_t(\omega) \neq 0, \Delta H^+_t(\omega) = 0\} \subset \{t > 0 : \Delta J^+_t(\omega) \neq 0\},$$

and the set on the right-hand side is countable thanks to $J^+_{\cdot}(\omega)$ being càdlàg, it follows that $\Lambda'_2(\omega)$ is countable as well. The above discussion therefore yields that the set of times $t > 0$ for which $J^{+}_{T^{+,\varepsilon}_t}(\omega)$ does not converge to $\mathcal{J}^+_t(\omega)$ is given by

$$\begin{aligned}\Lambda''(\omega) &:= \{t > 0 : \Delta J^+_{T^+_t}(\omega) \neq 0, H^+_{T^+_t -}(\omega) = t < H^+_{T^+_t}(\omega)\} \cup \Lambda'_2(\omega) \\ &\subset \{t > 0 : \Delta \mathcal{O}^+_t(\omega) > 0\} \cup \Lambda'_2(\omega),\end{aligned}$$

which is countable and therefore has Lebesgue measure 0 as well. It follows that for any $\omega \in \Xi$ we have for $f \in \mathcal{C}_b(\mathbb{R}_+ \times \Theta)$ by dominated convergence

$$\begin{aligned}\lim_{\varepsilon \downarrow 0} \int_0^\infty f({}^{\varepsilon}\mathcal{O}^+_t(\omega), {}^{\varepsilon}\mathcal{J}^+_t(\omega))\,\mathrm{d}t &= \int_{(\Lambda(\omega)\cup\Lambda''(\omega))^{\mathrm{c}}} \lim_{\varepsilon\downarrow 0} f({}^{\varepsilon}\mathcal{O}^+_t(\omega), {}^{\varepsilon}\mathcal{J}^+_t(\omega))\,\mathrm{d}t \\ &= \int_{(\Lambda(\omega)\cup\Lambda''(\omega))^{\mathrm{c}}} f(\mathcal{O}^+_t(\omega), \mathcal{J}^+_t(\omega))\,\mathrm{d}t \\ &= \int_0^\infty f(\mathcal{O}^+_t(\omega), \mathcal{J}^+_t(\omega))\,\mathrm{d}t.\end{aligned}$$

Consequently, if we denote by $\mathcal{U}^{\varepsilon}_{\lambda}$ the λ-resolvent for $({}^{\varepsilon}\boldsymbol{\mathcal{O}}^+, {}^{\varepsilon}\boldsymbol{\mathcal{J}}^+)$, the set Ξ having $\mathbb{P}$-measure 1 implies that for any $f \in \mathcal{C}_b(\mathbb{R}_+ \times \Theta)$

$$\begin{aligned}&(\mathcal{U}_\lambda f(0,i))_{i=1,\dots,n} \\ &\quad= \Big(\int_\Xi \lim_{\varepsilon\downarrow 0} \int_0^\infty f({}^{\varepsilon}\mathcal{O}^+_t(\omega), {}^{\varepsilon}\mathcal{J}^+_t(\omega))\,\mathrm{d}t\,\mathbb{P}^{0,i}(\mathrm{d}\omega)\Big)_{i=1,\dots,n} \\ &\quad= \lim_{\varepsilon\downarrow 0}\Big(\int_\Xi \int_0^\infty f({}^{\varepsilon}\mathcal{O}^+_t(\omega), {}^{\varepsilon}\mathcal{J}^+_t(\omega))\,\mathrm{d}t\,\mathbb{P}^{0,i}(\mathrm{d}\omega)\Big)_{i=1,\dots,n} \\ &\quad= \lim_{\varepsilon\downarrow 0}(\mathcal{U}^{\varepsilon}_\lambda f(0,i))_{i=1,\dots,n} \\ &\quad= \lim_{\varepsilon\downarrow 0} {}^{\varepsilon}\mathbf{\Phi}^+(\lambda)^{-1}\cdot\Big((d^+_i + \varepsilon) f(0,i) + \int_\varepsilon^\infty Q_\lambda f(x,i)\,\Pi^+_i(\mathrm{d}x) + \sum_{j\neq i} q^+_{i,j}\mathbb{E}[Q_\lambda f(\Delta^+_{i,j}, j)]\Big)^{\top}_{i=1,\dots,n} \\ &\quad= \mathbf{\Phi}^+(\lambda)^{-1}\cdot\Big(d^+_i f(0,i) + \int_0^\infty Q_\lambda f(x,i)\,\Pi^+_i(\mathrm{d}x) + \sum_{j\neq i} q^+_{i,j}\mathbb{E}[Q_\lambda f(\Delta^+_{i,j}, j)]\Big)^{\top}_{i=1,\dots,n} \\ &\quad=: \mathbf{\Upsilon}(\lambda),\end{aligned}$$

where we used dominated convergence for the second and (4.33) for the fourth equality. It remains to extend this result to any $f \in \mathcal{B}_+(\mathbb{R}_+ \times \Theta) \cup \mathcal{B}_b(\mathbb{R}_+ \times \Theta)$. To this end, let

$$\mathcal{M} := \left\{f \in \mathcal{B}_b(\mathbb{R}_+ \times \Theta) : (\mathcal{U}_\lambda f(0,i))_{i=1,\dots,n} = \Upsilon(\lambda)\right\}.$$

Clearly, $\mathcal{M}$ is a vector space over $\mathbb{R}_+$ by linearity of the Lebesgue integral and since $\mathcal{C}_b(\mathbb{R}_+ \times \Theta) \subset \mathcal{M}$, the constant function $\mathbb{1}_{\mathbb{R}_+ \times \Theta}$ is contained in $\mathcal{M}$. Moreover, dominated convergence shows that $\mathcal{M}$ is closed under convergence of an increasing family of functions f_n converging to some $f \in \mathcal{B}_b(\mathbb{R}_+ \times \Theta)$. Hence, $\mathcal{M}$ is a monotone vector space and since $\mathcal{C}_b(\mathbb{R}_+ \times \Theta)$ is closed under multiplication and contained in $\mathcal{M}$, the functional Monotone Class Theorem A.0.6 from [152] implies that all bounded $\sigma(\mathcal{C}_b(\mathbb{R}_+ \times \Theta))$-measurable functions are contained in $\mathcal{M}$. But since $\mathbb{R}_+ \times \Theta$ is a locally compact metric space with countable base, $\mathcal{C}_b(\mathbb{R}_+ \times \Theta)$ generates $\mathcal{B}(\mathbb{R}_+ \times \Theta)$ and hence $\mathcal{M} = \mathcal{B}_b(\mathbb{R}_+ \times \Theta)$ follows. For general $f \in \mathcal{B}_+(\mathbb{R}_+ \times \Theta)$ let $f_n := f\mathbb{1}_{\{f \in [0,n]\}} \in \mathcal{B}_b(\mathbb{R}_+ \times \Theta)$ and apply monotone convergence to deduce that (4.33) also holds for $f \in \mathcal{B}_+(\mathbb{R}_+ \times \Theta)$. This finishes the proof. ∎

4.B SUMMARY OF RESULTS FOR THE SPECIAL CASE OF LÉVY PROCESSES

This section is devoted to giving a (very) brief summary of Lévy processes and their overshoots that contains the main contributions of this chapter for the particular case of Lévy processes but can be read independently without any prior knowledge on MAPs. Moreover, this section prepares the reader for the developments in Chapter 5 by sketching how our convergence and mixing results on overshoots will be useful for developing data-driven ergodic control strategies for Lévy processes.

As seen before, talking about overshoots quite naturally guides us into fluctuation theory of Lévy processes, which is based on a rigorous treatment of excursions of Lévy processes from its extrema. For an extensive textbook treatment of fluctuation theory, we refer to [109] with basic properties of overshoots being discussed in Chapter 5. A general account on Lévy processes is given in the monographs [25] and [147].

We consider a Lévy process $\boldsymbol{X}$ with underlying natural filtration $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$ augmented in the usual way, which is equipped with a family of probability measures $(\mathbb{P}^x)_{x\in\mathbb{R}}$ such that $(\boldsymbol{X}, \mathbb{F}, (\mathbb{P}^x)_{x\in\mathbb{R}})$ is a Markov process. Thus, $\boldsymbol{X}$ has càdlàg paths, almost surely starts in x under $\mathbb{P}^x$, has stationary and independent increments under $\mathbb{P}^0$ and its semigroup $(P_t)_{t\geq 0}$ is given by

$$P_t(x, B) = \mathbb{P}^x(X_t \in B) = \mathbb{P}^0(X_t + x \in B), \quad x \in \mathbb{R}, B \in \mathcal{B}(\mathbb{R}).$$

From the last equality, it follows that $\boldsymbol{X}$ is *spatially homogeneous*, i.e. $\{\boldsymbol{X} + x, \mathbb{P}^0\} \overset{\mathrm{d}}{=} \{\boldsymbol{X}, \mathbb{P}^x\}$, and one easily derives that $\boldsymbol{X}$ is a Feller process, that is $P_t\mathcal{C}_0(\mathbb{R}) \subset \mathcal{C}_0(\mathbb{R})$ for any $t \geq 0$ and for any $f \in \mathcal{C}_0(\mathbb{R})$, $P_t f \longrightarrow f$ strongly as $t \longrightarrow 0$.

While the Fellerian nature of Lévy processes is still fairly general from a Markovian perspective, it is the spatial homogeneity which gives rise to a quite unique and powerful theory. The fundamental starting point to the analysis of Lévy processes is the Lévy–Khintchine formula, which identifies the characteristic function of the marginals of the process and hence uniquely describes the complete process in terms of a *characteristic triplet* (a, σ, Π), where $a \in \mathbb{R}$, $\sigma \geq 0$ and Π is a measure on $\mathbb{R}$ (called Lévy measure) having no atom at 0 and being such that

$\int_{\mathbb{R}} 1 \wedge x^2 \, \Pi(\mathrm{d}x) < \infty$. The Lévy–Khintchine formula then states that $\mathbb{E}^0[\exp(\mathrm{i}\theta X_t)] = \exp(t\Psi(\theta))$, $\theta \in \mathbb{R}$, $t \geq 0$, with the *characteristic exponent* Ψ satisfying

$$\Psi(\theta) = \mathrm{i}a\theta - \frac{\sigma^2\theta^2}{2} + \int_{\mathbb{R}} \left(\mathrm{e}^{\mathrm{i}\theta x} - 1 - \mathrm{i}\theta x \mathbb{1}_{[-1,1]}(x)\right) \Pi(\mathrm{d}x).$$

On the level of processes, the Lévy–Khintchine representation can be translated into a partition of the process into a linear Brownian motion $(at + \sigma B_t)_{t \geq 0}$ and an independent pure jump process characterized by Π, which itself decomposes into an independent compound Poisson process and a zero mean L^2 martingale with infinitely many jumps bounded by 1 on any finite time interval. If $\int_1^\infty x \, \Pi(\mathrm{d}x) < \infty$, it follows from the Lévy–Khintchine representation that the first moment of X_t is finite for any $t \geq 0$ and $\mathbb{E}^0[X_t] = t\eta$ with

$$\eta = \mathbb{E}^0[X_1] = a + \int_{\mathbb{R}\setminus[-1,1]} x \, \Pi(\mathrm{d}x)$$

determining the long-time behaviour of $\boldsymbol{X}$ in the sense that

(i) $\eta > 0 \implies \lim_{t\to\infty} X_t = \infty$, $\mathbb{P}^0$-a.s.;

(ii) $\eta < 0 \implies \lim_{t\to\infty} X_t = -\infty$, $\mathbb{P}^0$-a.s.;

(iii) $\eta = 0 \implies \limsup_{t\to\infty} X_t = -\liminf_{t\to\infty} X_t = \infty$, $\mathbb{P}^0$-a.s..

and

$$\lim_{t\to\infty} \frac{X_t}{t} = \eta, \quad \mathbb{P}^0\text{-a.s..}$$

With this basic preparation on the characteristics of Lévy processes, let us now come to their fluctuation theory, with a certain emphasis on the so called Wiener–Hopf factorization. This commands a discussion of the ascending ladder height process, which will be central to our analysis of data-driven solutions to ergodic control problems associated to Lévy processes in Chapter 5. This process is derived from the local time at the supremum $\mathsf{L} = (\mathsf{L}_t)_{t\geq 0}$, which is a stochastic process measuring the time that $\boldsymbol{X}$ spends at its running supremum $\overline{X}_t = \sup_{0\leq s\leq t} X_s$, $t \geq 0$. Its construction is based on the observation that $\mathbf{Y} = (\overline{X}_t - X_t)_{t\geq 0}$, which can be interpreted as the process obtained from reflecting $\boldsymbol{X}$ at its supremum, is a strong Markov process and hence one can define L as the local time at 0 for $\mathbf{Y}$, which means that L is an additive functional of $\mathbf{Y}$ which almost surely increases on the closure of $\{t \geq 0 : Y_t = 0\} = \{t \geq 0 : X_t = \overline{X}_t\}$. In case that $\boldsymbol{X}$ is *upward regular*, i.e., for $T_0 \coloneqq \inf\{t \geq 0 : X_t > 0\}$ we have $\mathbb{P}^0(T_0 = 0) = 1$, L can be constructed as a process with almost surely continuous paths, which entails that its right-continuous inverse $\mathsf{L}_t^{-1} = \inf\{s \geq 0 : L_s > t\}$, $t \geq 0$, is almost surely *strictly* increasing. In this case, the ascending ladder height process $\mathbf{H} = (H_t)_{t\geq 0}$, defined by

$$H_t = \begin{cases} X_{\mathsf{L}_t^{-1}}, & \text{if } 0 \leq t < \mathsf{L}_\infty \\ \infty, & \text{if } t \geq \mathsf{L}_\infty, \end{cases}$$

is a *killed* subordinator, strictly increasing up to its lifetime L_∞, i.e., $\boldsymbol{H}$ is equal in law to a strictly increasing Lévy process, which is sent to the cemetery state ∞ after an independent exponentially distributed random time with expectation $\mathbb{E}^0[\mathsf{L}_\infty]$. If $\limsup_{t\to\infty} X_t = \infty$, which as seen before is guaranteed if $\mathbb{E}^0[|X_1|] < \infty$ and $\eta = \mathbb{E}^0[X_1] \geq 0$, it follows that $\mathsf{L}_\infty = \infty$ almost surely and hence $\boldsymbol{H}$ is unkilled. Moreover, when $\eta > 0$, which is the setting that we will be working with in Chapter 5, it holds that $0 < \mathbb{E}^0[H_1] < \infty$ as well, which can be deduced from (4.37) below. It is important to note that L is only characterized uniquely up to a multiplicative constant and hence the definition of $\boldsymbol{H}$ depends on the chosen scaling of local time. For our purposes, it will be convenient to choose a scaling of local time such that $\mathbb{E}^0[\mathsf{L}_1^{-1}] = 1$ and hence, by Wald's equality, $\mathbb{E}^0[H_1] = \mathbb{E}^0[X_1]\mathbb{E}^0[\mathsf{L}_1^{-1}] = \mathbb{E}^0[X_1]$.

When upward regularity does not hold, $(\mathsf{L}_t^{-1})_{t\geq 0}$ will not be strictly increasing and consequently, the ascending ladder height process $\boldsymbol{H}$ is a (possibly killed) compound Poisson subordinator. In any scenario, the Laplace exponent $\Phi_H(\lambda)$, given by

$$\Phi_H(\lambda) = q + d_H\lambda + \int_0^\infty (1 - \mathrm{e}^{-\lambda x})\, \Pi_H(\mathrm{d}x), \quad \lambda \geq 0,$$

satisfies $\mathbb{E}^0[\exp(-\lambda H_t)] = \exp(-t\Phi_H(\lambda))$ and we refer to d_H as the drift and Π_H as the Lévy measure of $\boldsymbol{H}$.

In the same vein, we can construct the ascending ladder height process $\widehat{\boldsymbol{H}} = (\widehat{H}_t)_{t\geq 0}$ for the dual Lévy process $\widehat{\boldsymbol{X}} = -\boldsymbol{X}$, which corresponds to time changing $\boldsymbol{X}$ by the right continuous inverse of local time $\widehat{\mathsf{L}}$ at the *infimum* of $\boldsymbol{X}$. Therefore, $\widehat{\boldsymbol{H}}$ is referred to as the *descending* ladder height process. If we denote by $\widehat{\Phi}_H$ the Laplace exponent of $\widehat{\boldsymbol{H}}$, then the Wiener–Hopf factorization tells us that $\boldsymbol{X}$ is fully characterized by means of the ascending and descending ladder height processes since the Lévy–Khintchine exponent of $\boldsymbol{X}$ can be expressed as a factorization of the Laplace exponents of $\boldsymbol{H}$ and $\widehat{\boldsymbol{H}}$,

$$\Psi(\theta) = -\mathrm{c}\Phi_H(-\mathrm{i}\theta)\widehat{\Phi}_H(\mathrm{i}\theta), \quad \theta \in \mathbb{R}, \tag{4.36}$$

where the constant c depends on the scaling of local time at the supremum and infimum. Among others, this factorization allows to express the characteristics of $\boldsymbol{H}$ in terms of the characteristics of $\boldsymbol{X}$ and $\widehat{\boldsymbol{H}}$. A particularly useful identity for understanding the ascending ladder height Lévy measure, usually referred to as Vigon's équation amicale inversé, was demonstrated in [172] and generalized in Theorem 4.30 for MAPs:

$$\Pi_H(\mathrm{d}x) = \int_0^\infty \Pi(y + \mathrm{d}x)\, \widehat{U}_H(\mathrm{d}y), \quad x > 0. \tag{4.37}$$

Here, $\widehat{U}_H(\mathrm{d}x) = \mathbb{E}^0[\int_0^\infty \mathbb{1}_{\{\widehat{H}_t \in \mathrm{d}x\}}\,\mathrm{d}t]$, $x \geq 0$, denotes the potential measure of $\widehat{\boldsymbol{H}}$ and without loss of generality, the constant c in (4.36) is set to unity.

While the theoretical solution strategy of the Lévy driven impulse control problem from Chapter 4 will be driven by the generator functional $f = \mathcal{A}_H\gamma$ with $\mathcal{A}_H$ denoting the generator of the ascending ladder height process $\boldsymbol{H}$, the data-driven reflection strategy that we shall develop makes use of the link between $\boldsymbol{H}$ and overshoots $\boldsymbol{\mathcal{O}} = (\mathcal{O}_t)_{t\geq 0}$ of $\boldsymbol{X}$ to find an estimator $\widehat{f}_T$, from which the optimal reflection boundary will be approximated by a greedy strategy. Let us first discuss classical properties of overshoots and then make the transition to the results from the main part of this chapter, which will be fundamental for the approximation properties of the statistical procedure.

Let $t \geq 0$ be a given *level* and consider the overshoot $\mathcal{O}_t$ over t, given by

$$\mathcal{O}_t = X_{T_t} - t$$

on $\{T_t < \infty\}$, where $T_t := \inf\{s \geq 0 : X_s > t\}$. Let us assume from here on that

($\mathscr{L}0$) $\boldsymbol{X}$ is upward regular, i.e.,

$$\mathbb{P}^0(\inf\{t \geq 0 : X_t > 0\} = 0) = 1,$$

and moreover $0 < \mathbb{E}^0[X_1] =: \eta < \infty$.

is in place, which in particular implies $T_t < \infty$ almost surely, $0 < \mathbb{E}^0[H_1] < \infty$ and that $\boldsymbol{H}$ is an unkilled, strictly increasing subordinator. The fundamental link between $\mathfrak{O}$ and $\boldsymbol{H}$ now stems from the observation that, due to its construction, the range of $\boldsymbol{H}$ almost surely coincides with the range of the running supremum process $(\overline{X}_t)_{t\geq 0}$. As a consequence, the overshoot process $\mathfrak{O}^H$ associated to $\boldsymbol{H}$ is indistinguishable from the overshoot process $\mathfrak{O}$ associated to $\boldsymbol{X}$. Hence, if we want to estimate the characteristics of $\boldsymbol{H}$ (which cannot be observed based on a sample of $\boldsymbol{X}$ due to a lack of explicitness of the local time L), one could hope for utilizing the overshoot link to get hold of $\boldsymbol{H}$ based on observations of $\boldsymbol{X}$, provided that $\mathfrak{O}$ has some kind of regularity structures. Indeed, making use of the compensation formula (cf. [109, Theorem 4.4]), it can be shown that the law of $\mathcal{O}_t = \mathcal{O}_t^H$ is given by

$$\mathbb{P}^x(\mathcal{O}_t \in \mathrm{d}y) = \delta_{x-t}(\mathrm{d}y)\mathbb{1}_{[0,x]}(t) + \int_{[0,t-x]} \Pi_H(u+\mathrm{d}y)\, U_H(t-x-\mathrm{d}u)\mathbb{1}_{(x,\infty)}(t), \quad y > 0,$$

(see [109, Theorem 5.6]) and resorting to classical renewal arguments it can be deduced from this formula that (cf. [109, Theorem 5.7])

$$\mathbb{P}^x(\mathcal{O}_t \in \mathrm{d}y) \xrightarrow[t\to\infty]{\mathrm{w}} \frac{1}{\mathbb{E}^0[H_1]}\big(d_H\delta_0(\mathrm{d}y) + \mathbb{1}_{(0,\infty)}(y)\Pi_H((y,\infty))\,\mathrm{d}y\big) =: \mu(\mathrm{d}y), \quad y \geq 0. \tag{4.38}$$

As discussed before, similar results can be obtained for overshoots associated to MAPs, as a natural regime switching generalization of Lévy processes. By considering a MAP with a trivial underlying Markov chain, we obtain a Lévy process by projection and hence results on overshoots of MAPs have direct analogues for overshoots of Lévy processes and thus, our results from this chapter can be directly translated to the Lévy case. Let

($\mathscr{L}1$) either, $d_H > 0$, or there exists $(a,b) \subset (0,\infty)$ such that $\lambda|_{(a,b)} \ll \Pi_H|_{(a,b)}$.

Then Theorem 4.19 shows that under assumptions ($\mathscr{L}0$) and ($\mathscr{L}1$), weak convergence in (4.38) can be improved to total variation convergence—or said differently to ergodicity of the Markov process $\mathfrak{O}$.

Proposition 4.39 (Theorem 4.19, Lemma 4.34 and Theorem 7.11 in [109]). *Given ($\mathscr{L}1$), it holds that, for any $x \geq 0$,*

$$\mathbb{P}^x(\mathcal{O}_t \in \cdot) \xrightarrow[t\to\infty]{\mathrm{TV}} \mu. \tag{4.39}$$

Moreover, ($\mathscr{L}1$) is fulfilled if one of the following conditions hold:

(i) $\exists (a,b) \subset \mathbb{R}_+$ *s.t.* $\lambda|_{(a,b)} \ll \Pi|_{(a,b)}$;

(ii) $\boldsymbol{X}$ *has bounded variation with Lévy–Khintchine exponent*

$$\Psi(\theta) = \mathrm{i}\delta\theta + \int_{\mathbb{R}} \big(\mathrm{e}^{\mathrm{i}\theta x} - 1\big)\,\Pi(\mathrm{d}x),$$

and $\delta > 0$;

(iii) *$\boldsymbol{X}$ has a Gaussian component;*

(iv) *$\boldsymbol{X}$ has positive jumps, unbounded variation, no Gaussian component and its Lévy measure Π satisfies*

$$\int_0^1 \frac{x\Pi((x,\infty))}{\int_0^x \int_y^1 \Pi((-1,-u))\,\mathrm{d}u\,\mathrm{d}y}\,\mathrm{d}x < \infty.$$

Remark 4.40. The mutually exclusive conditions (ii)-(iv) are necessary and sufficient criteria for $d_H > 0$, which implies total variation convergence.

To establish exponential rates of convergence in (4.39) under the natural assumption that H_1 possesses an exponential moment, a Lyapunov-type drift criterion given in [74, Theorem 5.2] is combined with an explicit calculation of the resolvent kernel

$$\mathcal{R}_\lambda(x,\cdot) := \int_0^\infty \lambda \mathrm{e}^{-\lambda t}\mathbb{P}^x(\mathcal{O}_t \in \cdot)\,\mathrm{d}t,$$

in Theorem 4.7. Building on exponential ergodicity it is then shown that under

($\mathscr{L}$2) there is $\lambda > 0$ such that $\mathbb{E}^0[\exp(\lambda H_1)] < \infty$,

$\mathfrak{O}$ is exponentially β-mixing for any initial distribution possessing an exponential moment, which in particular includes the stationary distribution μ.

PROPOSITION 4.41 (Theorem 4.22, Theorem 4.25 and Lemma 4.34/Théorème 6.2.3 in [171]). *Grant ($\mathscr{L}$1) and ($\mathscr{L}$2). Then, convergence in* (4.39) *takes place at exponential rate. More precisely, for any $\delta \in (0,1)$ there exists a constant $c(\delta) > 0$ such that*

$$\left\|\mathbb{P}^x(\mathcal{O}_t \in \cdot) - \mu\right\|_{\mathrm{TV}} \leq c(\delta)\mathcal{R}_\lambda \exp(\lambda\cdot)(x)\mathrm{e}^{-t/(2+\delta)}, \quad t \geq 0.$$

Moreover, if η is some distribution on $(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$ such that $\eta(\exp(\lambda\cdot)) < \infty$, then, if $\boldsymbol{X}$ is started in η, $\mathfrak{O}$ is exponentially β-mixing with rate

$$\beta_{\mathbb{P}^\eta}(t) \leq 2\varrho(\eta,\lambda,\delta)\mathrm{e}^{-t/(2+\delta)},$$

where

$$\varrho(\eta,\lambda,\delta) = c(\delta)\sup_{t\geq 0}\mathbb{E}^\eta\left[\mathcal{R}_\lambda \exp(\lambda\cdot)(\mathcal{O}_t)\right] < \infty.$$

Finally, ($\mathscr{L}$2) is satisfied if and only if

$$\int_1^\infty \mathrm{e}^{\lambda x}\,\Pi(\mathrm{d}x) < \infty.$$

Data-driven control strategies for diffusions and Lévy processes

5

From a purely mathematical point of view, the field of statistics of stochastic processes is very appealing as it lives from the combination of different techniques and findings from diverse mathematical areas, in particular statistics, probability theory or functional analysis. The fundamental motivation of this branch of statistics, however, results from concrete applications. Thus, besides mathematical elegance and completeness, the developments and results in this area should always be tested in terms of their applicability.

An important area in which stochastic processes (especially of diffusion-type) are used by default to account for random impacts is stochastic control theory. Whereas the theory itself is very well developed and offers concrete decision strategies for a variety of problems, these are usually based on the assumption that the decision maker has full knowledge of the dynamics of the underlying random process. In [50], the authors already presented an approach to overcome this constraint by means of nonparametric estimation methods and proposed a fully data-driven approach to solving a concrete impulse control problem. In this chapter, we are expanding the view and approaching the problem from a general perspective. Basic components for the data-based solution of a large class of stochastic control problems are

» the control of the sup-norm risk for the estimation of certain (functionals of) characteristics of the random process, in particular

» the derivation of upper bounds on the convergence rate.

In Section 5.1.1, we describe the nature of the control problems and how they naturally lead to associated nonparametric estimation problems. Based on this, in Section 5.1.2, we briefly formulate our general statistical modelling framework.

5.1 Introduction

5.1.1 *The motivating control problems*

The stochastic control problems we consider in this chapter are—under the assumption that the decision maker has access to the underlying dynamics—classical, and variants are well-studied. They have in common that a decision maker controls a continuous-time process $\boldsymbol{X}$ on the real line, but the controls do not change continuously over time, but are of a singular type. More precisely, it turns out that the optimal strategies call for reflecting the underlying process at certain boundaries. These optimal boundaries can be found (semi-) explicitly as optimizers of certain (deterministic) auxiliary functions, based on the dynamics of the underlying uncontrolled process. In this chapter, we consider the more realistic situation that the decision maker has to estimate the underlying dynamics while controlling the process. The main key for such a statistical treatment is that, for an underlying ergodic scalar diffusion $\boldsymbol{X}$, the corresponding auxiliary function can be described explicitly in terms of the invariant density, as detailed in Section 5.2. For estimating the optimizer in this case, the sup-norm risk of invariant density estimators has to be studied. In Section 5.3, we then turn our attention to a control problem for underlying Lévy processes. In this case, the auxiliary function is identified as a generator functional $\mathcal{A}_H\gamma$ of the ascending ladder height process $\boldsymbol{H}$ belonging to a Lévy process $\boldsymbol{X}$, which we encountered in Chapter 4. Again, a sup-norm estimation procedure for such functionals has to be found.

A data-driven solution method for the control problems therefore naturally leads to the challenging statistical problem of setting set up a framework such that the seemingly different issues of sup-norm estimation of the invariant density of an ergodic diffusion on the real axis and sup-norm estimation of ladder height generator functionals $\mathcal{A}_H\gamma$ for a Lévy process $\boldsymbol{X}$ can be integrated into.

5.1.2 *Nonparametric analysis: Controlling the sup-norm risk of Markovian functionals*

The identification of an appropriate technical framework is a crucial issue for the statistical analysis of stochastic processes. Specific model choices such as scalar diffusion processes or multivariate reversible processes with continuous trajectories permit the application of particular technical tools (e.g., associated to diffusion local time or to the symmetry of the semigroup), but generally do not provide any information about the robustness of the used statistical methods beyond the chosen framework. In contrast, exponential β-mixing of general continuous-time Markov processes $\boldsymbol{X} = (X_t)_{t\geq 0}$ has been identified in Chapter 3 as a criterion which, on the one hand, is strong enough to serve as a central building block of a robust statistical analysis while, on the other hand, providing sufficient generality to allow to include an exhaustive list of Markov processes in the framework. Statistical properties of such processes can thus be studied based on fairly general results rooted in stability theory of Markov processes. Our technical main result in this regard was Theorem 3.7, which gives nonasymptotic bounds on the moments of suprema of empricial processes associated to Markovian integral functionals. More precisely, if $\boldsymbol{X}$ is a stationary non-explosive Borel right process with invariant distribution μ having the exponential β-mixing property, i.e., there exist constants $\kappa, c_\kappa > 0$ such that

$$\beta(t) = \int_{\mathcal{X}} \|P_t(x,\cdot) - \mu\|_{\mathrm{TV}}\, \mu(\mathrm{d}x) \leq c_\kappa \mathrm{e}^{-\kappa t}, \quad t \geq 0, \tag{5.1}$$

and $\mathcal{G}$ is a countable class of bounded real-valued functions g satisfying $\mu(g) = 0$, then if we define

$$\mathbb{G}_T(g) = \frac{1}{\sqrt{T}} \int_0^T g(X_t)\,\mathrm{d}t, \quad T > 0, g \in \mathcal{G},$$

and let $m_t \in (0, t/4]$, there exist $\tau \in [m_t, 2m_t]$ and universal constants $C_1, C_2, c_1, c_2 > 0$ such that, for any $1 \leq p < \infty$,

$$\begin{aligned}\left(\mathbb{E}\left[\sup_{g\in\mathcal{G}} |\mathbb{G}_t(g)|^p\right]\right)^{1/p} &\leq C_1 \int_0^\infty \log \mathcal{N}\big(u, \mathcal{G}, \tfrac{2m_t}{\sqrt{t}} d_\infty\big)\,\mathrm{d}u + C_2 \int_0^\infty \sqrt{\log \mathcal{N}(u, \mathcal{G}, d_{\mathbb{G},\tau})}\,\mathrm{d}u \\ &\quad + 4 \sup_{g\in\mathcal{G}} \Big(\frac{2m_t}{\sqrt{t}} \|g\|_\infty c_1 p + \|g\|_{\mathbb{G},\tau} c_2 \sqrt{p} + \frac{1}{2}\|g\|_\infty c_\kappa \sqrt{t} \mathrm{e}^{-\frac{\kappa m_t}{p}}\Big),\end{aligned} \tag{5.2}$$

where

$$d^2_{\mathbb{G},T}(f,g) := \mathrm{Var}\Big(\frac{1}{\sqrt{T}} \int_0^T (f-g)(X_t)\,\mathrm{d}t\Big), \quad T > 0, f, g \in \mathcal{G}$$

is the semidistance associated to the variance of the integral functionals wrt. $\mathbb{P}^{\mu}$. As demonstrated in Chapter 3, this result covers a wide range of potential applications. For example, it can be used to find optimal upper bounds (regarding the sup-norm risk over bounded domains) for nonparametric estimation of the invariant density for $\mathbb{R}^d$-valued Markov processes with transition densities (cf. Sections 3.3 and 3.4.) Recall that such results are derived from Theorem 3.7 by bounding the (pseudo-) norms and thus the associated entropy integrals for the function class $\mathcal{G}$ related to the chosen estimation procedure. For d_∞, this can be achieved by using the analytical properties of $\mathcal{G}$, while bounds on the pseudo-metric $d_{\mathbb{G},\tau}$ are based on suitable bounds for the variance of integral functionals of $\boldsymbol{X}$. As shown in Section 3.1.2, for $d \geq 2$ this is taken care of using the exponential β-mixing property of $\boldsymbol{X}$ once we assume that, additionally, an on-diagonal heat kernel estimate is in place for the densities $(p_t)_{t\geq 0}$ of the Markov semigroup, i.e., there exists some constant $C > 0$ such that

$$\forall t \in (0,1]: \qquad \sup_{x,y\in\mathbb{R}^d} p_t(x,y) \leq Ct^{-d/2}.$$

Application in the scalar setting

In Section 3.1.1 we introduced a related condition on local uniform convergence, which is satisfied whenever the process has bounded transition densities and is exponentially ergodic. Together with heat-kernel bounds on the short time transitional behavior of the process, this allowed us to prove optimal convergence rates for sup-norm estimation of the invariant density in any dimension. In the following two types of scalar estimation problems will be of central importance: one which fits perfectly into the above frame, and another, which must be tackled differently because the ergodic process under consideration does not necessarily have transition densities and even when this is the case they are not bounded in general.

In Section 5.2.1, we study kernel invariant density estimation (for scalar ergodic diffusions) which requires a careful balancing of *bias* and stochastic error of the estimator (in case of pointwise risk, the well-known bias–variance tradeoff) by choosing an appropriate bandwidth h. In dimension $d = 1$, the exponential β-mixing property is not quite sufficient per se, but exponential ergodicity with locally bounded penalty function together with (a relaxation of) the on-diagonal heat kernel estimate of the semigroup guarantee variance bounds that are tight enough for proving optimal upper bounds on the convergence rates. Both of these properties hold under classical coefficient assumptions that we will impose on the diffusion process. This is of considerable independent interest since, in contrast to the local time arguments usually employed for the statistical analysis of scalar diffusions, the techniques generalize without much effort to the multivariate diffusion case. Our framework therefore arguably closes the gap between the relatively distinct approaches to statistical estimation of scalar and multivariate diffusions (see, e.g., [60] vs. [59] or [57, 58] vs. [159, 160]). Moreover, it potentially extends results obtained exclusively for symmetric diffusions to the general case since it is not reliant on functional inequalities, which are not well-suited to the non-reversible setup, see also the discussion in Chapter 3.

Suppose that we can find an *unbiased* estimator of the characteristic we are interested in. In this situation, a fine analysis of the variance of Markovian integral functionals is not necessarily needed. This is, e.g., the case if we can express the quantity of interest as an integral wrt. the stationary distribution of some stationary Markov process $\boldsymbol{X}$, since then the continuous-time mean estimator

$$\frac{1}{T}\int_0^T f(X_t)\,\mathrm{d}t$$

is unbiased. If $\boldsymbol{X}$ is moreover β-mixing, we can make use of Theorem 3.7 based on purely analytical arguments. This will become clear in Section 5.3.1, where we are investigating sup-norm estimation of generator functionals $\mathcal{A}_H\gamma$ of the ascending ladder height process $\boldsymbol{H}$ belonging to a Lévy process $\boldsymbol{X}$ via an unbiased mean estimator based on overshoots of $\boldsymbol{X}$. The thereby established sup-norm bounds will be of central importance for the procedure in Section 5.3.2.

5.1.3 *Overview*

In Section 5.2 we develop a data-driven strategy for a singular control problem associated to a scalar diffusion process. The construction and error anaylsis is given in Section 5.2.2 based on a minimax optimal estimation procedure for the stationary density under exponential ergodicity assumptions, which is carried out in Section 5.2.1. In Section 5.3 a data driven strategy for an impulse control problem with an underlying Lévy process is constructed. The statistical foundations for the estimation strategy of Section 5.3.2 is presented in Section 5.3.1.

5.2 DATA-DRIVEN SINGULAR CONTROLS FOR DIFFUSIONS ON THE REAL LINE

We now introduce the singular control problem for underlying scalar diffusion processes, given as a solution of the Itō-type SDE

$$\mathrm{d}X_t \;=\; b(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t, \tag{5.3}$$

$b,\sigma\colon\mathbb{R}\longrightarrow\mathbb{R}$ some measurable functions and $(W_t)_{t\geq 0}$ a standard Brownian motion on a probability space $(\Omega,\mathcal{F},\mathbb{P}_b)$. One motivation for considering such problems comes from investigating optimal dividend distributions [10, 16, 38]. Another stream of literature deals with determining a policy that optimizes the expected cumulative present value of the harvesting [9, 92, 112]. In particular for the latter application, it is natural to study an ergodic formulation, as it reflects the idea of considering sustainable harvesting guidelines, which we will also use here.

Assume that for some constants $\bar{\nu},\underline{\nu}\in(0,\infty)$, σ is continuously differentiable with bounded derivative (thus globally Lipschitz) and satisfies $\underline{\nu}\leq|\sigma(x)|\leq\bar{\nu}$ for all $x\in\mathbb{R}$. For fixed constants $A,\gamma>0$ and $\mathfrak{C}\geq 1$, define the set $\boldsymbol{\Sigma}=\boldsymbol{\Sigma}(\mathfrak{C},A,\gamma,\sigma)$ as

$$\boldsymbol{\Sigma} := \left\{b\in\operatorname{Lip}(\mathbb{R}) : |b(x)|\leq\mathfrak{C}(1+|x|),\ \forall|x|>A\colon\ \frac{b(x)}{\sigma^2(x)}\operatorname{sgn}(x)\leq-\gamma\right\}.$$

Note that a linear growth condition for Lipschitz drift b is always satisfied, but the class $\boldsymbol{\Sigma}$ specifies a global magnitude of this maximal growth in terms of the constant $\mathfrak{C}$. Moreover, given σ as above and any $b\in\boldsymbol{\Sigma}$, an immediate consequence is that there exists a strong solution $\boldsymbol{X}$ of the SDE (5.3) for given initial value X_0 independent of $\boldsymbol{W}$. If we let $\mathbb{P}_b^x=\mathbb{P}_b(\cdot|X_0=x)$, then $(\boldsymbol{X},(\mathbb{P}_b^x)_{x\in\mathbb{R}})$ defines a non-explosive Feller Markov process [150, Theorem 19.9] and thus in particular a Borel right process. Moreover, $\boldsymbol{X}$ has a unique stationary distribution $\mu=\mu_b$ having invariant density

$$\rho(x)=\rho_b(x):=\frac{1}{C_{b,\sigma}\sigma^2(x)}\exp\left(\int_0^x\frac{2b(y)}{\sigma^2(y)}\,\mathrm{d}y\right),\quad x\in\mathbb{R}, \tag{5.4}$$

with normalizing constant $C_{b,\sigma} := \int_{\mathbb{R}} \frac{1}{\sigma^2(u)} \exp\Big(\int_0^u \frac{2b(y)}{\sigma^2(y)}\, \mathrm{d}y\Big)\, \mathrm{d}u$. This follows from [84, §18, Lemma 3], but can also readily be deduced from the generator characterization of stationary distributions given in Section 2.2, see also [4, Section 3.2]. In the following, we will abbreviate $\mathbb{P}_b^{\mu_b} = \mathbb{P}_b$ and, if there is no room for confusion, also just write $\mathbb{P}$ instead. For any $b \in \boldsymbol{\Sigma}$, $\sigma^2\rho_b$ is continuously differentiable and there exists a constant $\rho^* > 0$ (depending only on $\mathfrak{C}, A, \gamma, \underline{\nu}, \bar{\nu}$) such that

$$\sup_{b\in\boldsymbol{\Sigma}(\mathfrak{C},A,\gamma,\sigma)} \max\big\{\|\rho_b\|_\infty,\ \|(\sigma^2\rho_b)'\|_\infty\big\} < \rho^*. \tag{5.5}$$

Furthermore, for any fixed bounded set $D \subset \mathbb{R}$, there exists some $\rho_* > 0$ (depending again only on $\boldsymbol{\Sigma}$) such that

$$\forall x \in D, \qquad \inf_{b\in\boldsymbol{\Sigma}(\mathfrak{C},A,\gamma,\sigma)} \rho_b(x) \geq \rho_*. \tag{5.6}$$

The controls used to formulate the problem are of the form $\boldsymbol{Z} = (U_t, D_t)_{t\geq 0}$ for non-decreasing, right-continuous and adapted processes $\boldsymbol{U}$ and $\boldsymbol{D}$. Here, U_t and D_t denote the cumulative upwards and downwards controls, resp. These processes can be decomposed into singular and jump part as

$$U_t = U_t^c + \sum_{0\leq s\leq t} (U_s - U_{s-}),\ \ D_t = D_t^c + \sum_{0\leq s\leq t} (D_s - D_{s-}),$$

where $\boldsymbol{U}^c$ and $\boldsymbol{D}^c$ are continuous. In the following, we will mostly deal with a special class of controls for which the jump part is absent (with a possible exception at $t = 0$): $\boldsymbol{U}$ and $\boldsymbol{D}$ are associated to the local times at certain fixed points ξ, θ.

We denote the set of all controls by $\boldsymbol{\Lambda}$ and, for each $\boldsymbol{Z} \in \boldsymbol{\Lambda}$, we define the controlled process $\boldsymbol{X}^Z$ as the solution to

$$\mathrm{d}X_t^Z \ = \ b(X_t^Z)\,\mathrm{d}t + \sigma(X_t^Z)\,\mathrm{d}W_t + \mathrm{d}U_t - \mathrm{d}D_t,$$

where we work under the assumption that $b \in \boldsymbol{\Sigma}$, implying in particular that the uncontrolled process $\boldsymbol{X} = \boldsymbol{X}^0$ has a stationary distribution $\rho = \rho_b$.

The problem to be studied is now to determine the minimal value and the minimizer of

$$\limsup_{T\to\infty} \frac{1}{T}\bigg(\int_0^T c(X_s^Z)\,\mathrm{d}s + q_u U_T + q_d D_T\bigg), \tag{5.7}$$

where c is a continuous, nonnegative function with

$$\sup_{b\in\boldsymbol{\Sigma}} \int c(x)\rho_b(x)\,\mathrm{d}x < \infty$$

modelling the running costs and q_u, q_d are positive constants describing the (proportional) costs associated with applying a control. We can interpret our goal as keeping $\boldsymbol{X}$ close to the target state 0, say, and therefore assume that c has a minimum in 0. The goal in the sequel is to find a data-driven strategy for problem (5.7) when the drift b of the underlying process is unknown. While parts of the following analysis are similar to the one in [50], it here turns out to be essential to control the sup-norm risk of estimators of the characteristics (precisely, the invariant density ρ_b) of $\boldsymbol{X}$ solving (5.3).

5.2.1 *Estimating the stationary density of ergodic diffusion processes*

We first show how the underlying statistical problem for the uncontrolled process can be integrated into our general framework recalled in Section 5.1.2. For a large class of ergodic scalar diffusion processes $\boldsymbol{X}$ solving (5.3) it is known that, given continuous observations $(X_t)_{0\leq t\leq T}$, the invariant density can be estimated with a parametric rate of convergence. For an overview, we refer to Sections 1.3.2 and 4.2 in [106]. It is however not straightforward to extend bounds on the pointwise or L^2 risk to the sup-norm bounds required for our application. A corresponding result is given in Corollary 13 in [3] whose proof, however, is deeply rooted in continuous martingale techniques. We show how this behaviour can also be deduced from mixing properties of the diffusion.

Given some fixed domain $D\subset\mathbb{R}$, constants $\beta,\boldsymbol{\vartheta},\boldsymbol{\varrho},\boldsymbol{\chi}>0,\mathsf{L}>0$ and a measurable function $V\geq 1$, introduce the set $\widetilde{\Sigma}_D(\beta)=\widetilde{\Sigma}_D(\beta,\mathsf{L},\boldsymbol{\vartheta},\boldsymbol{\varrho},\boldsymbol{\chi})$

$$\widetilde{\Sigma}_D(\beta):=\left\{b\in\Sigma:\ \|P_t(x,\cdot)-\mu_b\|_{\mathrm{TV}}\leq\boldsymbol{\varrho}V(x)\mathrm{e}^{-\boldsymbol{\vartheta}t}\text{ with }\mu_b(V)\leq\boldsymbol{\chi}\text{ and }\rho_b\in\mathcal{H}_D(\beta,\mathsf{L})\right\}.\tag{5.8}$$

Note that for diffusions $\boldsymbol{X}$ with drift $b\in\widetilde{\Sigma}_D(\beta)$ it holds that $\boldsymbol{X}$ is exponentially ergodic, i.e., the total variation distance between the marginal laws of $\boldsymbol{X}$ and the invariant distribution decreases exponentially fast in time, and $\boldsymbol{X}$ is exponentially β-mixing with mixing coefficient $\beta(t)\leq\boldsymbol{\varrho}\boldsymbol{\chi}\mathrm{e}^{-\boldsymbol{\vartheta}t}$, which is independent of b. As demonstrated in Proposition 5.2 below, exponential ergodicity and the exponential β-mixing property are satisfied for any $\boldsymbol{X}$ such that the coefficients $b\in\Sigma$ and σ are globally Lipschitz. Apart from the assumption on the Hölder continuity of ρ_b, restricting the class Σ to $\widetilde{\Sigma}_D$ should therefore be understood as a technical device to obtain *uniform* control on the coefficients in the exponential β-mixing bound, which is needed for the upper bound in the minimax sense provided in Theorem 5.3. Let us briefly demonstrate that our coefficient assumptions guarantee that $\boldsymbol{X}$ fits into the setting of Chapter 3 by verifying that $\boldsymbol{X}$ is exponentially β-mixing and that assumptions $(\mathscr{A}1)$ and $(\mathscr{A}2)$ are satisfied. To this end, let us first remark that in our model, every point $x\in\mathbb{R}$ is reachable in finite time almost surely (cf. [106, Proposition 1.15]), i.e., for any initial distribution η, $\mathbb{P}^\eta(\tau_x<\infty)=1$, where $\tau_x=\inf\{t\geq 0: X_t=x\}$. Thus, it is clear that $\boldsymbol{X}$ is Harris recurrent wrt. the Lebesgue measure $\boldsymbol{\lambda}$.

LEMMA 5.1. *Suppose that $b\in\Sigma$. Then, $\boldsymbol{X}$ possesses transition densities $(p_t)_{t\geq 0}$ such that $p_t(x,y)>0$ for any $x,y\in\mathbb{R}^d$ and there exists some constant $c>0$ such that for any $t\in(0,1]$, we have*

$$\sup_{x,y\in\mathbb{R}} p_t(x,y)\leq ct^{-1/2.}$$

In particular, $(\mathscr{A}1)$ holds.

Proof. Let $f(x)=\int_0^x 1/\sigma(y)\,\mathrm{d}y$, $x\in\mathbb{R}$. Our assumptions on σ guarantee that f is strictly monotone and finite at any fixed point $x\in\mathbb{R}$ and thus has a continuous, strictly monotone inverse f^{-1}. Moreover, since $\sigma\in\mathcal{C}_b^1$, $f\in\mathcal{C}^2(\mathbb{R})$ and thus Itō's formula yields that $\widetilde{X}_t=f(X_t)$, $t\geq 0$, is a strong solution to the SDE

$$\mathrm{d}\widetilde{X}_t=\left(\frac{b}{\sigma}-\frac{1}{2}\sigma'\right)\circ f^{-1}(\widetilde{X}_t)\,\mathrm{d}t+\mathrm{d}W_t,\quad t\geq 0.$$

By the linear growth condition on b, boundedness of σ' and uniform ellipticity of σ—which also implies that $|f^{-1}(x)| \leq |x|/\underline{\nu}$—it follows for the drift $\widetilde{b} = (b/\sigma - \sigma'/2) \circ f^{-1}$ of the SDE that

$$|\widetilde{b}(x)| \lesssim 1 + |f^{-1}(x)| \lesssim 1 + |x|,$$

that is, $\widetilde{b}$ satisfies a linear growth condition. Since additionally the SDE corresponding to $\widetilde{\boldsymbol{X}}$ has unit diffusion coefficient, Theorem 3.1 and Theorem 3.2 in [140] yield that $\widetilde{\boldsymbol{X}}$ possesses transition densities $(\widetilde{p}_t)_{t\geq 0}$ such that $\widetilde{p}_t(x,y) > 0$ for any $x, y \in \mathbb{R}$ and

$$\sup_{x,y\in\mathbb{R}} \widetilde{p}_t(x,y) \leq ct^{-1/2}, \quad t \in (0,1],$$

for some constant $c > 0$. Consequently, the assertion follows from the representation $p_t(x,y) = \frac{1}{\sigma(y)}\widetilde{p}_t(f(x), f(y))$, $x, y \in \mathbb{R}$, and boundedness of $1/\sigma$ by uniform ellipticity. ∎

Next, we turn to exponential ergodicty/mixing of the diffusion process when the drift condition from $\boldsymbol{\Sigma}$ is satisfied. This is a classical result obtained in [167] in any dimension, which was refined in [169] and [121, 168] to polynomial and subexponential ergodicity, resp., by a relaxation of the drift condition. For the sake of completeness we give a full and more compact proof embedded into the general Meyn and Tweedie theory for stability of Markov processes presented in Chapter 2.

Proposition 5.2. *Suppose that $b \in \boldsymbol{\Sigma}$. Then, $\boldsymbol{X}$ is exponentially ergodic, i.e., there exist constants $\varrho, \vartheta > 0$ such that*

$$\|P_t(x,\cdot) - \mu\|_{\mathrm{TV}} \leq \varrho V(x)\mathrm{e}^{-\vartheta t}, \quad t \geq 0, x \in \mathbb{R}, \tag{5.9}$$

where $V \geq 1$ is a $\mathcal{C}^2$-function, which is equal to $\exp(\gamma|x|)$ for $|x| > A$. Moreover, $\boldsymbol{X}$ is exponentially β-mixing and ($\mathscr{A}2$) holds true with

$$r_{\mathcal{S}}(t) := \varrho\|p_1\|_\infty \sup_{x\in\mathcal{S}} V(x)\mathrm{e}^{-\vartheta(t-1)}, \quad t > 1. \tag{5.10}$$

Proof. Denote by $\mathcal{A}$ the extended generator of $\boldsymbol{X}$ with domain $\mathcal{D}(\mathcal{A})$. By Itō's formula, $\mathcal{C}^2(\mathbb{R}) \subset \mathcal{D}(\mathcal{A})$ and for any $f \in \mathcal{C}^2(\mathbb{R})$,

$$\mathcal{A}f = bf' + \frac{1}{2}\sigma^2 f'' = \frac{1}{2\rho}(\sigma^2\rho f')'. \tag{5.11}$$

Let $V \geq 1$ be a function as in the statement of the proposition. Using the definition of $\boldsymbol{\Sigma}$, it follows for $|x| > A$ that

$$\mathcal{A}V(x) = \gamma\sigma^2(x)V(x)\Big(\operatorname{sgn}(x)\frac{b(x)}{\sigma^2(x)} + \frac{\gamma}{2}\Big) \leq -\frac{\gamma^2\bar{\nu}^2}{2}V(x).$$

Thus, V satisfies the Lyapunov-type inequality

$$\mathcal{A}V(x) \leq -\frac{\gamma^2\bar{\nu}^2}{2}V(x) + \zeta, \quad x \in \mathbb{R}, \tag{5.12}$$

where $\zeta = \sup_{x\in[-A,A]}\big(|b(x)V'(x)| + \frac{1}{2}|\sigma^2(x)V''(x)|\big) < \infty$. Since $\boldsymbol{X}$ is irreducible by existence of a stationary distribution and also a T-process by open set irreducibility implied by Lemma 5.1 it follows from [164, Theorem 5.1]that any compact set is petite and, hence, V is unbounded off petite sets. Moreover, for any $\Delta > 0$, the skeleton chain $\boldsymbol{X}^\Delta = (X_{n\Delta})_{n\in\mathbb{N}_0}$ is irreducible, which follows again from $p_t(x,y) > 0$ for all $t \geq 0$ and $x, y \in \mathbb{R}$ by Lemma 5.1, implying that any skeleton is $\boldsymbol{\lambda}$-irreducible. Existence of an irreducible skeleton chain and Harris recurrence give aperiodicity of $\boldsymbol{X}$ as defined in [74] and, hence, we conclude from [74, Theorem 5.2] that (5.12) indeed implies exponential ergodicity in the sense of (5.9). This also gives that the stationary process $\boldsymbol{X}$ is exponentially β-mixing (recall from Chapter 3 that V can always be chosen such that this is implied by exponential ergodicity, or alternatively, use Proposition 2.11) and ($\mathscr{A}2$) with the representation (5.10) follows from Lemma 3.4. ∎

With the above results, it follows from Theorem 3.11 that for fixed $b \in \boldsymbol{\Sigma}$ the invariant density ρ_b can be estimated with parametric rate $1/\sqrt{T}$ wrt. sup-norm risk $\|f\|_\infty = \|f\|_{L^\infty(D)}$ on bounded, open sets D since we have access to the optimal variance bounds from Proposition 3.1. In the following, we aim to obtain the related uniform result over the whole class $\widetilde{\boldsymbol{\Sigma}}(\beta+1)$. To this end, it is advantageous to estimate the variance of the kernel estimator with the help of a local time technique taken from [60].

Theorem 5.3 (concentration of invariant density estimators). *Fix some open and bounded set $D \subset \mathbb{R}$, assume that $b \in \widetilde{\boldsymbol{\Sigma}}_D(\beta+1)$, for some $\beta > 0$, and let Q be a Lipschitz-continuous kernel function of order $\lfloor \beta + 1 \rfloor$ with support* $[-1/2, 1/2]$. *Define the estimator*

$$\widehat{\rho}_T(x) := \frac{1}{\sqrt{T}(\log T)^2} \int_0^T Q\left(\frac{\sqrt{T}(x - X_u)}{(\log T)^2}\right) \mathrm{d}u, \quad x \in D. \tag{5.13}$$

Then, for any $p \geq 1$,

$$\sup_{b \in \widetilde{\boldsymbol{\Sigma}}_D(\beta+1)} \left(\mathbb{E}_b^{\mu_b}\left[\|\widehat{\rho}_T - \rho_b\|_\infty^p\right]\right)^{1/p} \in \mathcal{O}\left(\sqrt{\frac{\log T}{T}}\right). \tag{5.14}$$

Proof. Fix $p \in [1, \infty)$ and $b \in \widetilde{\boldsymbol{\Sigma}}(\beta+1)$, and denote $h = h_T := (\log T)^2/\sqrt{T}$, $m_T := p \log T/\boldsymbol{\vartheta}$, $\mathbb{E}_b = \mathbb{E}$, $\rho_b = \rho$

$$\begin{aligned}
\mathcal{G} &:= \left\{Q\left(\frac{x - \cdot}{h}\right) - \mathbb{E}\left[Q\left(\frac{x - X_0}{h}\right)\right] : x \in D \cap \mathbb{Q}\right\}, \\
\mathbb{H}_T(x) &:= \widehat{\rho}_T(x) - \mathbb{E}\left[\widehat{\rho}_T(x)\right] = \frac{1}{\sqrt{T}h} \mathbb{G}_T\left(Q\left(\frac{x - \cdot}{h}\right) - \mathbb{E}\left[Q\left(\frac{x - X_0}{h}\right)\right]\right).
\end{aligned}$$

Given any $b \in \widetilde{\boldsymbol{\Sigma}}_D(\beta+1)$, Proposition 5.2 shows that the associated diffusion process solving (5.3) is exponentially β-mixing. Thus, we may apply Theorem 3.7 for bounding

$$\left(\mathbb{E}\left[\sup_{x\in D} |\mathbb{H}_T(x)|^p\right]\right)^{1/p} = \left(\mathbb{E}\left[\sup_{x\in D\cap\mathbb{Q}} |\mathbb{H}_T(x)|^p\right]\right)^{1/p}. \tag{5.15}$$

Let τ be as in Theorem 3.7, and denote by $L_T(y)$ the local time of $\boldsymbol{X}$ at the point $y \in \mathbb{R}$ up to time $T \geq 0$, fulfilling in particular $\mathbb{E}[L_T(y)] = T\rho(y)\sigma^2(y)$. Using the occupation times formula and Minkowski's integral inequality, one obtains

$$\begin{aligned}\operatorname{Var}\Big(\int_0^\tau Q\Big(\frac{x-X_u}{h}\Big)\,\mathrm{d}u\Big) &= \mathbb{E}\Big[\Big(\int_{\mathbb{R}} Q\Big(\frac{x-y}{h}\Big)\Big(\frac{L_\tau(y)}{\sigma^2(y)}-\tau\rho(y)\Big)\,\mathrm{d}y\Big)^2\Big]\\ &= h^2\,\mathbb{E}\Big[\Big(\int_{\mathbb{R}} Q(v)\Big(\frac{L_\tau(x-hv)}{\sigma^2(x-hv)}-\tau\rho(x-hv)\Big)\,\mathrm{d}v\Big)^2\Big]\\ &\le h^2\int_{\mathbb{R}} Q(v)\,\mathbb{E}\Big[\Big(\frac{L_\tau(x-hv)}{\sigma^2(x-hv)}-\tau\rho(x-hv)\Big)^2\Big]\,\mathrm{d}v\\ &\le h^2\sup_{v\in\operatorname{supp}(Q)}\frac{\operatorname{Var}(L_\tau(x-hv))}{\sigma^4(x-hv)}\\ &\le\; h^2\underline{\nu}^{-4}C_0\tau\sup_{v\in\operatorname{supp}(Q)}\rho(x-hv),\end{aligned}\tag{5.16}$$

where the last estimate follows from Proposition 5.1 in [60] and $C_0>0$ is a constant depending only on the class $\boldsymbol{\Sigma}$. Thus,

$$\sup_{f,g\in\mathcal{G}} d_{\mathbb{G},\tau}(f,g) = \sup_{f,g\in\mathcal{G}}\sqrt{\operatorname{Var}\Big(\frac{1}{\sqrt{\tau}}\int_0^\tau (f-g)(X_s)\,\mathrm{d}s\Big)} \le h\mathfrak{D},\qquad \text{for } \mathfrak{D}:=\underline{\nu}^{-2}\sqrt{C_0\rho^*},$$

such that $\mathcal{N}(\varepsilon,\mathcal{G},d_{\mathbb{G},\tau})=1$ for $\varepsilon\ge h\mathfrak{D}$. Similarly, for any $g\in\mathcal{G}$,

$$\operatorname{Var}\Big(\frac{1}{\sqrt{\tau}}\int_0^\tau g(X_u)\,\mathrm{d}u\Big) \le \sup_{x\in\operatorname{supp}(g)}\frac{\operatorname{Var}(L_\tau(x))}{\tau\sigma^4(x)\rho(x)}\|g\|^2_{L^2(\mu)} \le \mathfrak{D}^2\rho_*^{-1}\|g\|^2_{L^2(\mu)},$$

and hence

$$\mathcal{N}(\varepsilon,\mathcal{G},\|\cdot\|_{\mathbb{G},\tau}) \;\le\; \mathcal{N}(\varepsilon\sqrt{\rho^*}\mathfrak{D}^{-1},\mathcal{G},\|\cdot\|_{L^2(\mu)}).\tag{5.17}$$

Under the given assumptions on K, $\mathcal{G}$ is a countable class of real-valued functions fulfilling the entropy bound

$$\mathcal{N}(\varepsilon,\mathcal{G},\|\cdot\|_{L^2(\mathbf{Q})}) \le (\mathcal{A}/\varepsilon)^v,$$

for some constants $\mathcal{A}\in(\mathrm{e}^2,\infty)$, $v>0$ and any probability measure $\mathbf{Q}$, cf. Proposition 3.6.12 in [86]. With (5.17) it follows that

$$\mathcal{N}(\varepsilon,\mathcal{G},\|\cdot\|_{\mathbb{G},\tau}) \le \Big(\frac{\mathcal{A}\mathfrak{D}}{\sqrt{\rho^*}\varepsilon}\Big).$$

Consequently, as in the proof of Theorem 3.11, we obtain for $h\le \mathrm{e}^{-2}\mathcal{A}/\sqrt{\rho^*}$ that

$$\int_0^\infty\sqrt{\log\mathcal{N}(\varepsilon,\mathcal{G},d_{\mathbb{G},\tau})}\,\mathrm{d}\varepsilon \le \int_0^{h\mathfrak{D}}\sqrt{v\log\Big(\frac{\mathcal{A}\mathfrak{D}}{\sqrt{\rho_*}\varepsilon}\Big)}\,\mathrm{d}\varepsilon \;\le\; 4\mathfrak{D}h\sqrt{v\log\Big(\frac{\mathcal{A}}{h\sqrt{\rho_*}}\Big)},$$

and Lemma 3.19 implies that

$$\int_0^\infty \log\mathcal{N}\big(\varepsilon,\mathcal{G},\tfrac{2m_T}{\sqrt{T}}d_\infty\big)\,\mathrm{d}\varepsilon \le \frac{8m_T}{\sqrt{T}}\|Q\|_\infty\Big(1+\log\Big(\frac{L\operatorname{diam}(D)}{\|Q\|_\infty h}\Big)\Big).$$

Thus, Theorem 3.7 gives that (5.15) is upper-bounded by

$$\begin{aligned}\frac{1}{\sqrt{T}h}\Bigg(8C_1\frac{m_T}{\sqrt{T}}\|Q\|_\infty\left(1+\log\left(\frac{\mathbb{L}\mathrm{diam}(D)}{\|Q\|_\infty h}\right)\right)+4C_2\mathfrak{D}h\sqrt{v\log\left(\frac{\mathfrak{A}}{h\sqrt{\rho_*}}\right)}\\ +\frac{16m_T}{\sqrt{T}}\|Q\|_\infty c_1 p+2h\mathfrak{D}c_2\sqrt{p}+4\|Q\|_\infty\varrho\chi\sqrt{T}\mathrm{e}^{-\frac{\vartheta m_T}{p}}\Bigg)\in\mathcal{O}\left(\sqrt{\tfrac{\log T}{T}}\right),\end{aligned}\tag{5.18}$$

where the last implication follows from the choice of $h = h_T$ and m_T. The conditions on the order of the kernel function Q and the fact that $\beta > 0$ further imply that, for any $x \in D$,

$$|\mathbb{E}[\widehat{\rho}_T(x)]-\rho(x)| = |(\rho * Q_h-\rho)(x)| \lesssim h^{\beta+1}\in\mathcal{O}\left(\sqrt{\tfrac{\log T}{T}}\right).$$

In combination with the upper bound on the stochastic error stated in (5.18), we thus obtain (5.14). ■

Although assuming stationarity of the process is standard in the statistical literature, this assumption can be slightly problematic for practical purposes. In the present scenario this will become evident for our proof technique of the data-driven control strategy, where we require sup-norm bounds under $\mathbb{P}_b^0$ instead of $\mathbb{P}_b^{\mu_b}$. To extend the rate result (5.14) from the stationary regime to the non-stationary case, we use the following auxiliary result, which shows that exponential convergence allows to exactly quantify the loss imposed by non-stationarity for nonparametric estimation. We focus on the case $p = 1$ as the relevant result for our purposes.

LEMMA 5.4. *Let $\mathcal{X}$ be a topological space and $(\mathbf{Y},(\mathbb{P}^x)_{x\in\mathcal{X}})$ an $\mathcal{X}$-valued exponentially ergodic Markov process, i.e., there exist a function $V\colon\mathcal{X}\to[1,\infty)$ and constants $c,\kappa>0$ such that, for any $x\in\mathcal{X}$,*

$$\left\|\mathbb{P}^x(Y_t\in\cdot)-\mu\right\|_{\mathrm{TV}}\le cV(x)\mathrm{e}^{-\kappa t},\quad t\ge 0,$$

where μ is the invariant distribution of $\mathbf{Y}$. Then, for any bounded $g\in\mathcal{B}(\mathcal{X}^2)$, $x\in\mathcal{X}$ and T large enough such that $T>\kappa^{-1}\log T$, it holds that

$$\begin{aligned}&\left|\mathbb{E}^x\left[\sup_{y\in\mathcal{X}}\left|\frac{1}{T}\int_0^T g(y,Y_s)\,\mathrm{d}s\right|\right]-\mathbb{E}^\mu\left[\sup_{y\in\mathcal{X}}\left|\frac{1}{T}\int_0^T g(y,Y_s)\,\mathrm{d}s\right|\right]\right|\\ &\quad\le\|g\|_\infty\left(\frac{2\log T}{\kappa T}+cV(x)\frac{1}{T}\right).\end{aligned}$$

Proof. Let

$$\theta(y,u,v)=\int_u^v g(y,X_s)\,\mathrm{d}s,\quad 0\le u\le v, y\in\mathcal{X},$$

and

$$\varphi(x):=\mathbb{E}^x\left[\|\theta(\cdot,0,T-\kappa^{-1}\log T)\|_\infty\right],\quad x\in\mathcal{X}.$$

Then,

$$\begin{aligned}&\left|\mathbb{E}^x\left[\sup_{y\in\mathcal{X}}\left|\frac{1}{T}\int_0^T g(y,Y_s)\,\mathrm{d}s\right|\right]-\mathbb{E}^\mu\left[\sup_{y\in\mathcal{X}}\left|\frac{1}{T}\int_0^T g(y,Y_s)\,\mathrm{d}s\right|\right]\right|\\ &\le\frac{1}{T}\left(\mathbb{E}^x\left[\left|\|\theta(\cdot,0,T)\|_\infty-\|\theta(\cdot,\kappa^{-1}\log T,T)\|_\infty\right|\right]+\mathbb{E}^\mu\left[\left|\|\theta(\cdot,0,T)\|_\infty-\|\theta(\cdot,\kappa^{-1}\log T,T)\|_\infty\right|\right]\right)\end{aligned}$$

$$
\begin{aligned}
&+ \frac{1}{T}\Big|\mathbb{E}^x\big[\|\theta(\cdot,\kappa^{-1}\log T, T)\|_\infty\big] - \mathbb{E}^\mu\big[\|\theta(\cdot,\kappa^{-1}\log T, T)\|_\infty\big]\Big| \\
&\leq \frac{1}{T}\Big(\mathbb{E}^x\big[\|\theta(\cdot,0,\kappa^{-1}\log T)\|_\infty\big] + \mathbb{E}^\mu\big[\|\theta(\cdot,0,\kappa^{-1}\log T)\|_\infty\big]\Big) \\
&\quad + \frac{1}{T}\Big|\mathbb{E}^x\big[\|\theta(\cdot,\kappa^{-1}\log T, T)\|_\infty\big] - \mathbb{E}^\mu\big[\|\theta(\cdot,\kappa^{-1}\log T, T)\|_\infty\big]\Big| \\
&\leq 2\|g\|_\infty \frac{\log T}{\kappa T} + \frac{1}{T}\Big|\mathbb{E}^x\big[\|\theta(\cdot,\kappa^{-1}\log T, T)\|_\infty\big] - \mathbb{E}^\mu\big[\|\theta(\cdot,\kappa^{-1}\log T, T)\|_\infty\big]\Big| \\
&= 2\|g\|_\infty \frac{\log T}{\kappa T} + \frac{1}{T}\Big|\mathbb{E}^x\big[\varphi(Y_{\kappa^{-1}\log T})\big] - \mu(\varphi)\Big|,
\end{aligned}
$$

where for the second inequality we used reverse triangle inequality for the first two summands and the last equality is a consequence of the Markov property of $\boldsymbol{Y}$ and stationarity of $\boldsymbol{Y}$ under $\mathbb{P}^\mu$. Using that $\|\varphi\|_\infty \leq \|g\|_\infty T$, exponential ergodicity of $\boldsymbol{Y}$ yields

$$
\begin{aligned}
&\Big|\mathbb{E}^x\Big[\sup_{y\in\mathcal{X}}\Big|\frac{1}{T}\int_0^T g(y,Y_s)\,\mathrm{d}s\Big|\Big] - \mathbb{E}^\mu\Big[\sup_{y\in\mathcal{X}}\Big|\frac{1}{T}\int_0^T g(y,Y_s)\,\mathrm{d}s\Big|\Big]\Big| \\
&\leq 2\|g\|_\infty \frac{\log T}{\kappa T} + cV(x)\|g\|_\infty \mathrm{e}^{-\kappa(\kappa^{-1}\log T)} \\
&= 2\|g\|_\infty \frac{\log T}{\kappa T} + cV(x)\|g\|_\infty \frac{1}{T},
\end{aligned}
$$

as claimed. ■

With this at hand, we obtain a non-stationary version of Theorem 5.3.

Corollary 5.5. *Given the assumptions from Theorem 5.3, it holds for any $x \in \mathbb{R}$ that*

$$
\sup_{b\in\widetilde{\Sigma}_D(\beta+1)} \mathbb{E}_b^x[\|\widehat{\rho}_T - \rho_b\|_\infty] \in \mathrm{O}\Big(\sqrt{\tfrac{\log T}{T}}\Big). \tag{5.19}
$$

Proof. Let

$$
g(x,y) := \frac{\sqrt{T}}{(\log T)^2} Q\Big(\frac{\sqrt{T}(x-y)}{(\log T)^2}\Big) - \rho_b(x), \quad x,y \in \mathbb{R}.
$$

Then, for T large enough such that $\sqrt{T}/(\log T)^2\|Q\|_\infty \geq \rho^*$, we have

$$
\|g\|_\infty \leq 2\sqrt{T}/(\log T)^2 \|Q\|_\infty.
$$

Applying Lemma 5.4 to g and $\boldsymbol{X}$, which is exponentially ergodic by construction of $\widetilde{\Sigma}_D(\beta+1)$, we obtain, for any $x \in \mathbb{R}$ and T large enough such that $T > \boldsymbol{\vartheta}^{-1}\log T$, that

$$
\big|\mathbb{E}_b^x[\|\widehat{\rho}_T - \rho_b\|_\infty] - \mathbb{E}_b^{\mu_b}[\|\widehat{\rho}_T - \rho_b\|_\infty]\big| \leq 2\|Q\|_\infty \Big(\frac{2}{\boldsymbol{\vartheta}\sqrt{T}\log T} + \frac{\boldsymbol{\varrho} V(x)}{\sqrt{T}(\log T)^2}\Big).
$$

Combining this with (5.14), we obtain (5.19) by triangle inequality. ■

5.2.2 *Application*

We now turn to analyzing the singular control problem. Given the literature on related problems, it is natural to expect reflecting barrier strategies, i.e., strategies which maintain the process between two constant thresholds ξ and θ and just grow at these points, to be optimal. We denote such strategies for upper and lower, resp., boundaries $\xi, \theta \in \mathbb{R}$, $\xi < \theta$, by $\boldsymbol{U}^\xi$ and $\boldsymbol{D}^\theta$ and refer to [84, §23] for an explicit characterization, which underlines the interpretation of these processes as local times at the boundaries ξ and θ, resp., of the process $\boldsymbol{X}^Z$.

Solution for known drift

Using classical ergodic results for one-dimensional linear diffusions, it is straightforward to show the following analytic expression for the expected costs when applying reflecting barrier strategies. We refer to [11, Proposition 2.1] for a detailed proof.

LEMMA 5.6. *Let $\xi, \theta \in \mathbb{R}$, $\xi < \theta$ and $x \in [\xi, \theta]$. Then, for $\boldsymbol{Z} = (\boldsymbol{U}^\xi, \boldsymbol{D}^\theta)$,*

$$\lim_{T\to\infty} \frac{1}{T}\mathbb{E}^x\left[\int_0^T c(X_s^Z)\,\mathrm{d}s + q_u U_T^\xi + q_d D_T^\theta\right] = C(\xi,\theta),$$

with

$$C(\xi,\theta) = \frac{1}{\int_\xi^\theta m(x)\,\mathrm{d}x}\left(\int_\xi^\theta c(x)m(x)\,\mathrm{d}x + \frac{q_u}{S'(\xi)} + \frac{q_d}{S'(\theta)}\right),$$

where m denotes the speed density and S the scale function of the underlying diffusion.

For our later purposes, the main observation is that—given the volatility σ—the expected cost function C can completely be described in terms of the invariant density ρ of the underlying diffusion. Indeed:

$$C(\xi,\theta) = \frac{1}{\int_\xi^\theta \rho(x)\,\mathrm{d}x}\left(\int_\xi^\theta c(x)\rho(x)\,\mathrm{d}x + \frac{q_u\sigma^2(\xi)}{2}\rho(\xi) + \frac{q_d\sigma^2(\theta)}{2}\rho(\theta)\right).$$

Therefore, minimizers of C correspond to optimizers of (5.7) in the class of reflecting barrier strategies. The next natural question is whether such minimizers are indeed optimal within the class of all admissible strategies, i.e., whether the minimal value in (5.7) is equal to

$$C^* := \min_{(\xi,\theta)} C(\xi,\theta).$$

This also holds under natural assumptions as can be proven, e.g., adapting the lines of argument in [49, 117] to the two-sided case. We, however, do not go into detail here, but restrict our attention to the class of reflecting barrier strategies in the following.

As 0 is our target state, it is furthermore natural that 0 is contained in the no-action-region which is assumed to be bounded. More precisely, we assume that there exists $B > 0$ such that the minimizer (ξ^*, θ^*) of C fulfill

$$(\xi^*, \theta^*) \in K_B := \{(\xi,\theta) : -B \le \xi \le -1/B,\ 1/B \le \theta \le B\}.$$

In [11], a natural set of assumptions is introduced to guarantee that (ξ^*, θ^*) is characterized as the unique critical point of the function C. We, however, do not need uniqueness for our purposes.

Construction of the estimators

We proceed by constructing estimators $\widehat{\xi}_T$ and $\widehat{\theta}_T$ for the optimal thresholds ξ^* and θ^* which are based on the estimator $\widehat{\rho}_T$ of the invariant density $\rho = \rho_b$ (see (5.13)). To this end, we fix some $\beta > 0$, set $D = K_B$, and write $\widetilde{\Sigma} := \widetilde{\Sigma}_D(\beta + 1)$. In principle, we just use the plug-in estimator, taking however into account that (cf. (5.6))

$$a := \inf_{b \in \widetilde{\Sigma}} \min_{x \in K_B} \rho_b(x) > 0.$$

This leads to the estimator

$$\widehat{C}_T(\xi, \theta) := \frac{1}{\int_\xi^\theta \widehat{\rho}_T(x) \vee a \, \mathrm{d}x} \left(\int_\xi^\theta c(x) \widehat{\rho}_T(x) \, \mathrm{d}x + \frac{q_u \sigma^2(\xi)}{2} \widehat{\rho}_T(\xi) + \frac{q_d \sigma^2(\theta)}{2} \widehat{\rho}_T(\theta) \right)$$

for the expected value $C(\xi, \theta)$ of a reflection strategy with barriers ξ, θ, yielding

$$(\widehat{\xi}_T, \widehat{\theta}_T) \in \arg \min_{(\xi,\theta) \in K_B} \widehat{C}_T(\xi, \theta)$$

as our estimator for the optimal thresholds. Using this, we obtain that the expected costs, when using the strategy based on the estimator after having observed the uncontrolled process for T time units, converge to the optimal value with rate $\sqrt{\log T / T}$.

Proposition 5.7. *For any $x \in \mathbb{R}$, there exists $C_1 > 0$ such that*

$$\sup_{b \in \widetilde{\Sigma}} \mathbb{E}_b^x \left[C(\widehat{\xi}_T, \widehat{\theta}_T) - C^* \right] \leq C_1 \sqrt{\frac{\log T}{T}}.$$

Proof. It holds that

$$\begin{aligned} C(\widehat{\xi}_T, \widehat{\theta}_T) - C^* &= C(\widehat{\xi}_T, \widehat{\theta}_T) - \widehat{C}_T(\widehat{\xi}_T, \widehat{\theta}_T) + \widehat{C}_T(\widehat{\xi}_T, \widehat{\theta}_T) - \min_{(\xi,\theta) \in K_B} C(\xi, \theta) \\ &= C(\widehat{\xi}_T, \widehat{\theta}_T) - \widehat{C}_T(\widehat{\xi}_T, \widehat{\theta}_T) + \min_{(\xi,\theta) \in K_B} \widehat{C}_T(\xi, \theta) - \min_{(\xi,\theta) \in K_B} C(\xi, \theta) \\ &\leq 2 \max_{(\xi,\theta) \in K_B} \left| C(\xi, \theta) - \widehat{C}_T(\xi, \theta) \right|. \end{aligned}$$

To analyze this quantity, we denote numerator and denominator of C and $\widehat{C}_T$ by A_ρ, B_ρ and $A_{\widehat{\rho}_T}, B_{\widehat{\rho}_T}$, resp., and obtain for all $(\xi, \theta) \in K_B$

$$\begin{aligned} \left| C(\xi, \theta) - \widehat{C}_T(\xi, \theta) \right| &\leq \left| \frac{A_\rho(\xi, \theta) - A_{\widehat{\rho}_T}(\xi, \theta)}{B_\rho(\xi, \theta)} \right| + \left| \frac{A_{\widehat{\rho}_T}(\xi, \theta)}{B_\rho(\xi, \theta)} - \frac{A_{\widehat{\rho}_T}(\xi, \theta)}{B_{\widehat{\rho}_T}(\xi, \theta)} \right| \\ &\leq \frac{B}{2a} \left| A_\rho(\xi, \theta) - A_{\widehat{\rho}_T}(\xi, \theta) \right| + |A_{\widehat{\rho}_T}(\xi, \theta)| \left| \frac{1}{B_\rho(\xi, \theta)} - \frac{1}{B_{\widehat{\rho}_T}(\xi, \theta)} \right| \end{aligned}$$

Now, (5.5) yields that we find an absolute constant $\mathfrak{M}$ such that

$$\sup_{b \in \widetilde{\Sigma}} \mathbb{E}_b^x \left[C(\widehat{\xi}_T, \widehat{\theta}_T) - C^* \right] \leq \mathfrak{M} \sup_{b \in \widetilde{\Sigma}} \mathbb{E}_b^x [\| \widehat{\rho}_T - \rho_b \|_\infty],$$

proving the claim by Corollary 5.5. ∎

Data-driven singular controls

In most real world applications, the decision maker is faced with the problem of collecting data about the underlying dynamics and finding the optimal strategy at the same time. Here, however, a classical trade-off between exploration and exploitation occurs. On the one hand, the decision maker wants to minimize her expected costs and therefore uses singular control strategies with an optimal estimated threshold. On the other hand, using such a greedy strategy all the time, the decision maker can't learn about the drift b of the underlying process outside the estimated control interval and therefore this procedure cannot even be expected to converge.

Our solution is to separate exploration and exploitation periods as follows (see Figure 5.1): At the beginning of every period except the first, the process is in the target state 0. In the exploration periods, we then let the process run uncontrolled and the period ends when the process again reaches 0 after having visited two predefined boundaries ξ_0, θ_0, $\xi_0 < 0 < \theta_0$.

In the exploitation periods, we use an estimator for ρ as defined in the previous section in order to choose suitable thresholds based on the observations. The exact specification for this estimator $(\widehat{\xi}_T, \widehat{\theta}_T)$ is given below. An exploitation period ends after the process has been reflected at both the upper and lower estimated boundary and has returned to 0. In the following, we will always set $\xi_0 = -B = -\theta_0$.

We combine exploration and exploitation periods using a (deterministic) sequence $(c_n)_{n\in\mathbb{N}} \in \{0,1\}^{\mathbb{N}}$, where $c_n = 0$ (and $c_n = 1$) means that the n-th period is of exploration-type (and exploitation-type, resp.) and denote the corresponding strategy by $\widehat{\boldsymbol{Z}} = (\widehat{\boldsymbol{U}}, \widehat{\boldsymbol{D}})$. By $\tau_0 = 0 < \tau_1 < \tau_2 < \ldots$ we denote the stopping times separating the periods defining $\widehat{\boldsymbol{Z}}$. The question now is how to balance the time spent for exploration and exploitation. A suitable choice can be made by taking into account the estimation error bounds from the previous section and balancing the errors from misspecifying $(\widehat{\xi}_T, \widehat{\theta}_T)$ due to the estimation error and the losses due to the lack of control in the exploration periods. As we will see below, a suitable choice are sequences $(c_n)_{n\in\mathbb{N}}$ such that there exists $\mathfrak{d} > 0$ with

$$n^{2/3} \leq \#\{i \leq n : c_i = 0\} \leq n^{2/3} + \mathfrak{d}. \tag{5.20}$$

Observe that for such a sequence there exists $\overline{M} > 0$ such that

$$\#\{i \leq n : c_i = 0\} \leq \overline{M} n^{2/3}. \tag{5.21}$$

Note that $\boldsymbol{W}$ is a Brownian motion for the filtration generated by $\boldsymbol{W}$ and the independent random variable X_0. With respect to this filtration, the times separating the different periods are stopping times. Therefore, the process $\widetilde{\boldsymbol{W}}$ which is constructed by putting together the paths of $\boldsymbol{W}$ in the exploration periods, is again a Brownian motion. As the process $\widetilde{\boldsymbol{X}}$ which is constructed by joining the paths of $\boldsymbol{X}$ in the exploration periods fulfills $\widetilde{X}_0 = X_0$ and solves the SDE

$$\mathrm{d}\widetilde{X}_s \;=\; b(\widetilde{X}_s)\,\mathrm{d}s + \sigma(\widetilde{X}_s)\,\mathrm{d}\widetilde{W}_s, \quad s \geq 0,$$

it has the same dynamics as the uncontrolled process $\boldsymbol{X}$.

We denote the estimator for the optimal threshold from Section 5.2.2 for the uncontrolled process $\widetilde{\boldsymbol{X}}$ until time s by $(\widetilde{\xi}_s, \widetilde{\theta}_s)$ and define $(\widehat{\xi}_T, \widehat{\theta}_T) := (\widetilde{\xi}_{S_T \wedge mT^{2/3}}, \widetilde{\theta}_{S_T \wedge mT^{2/3}})$, where S_T denotes the time that the controlled process $\boldsymbol{X}^{\widehat{\boldsymbol{Z}}}$ has spent in the exploration periods until T, and m is a constant specified in the following lemma. In other words, we base the estimator $(\widehat{\xi}_T, \widehat{\theta}_T)$ for

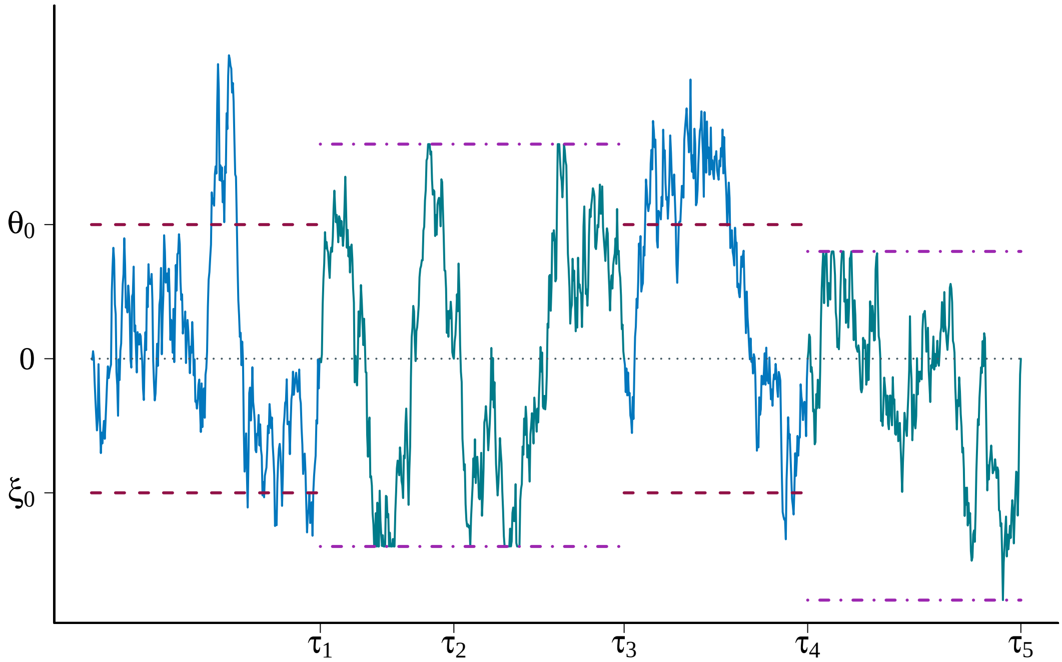

Figure 5.1: A path controlled using a data-driven reflection strategy with exploration (blue) and exploitation (turquoise) periods using $(c_n)_n = (0, 1, 1, 0, 1, \dots)$. The predefined boundaries ξ_0, θ_0 determining the length of the exploration periods are represented by red lines and the estimated optimal reflection boundaries by purple lines.

the threshold used in the exploitation periods just on the observations in the exploration periods and, in addition, just for technical reasons, ignore all observations after time $s = mT^{2/3}$.

We first observe that condition (5.20) implies that the time S_T spent in the exploration periods until time T is of order $T^{2/3}$. In particular, $S_T \to \infty$ and $S_T/T \to 0$. More precisely:

Lemma 5.8. *Let $(c_n)_{n\in\mathbb{N}} \in \{0,1\}^{\mathbb{N}}$ satisfy (5.20) with corresponding data-driven strategy $\widehat{\boldsymbol{Z}}$ as specified above. Then, there exist $m, M > 0$ such that*

$$\mathbb{P}_b^0(T^{-2/3}S_T \leq m) \lesssim T^{-1/3} \text{ and } \limsup_{T\to\infty} T^{-2/3}\mathbb{E}_b^0[N_T^0] \leq M,$$

where N_T^0 denotes the number of exploration periods until time T.

The proof, which is quite technical and based on renewal theoretic arguments, is deferred to Appendix 5.A. The main result of this section given below shows that, by employing the strategy $\widehat{\boldsymbol{Z}}$, we can guarantee that the expected regret per time unit vanishes with rate $\sqrt{\log T}/T^{1/3}$.

Theorem 5.9. *Let $(c_n)_{n\in\mathbb{N}} \in \{0,1\}^{\mathbb{N}}$ satisfy (5.20) with corresponding data-driven strategy $\widehat{\boldsymbol{Z}}$ as specified above. Then, the expected regret per time unit is of order $\mathrm{O}\left(\frac{\sqrt{\log T}}{T^{1/3}}\right)$. That is, for any $b \in \widetilde{\boldsymbol{\Sigma}}$,* *we have*

$$\limsup_{T\to\infty} \frac{T^{1/3}}{\sqrt{\log T}}\left(\frac{1}{T}\mathbb{E}_b^0\left[\int_0^T c(X_s^{\widehat{Z}})\,\mathrm{d}s + q_u\widehat{U}_T + q_d\widehat{D}_T\right] - C_b^*\right) < \infty.$$

Proof. We first consider the costs in the exploration periods. Using [15, Chapter VI, Theorem 1.2], we first see that in one exploration cycle starting and ending in 0, the expected costs are

$$\mathbb{E}_b^0[\tau_1]\int c(x)\rho_b(x)\,\mathrm{d}x,$$

with finiteness of (arbitrary) moments of τ_1 under $\mathbb{P}_b^0$ being demonstrated in Appendix 5.A. Hence, the expected costs per time unit in full exploration cycles are $\int c(x)\rho_b(x)\,\mathrm{d}x$ and the time spent in such cycles until T is bounded by S_T. If we consider the cumulative costs until time T, we have to take into account that the last exploration cycle may be cut off at the deterministic time T. Putting pieces together, we can bound the expected costs in the exploration period as follows:

$$\begin{aligned}
\mathbb{E}_b^0\left[\int_0^T c(X_t^{\widehat{Z}})\,\mathrm{d}S_t\right] &\leq \mathbb{E}_b^0\left[\sum_{\substack{n:\tau_n\leq T\\ \text{exploration period}}}\int_{\tau_n}^{\tau_{n+1}} c(X_t^{\widehat{Z}})\,\mathrm{d}t\right]\\
&= \sum_{n\in\mathbb{N}_0}\mathbb{E}_b^0\left[\mathbb{E}_b^0\left[\int_{\tau_n}^{\tau_{n+1}} c(X_t^{\widehat{Z}})\,\mathrm{d}t\,\middle|\,\mathcal{F}_{\tau_n}^{\widehat{Z}}\right]\mathbb{1}_{\{\tau_n\leq T,\ \text{exploration starts at }\tau_n\}}\right]\\
&= \mathbb{E}_b^0[N_T^0]\,\mathbb{E}_b^0[\tau_1]\int c(x)\rho_b(x)\,\mathrm{d}x\\
&\lesssim T^{2/3},
\end{aligned}$$

where we applied Lemma 5.8 with N_T^0 denoting the number of exploration periods until time T and $(\mathcal{F}_t^{\widehat{Z}})_{t\geq 0}$ is the filtration generated by the controlled process $\boldsymbol{X}^{\widehat{Z}}$. To analyze the costs in the exploitation periods, we write $R_t := t - S_t$ for the time spent in the exploitation periods and—again using [15, Chapter VI, Theorem 1.2]—similarly get

$$\begin{aligned}
&\mathbb{E}_b^0\left[\int_0^T c(X_t^{\widehat{Z}})\,\mathrm{d}R_t + q_u\widehat{U}_T + q_d\widehat{D}_T\right]\\
&\leq \mathbb{E}_b^0\left[\sum_{\substack{n:\tau_n\leq T\\ \text{exploitation period}}}\left(\int_{\tau_n}^{\tau_{n+1}} c(X_t^{\widehat{Z}})\,\mathrm{d}t + q_u(\widehat{U}_{\tau_{n+1}} - \widehat{U}_{\tau_n}) + q_d(\widehat{D}_{\tau_{n+1}} - \widehat{D}_{\tau_n})\right)\right]\\
&\leq \sum_{n\in\mathbb{N}_0}\mathbb{E}_b^0\left[\mathbb{E}_b^0\left[\int_{\tau_n}^{\tau_{n+1}} c(X_t^{\widehat{Z}})\,\mathrm{d}t + q_u(\widehat{U}_{\tau_{n+1}} - \widehat{U}_{\tau_n}) + q_d(\widehat{D}_{\tau_{n+1}} - \widehat{D}_{\tau_n})\,\middle|\,\mathcal{F}_{\tau_n}^{\widehat{Z}}\right]\mathbb{1}_{\{\tau_n\leq T,\ \text{exploit. starts at }\tau_n\}}\right]\\
&\leq \mathbb{E}_b^0\left[\sum_{n\in\mathbb{N}_0} C(\widehat{\xi}_{\tau_n},\widehat{\theta}_{\tau_n})\,\mathbb{E}_b^0[\tau_{n+1}-\tau_n\,|\,\mathcal{F}_{\tau_n}^{\widehat{Z}}]\,\mathbb{1}_{\{\tau_n\leq T,\ \text{exploitation starts at }\tau_n\}}\right]\\
&= \mathbb{E}_b^0\left[\sum_{n\in\mathbb{N}_0} C(\widehat{\xi}_{\tau_n},\widehat{\theta}_{\tau_n})(\tau_{n+1}-\tau_n)\,\mathbb{1}_{\{\tau_n\leq T,\ \text{exploitation starts at }\tau_n\}}\right]
\end{aligned}$$

$$
\begin{aligned}
&= \mathbb{E}_b^0\left[\sum_{n\in\mathbb{N}_0}\int_{\tau_n}^{\tau_{n+1}} C(\widehat{\xi}_t,\widehat{\theta}_t)\,\mathrm{d}t\mathbb{1}_{\{\tau_n\le T,\text{ exploitation starts at }\tau_n\}}\right]\\
&\le \int_0^T \mathbb{E}_b^0[C(\widehat{\xi}_t,\widehat{\theta}_t)]\,\mathrm{d}t + \max_{(\xi,\theta)\in K_B} C(\xi,\theta)\mathbb{E}_b^0[\bar{\eta}^1],
\end{aligned}
$$

where $\bar{\eta}^1$ denotes the length of an exploitation period with maximal length (i.e., a period with reflection in $\pm B$). On the event $\{S_t \ge mt^{2/3}\}$, we have that $(\widehat{\xi}_t,\widehat{\theta}_t) = (\widetilde{\xi}_{mt^{2/3}},\widetilde{\theta}_{mt^{2/3}})$, so that by Lemma 5.8 and Proposition 5.7, we have

$$
\begin{aligned}
\mathbb{E}_b^0[C(\widehat{\xi}_t,\widehat{\theta}_t)] &\le \max_{(\xi,\theta)\in K_B} C(\xi,\theta)\mathbb{P}_b^0(S_t < mt^{2/3}) + \mathbb{E}_b^0[C(\widehat{\xi}_t,\widehat{\theta}_t)\mathbb{1}_{\{S_t\ge mt^{2/3}\}}]\\
&\le c_1t^{-1/3} + \mathbb{E}_b^0[C(\widetilde{\xi}_{mt^{2/3}},\widetilde{\theta}_{mt^{2/3}})]\\
&\le c_1t^{-1/3} + C_b^* + c_2\sqrt{\frac{\log(mt^{2/3})}{mt^{2/3}}}\\
&\le C_b^* + c_3\frac{\sqrt{\log t}}{t^{1/3}}
\end{aligned}
$$

for certain constants c_1, c_2, c_3, hence

$$
\begin{aligned}
\mathbb{E}_b^0\left[\int_0^T c(X_t^{\widehat{Z}})\,\mathrm{d}R_t + q_u\widehat{U}_T + q_d\widehat{D}_T\right] &\le \int_0^T \mathbb{E}_b^0[C(\widehat{\xi}_t,\widehat{\theta}_t)]\,\mathrm{d}t + \max_{(\xi,\theta)\in K_B} C(\xi,\theta)\mathbb{E}_b^0[\bar{\eta}^1]\\
&\le C_b^*T + c_4\int_0^T \frac{\sqrt{\log t}}{t^{1/3}}\,\mathrm{d}t\\
&\le C_b^*T + c_4\sqrt{\log(T)}\int_0^T t^{-1/3}\,\mathrm{d}t\\
&\le C_b^*T + c_5T^{2/3}\sqrt{\log(T)}
\end{aligned}
$$

for certain constants c_4, c_5. Putting pieces together , we obtain

$$
\begin{aligned}
&\frac{1}{T}\mathbb{E}_b^0\left[\int_0^T c(X_s^{\widehat{Z}})\,\mathrm{d}s + q_u\widehat{U}_T + q_d\widehat{D}_T\right] - C_b^*\\
&= \frac{1}{T}\mathbb{E}_b^0\left[\int_0^T c(X_t^{\widehat{Z}})\,\mathrm{d}S_t\right] + \frac{1}{T}\mathbb{E}_b^0\left[\int_0^T c(X_t^{\widehat{Z}})\,\mathrm{d}R_t + q_u\widehat{U}_T + q_d\widehat{D}_T\right] - C_b^*\\
&\lesssim T^{-1/3} + \frac{\sqrt{\log(T)}}{T^{1/3}}.
\end{aligned}
$$

■

5.3 Data-driven controls for Lévy processes

We now turn our attention to another class of non-continuous control problems. The first main difference is that we consider a one-sided class of problems, that is, we just consider downward controls. Second, we assume the underlying processes to have jumps. More precisely, as our driving process, we take a Lévy process $\boldsymbol{X} = (X_t)_{t\ge 0}$, started in $x \in \mathbb{R}$ under $\mathbb{P}^x$, satisfying the basic assumption

($\mathcal{L}0$) $\boldsymbol{X}$ is upward regular, i.e.,

$$\mathbb{P}^0(\inf\{t \geq 0 : X_t > 0\} = 0) = 1,$$

and moreover $0 < \mathbb{E}^0[X_1] =: \eta < \infty$.

Let us note that any Lévy process with unbounded variation (i.e., either $\boldsymbol{X}$ has a non-trivial Gaussian part or $\int_{-1}^{1} |x| \, \Pi(\mathrm{d}x) = \infty$) satisfies the upward regularity assumption. For a full description of upward regularity in terms of necessary and sufficient conditions, also covering a subset of Lévy processes with bounded variation, see [109, Theorem 6.5].

Control problems with underlying jump processes are known to be much harder to analyze than their counterparts without jumps, see [135] for discussions and many examples. To formulate our problem, we fix a non-decreasing function $\gamma \in \mathcal{C}^2(\mathbb{R})$. In contrast to the singular controls discussed in Section 5.2.2, we now consider controls of impulse-type. These are sequences $S = (\tau_n, \zeta_n)_{n\in\mathbb{N}}$ of stopping times $\tau_1 < \tau_2 < \ldots \nearrow \infty$ and $\mathcal{F}_{\tau_n}$-measurable random variables ζ_n describing the times of the interventions and the state after exercising the control, respectively. The corresponding controlled process is given as

$$X_t^S = X_t - \sum_{n\in\mathbb{N}:\tau_n \leq t} (X_{\tau_n,-}^S - \zeta_n), \quad t \geq 0.$$

Here, the value at time τ_n, but with the control not having taken place yet, is denoted by

$$X_{\tau_n,-}^S = X_{\tau_n} - \sum_{m\in\mathbb{N}:m<n} (X_{\tau_m-}^S - \zeta_m).$$

In general, for processes with jumps, this quantity may deviate from both the value $X_{\tau_n}^S = \zeta_n$ at time τ_n after the control has taken place and the left limit $X_{\tau_n-}^S$. We can interpret $\gamma(\boldsymbol{X}^S)$ as the value of a natural resource we are managing. In most examples of interest, γ has a sigmoidal form, so that (without interventions) the value is expected to grow fast whenever $\boldsymbol{X}^S$ takes moderate values, while the value grows slowly whenever $\boldsymbol{X}^S$ has either large or small values. The stopping times τ_n describe the times of intervention. From the motivating problem, it is clear that we only assume downward controls to be admissible, i.e., we assume that $X_{\tau_n,-}^S \geq \zeta_n$ for all n.

Our aim is to find a maximizer and the corresponding value v of the expected rewards without fixed transaction costs, defined by

$$\liminf_{T\to\infty} \frac{1}{T} \mathbb{E}^x\Big[\sum_{n\in\mathbb{N}:\tau_n \leq T} \big(\gamma(X_{\tau_n,-}^S) - \gamma(\zeta_n)\big)\Big], \tag{5.22}$$

in the class of all admissible impulse control strategies $S = (\tau_n, \zeta_n)$.

We will argue in Section 5.3.2 below that the main tool for solving (5.22) is the ascending ladder height process. We remind the reader that the underlying concepts and main results from fluctuation theory for Lévy processes needed in the following are summarized in Appendix 4.B. Note that ($\mathcal{L}0$) implies $\lim_{t\to\infty} X_t = \infty$ almost surely. Let $H_t = X_{\mathsf{L}_t^{-1}}$, $t \geq 0$, be the ascending ladder height subordinator of $\boldsymbol{X}$, where $\mathsf{L} = (\mathsf{L}_t)_{t\geq 0}$ is a version of local time at the supremum and $\mathsf{L}^{-1} = (\mathsf{L}_t^{-1})_{t\geq 0}$ is its right-continuous inverse. Note that L can be chosen to be continuous by upward regularity of $\boldsymbol{X}$, which entails that $t \mapsto \mathsf{L}_t^{-1}$ is strictly increasing and thus $\boldsymbol{H}$ is a strictly increasing subordinator (or, put differently, is not compound Poisson). Moreover, for any $t \geq 0$, L_t^{-1} is an $\mathbb{F}$-stopping time, where $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$ denotes the usual completed natural filtration of $\boldsymbol{X}$. Motivated by the solution technique for the associated control problem, we choose a scaling of L such that $\mathbb{E}^0[\mathsf{L}_1^{-1}] = 1$ and hence, by Wald's equality (cf. [138, Corollary 2.5.2]),

$$\mathbb{E}^0[H_1] = \mathbb{E}^0[X_1].$$

Moreover, for $T_x := \inf\{t \geq 0 : X_t > x\}$ and $T_x^H := \inf\{t \geq 0 : H_t > x\}$, under ($\mathscr{L}0$) we have

$$T_x = \mathsf{L}^{-1}(\mathsf{L}_{T_x}) = \mathsf{L}^{-1}_{T_x^H}$$

almost surely for any $x \geq 0$ (see Proposition IV.7 and the proof of Theorem VI.19 in [25]). Since $(\mathsf{L}_t^{-1})_{t\geq 0}$ is a subordinator and T_x^H is a $(\mathcal{F}_{\mathsf{L}_t^{-1}})_{t\geq 0}$-stopping time, Wald's equality also yields that

$$\mathbb{E}^0[T_x] = \mathbb{E}^0[L_1^{-1}]\mathbb{E}^0[T_x^H] = \mathbb{E}^0[T_x^H]. \tag{5.23}$$

Let (d_H, Π_H) denote drift and Lévy measure of $\boldsymbol{H}$ and $\mathcal{D}(\mathcal{A}_H)$ be the domain of the extended generator $\mathcal{A}_H$ of $\boldsymbol{H}$. By Itō's formula for semimartingales applied to $\boldsymbol{H}$, see e.g. Theorem I.4.57 in [98], it follows that for $f \in \mathcal{C}^2(\mathbb{R})$ such that

$$g_f(x) = d_H f'(x) + \int_{0+}^{\infty} (f(x+y) - f(x))\, \Pi_H(\mathrm{d}y), \quad x \in \mathbb{R}, \tag{5.24}$$

is well-defined that $(f(H_t) - f(H_0) - \int_0^t g_f(H_s)\,\mathrm{d}s)_{t\geq 0}$ is a local martingale. For such functions $f \in \mathcal{C}^2(\mathbb{R})$ we set $\mathcal{A}_H f = g_f$. We will see in Section 5.3.2 below that the auxiliary function $f(x) = \mathcal{A}_H\gamma(x)$ is the key for the solution to (5.22). More precisely, $\mathcal{A}_H\gamma$ yields a maximum representation of the payoff that is needed to guarantee optimality of a threshold time, which can be derived from $\mathcal{A}_H\gamma$.

5.3.1 *Estimating generator functionals for the ascending ladder height Lévy process*

Motivated by this observation, to implement a data-driven strategy our goal is to find an estimator of $f(x) = \mathcal{A}_H\gamma(x)$ for an appropriate $\gamma \in \mathcal{C}^2(\mathbb{R})$, based on a continuously observed trajectory $(X_t)_{0\leq t\leq T}$ of $\boldsymbol{X}$ up to some fixed time horizon T, with good approximation properties wrt. the sup-norm risk. Estimating $\mathcal{A}_H\gamma$ is therefore of significant applied interest and, as it will turn out, establishing bounds for the sup-norm risk provides the right tool to infer estimates for the expected regret of data-driven control strategies. For our purposes, we will need to assume that γ' is bounded, which is clearly in line with a typically sigmoidal form of γ.

In order to construct an estimator for $\mathcal{A}_H\gamma$, a first intuitive approach would be to assume that Π_H is absolutely continuous with Lebesgue density π_H and reconstruct a path $(H_t)_{0\leq \mathsf{L}_t^{-1}\leq T}$ from the full observations $(X_t)_{0\leq t\leq T}$ to develop a nonparametric estimator $(\widehat{d}_H, \widehat{\pi}_H)$ of (d_H, π_H) and then analyze the plug-in estimator

$$\widehat{\mathcal{A}_H\gamma}(x) = \widehat{d}_H\gamma(x) + \int_{0+}^{\infty} (\gamma(x+y) - \gamma(x))\widehat{\pi}_H(y)\,\mathrm{d}y, \quad x \in \mathbb{R}, \tag{5.25}$$

based on convergence rates of $(\widehat{d}_H, \widehat{\pi}_H)$ as $T \to \infty$. An appropriate estimator for $\widehat{\pi}_H$ in this scenario is given by

$$\widehat{\pi}_H(x) = \frac{1}{\mathsf{L}_T} \sum_{0\leq t\leq \mathsf{L}_T} K_h(x - \Delta H_t)\mathbb{1}_{\{\Delta H_t > 0\}}, \quad x > 0,$$

for $K_h := h^{-1}K(\cdot/h)$, where $h = h(T) > 0$ is some bandwidth and K a high-order kernel function, see [154, 155].

However, under $(\mathscr{L}0)$—even with a full record of $\boldsymbol{X}$—local time L cannot be observed in general since its construction is not purely path dependent, see [25, Chapter 4]. Hence, in our framework such ansatz is hopeless, unless $\boldsymbol{X}$ is assumed to have a one sided jump structure, i.e., $\boldsymbol{X}$ is either a subordinator (increasing paths), spectrally negative (only negative jumps but non-monotone paths) or spectrally positive (only positive jumps but non-monotone paths). In the subordinator case we can simply choose $\mathsf{L}_t = t$ and hence $\boldsymbol{H} = \boldsymbol{X}$, i.e., the problem of estimating the generator functional via the plug-in estimator reduces to estimation of the Lévy measure and drift of $\boldsymbol{X}$. When $\boldsymbol{X}$ is spectrally negative, we can choose $\mathsf{L}_t = c\overline{X}_t$, where $\overline{X}_t = \sup_{0\leq s\leq t} X_s$ and c is some scaling factor. Then, $\mathsf{L}_t^{-1} = T_{t/c}$, where $T_x := \inf\{t \geq 0 : X_t > x\}$ is the first passage time of the level x. By exclusively negative jumps, $\boldsymbol{X}$ reaches its maxima continuously, hence $H_t = X_{T_{t/c}} = t/c$, where c must be chosen such that

$$1 = \mathbb{E}^0[\mathsf{L}_1^{-1}] = \mathbb{E}^0[T_{1/c}], \tag{5.26}$$

by our required scaling of local time. Thus, estimation of the generator functional in this case boils down to estimating the drift $d_H = 1/c$, which would require estimation of expected first passage times for different levels $x > 0$ and solving (5.26) for c with the expectation on the right hand side replaced by the constructed estimators. This is a non-trivial procedure and it is not clear how such issue should be efficiently attacked with a given dataset. For the case of spectrally positive processes, a similar issue would arise for the correct scaling of local time at the infimum.

Thus, the only direct estimation approach other than the one we introduce below, demands restricting to a subordinator. If its Lévy measure is finite (i.e., $\boldsymbol{X}$ must be a compound Poisson subordinator with positive drift since we require $(\mathscr{L}0)$), it follows from Theorem 3.1 in [155] that (ignoring the drift part) the L^2 risk of the estimator (5.25) is of order $1/\sqrt{T}$. At the end of this section, we will argue that even in this much more restricted setting, our estimator—which can be applied for arbitrary jump structures—matches this performance.

Let us therefore now show how to go a more sophisticated route, exploiting the probabilistic structure of the generator functional by making use of the stability results on overshoots of Lévy processes established in Chapter 4. This is in general a very natural approach for statistical inference of objects related to $\boldsymbol{H}$ due to its intimate connections with overshoots of $\boldsymbol{X}$, that we briefly recall in the sequel. Let $\mathcal{O}_x$ be the overshoot of $\boldsymbol{X}$ over the level $x \geq 0$, defined by

$$\mathcal{O}_x := X_{T_x} - x, \quad x \geq 0,$$

where $T_x := \inf\{t \geq 0 : X_t > x\}$ is again the first hitting time of (x, ∞). If we consider the spatial levels that $\boldsymbol{X}$ surpasses along its lifetime as time index, it can be shown that under $(\mathscr{L}0)$ the overshoot process $\boldsymbol{\mathcal{O}} := (\mathcal{O}_t)_{t\geq 0}$ is a Feller Markov process.

Its role in revealing the characteristics of the ascending ladder height process stems from the simple observation that the closure of the range of $\boldsymbol{H}$ almost surely is identical to the range of the running supremum process $\overline{X}_t := \sup_{0\leq s\leq t} X_s$, $t \geq 0$, and hence the overshoot process $\boldsymbol{\mathcal{O}}^H$ of $\boldsymbol{H}$ is indistinguishable from $\boldsymbol{\mathcal{O}}$. As seen in Chapter 4 the unique invariant distribution of $\boldsymbol{\mathcal{O}}$ is given by

$$\mu(\mathrm{d}y) = \frac{1}{\mathbb{E}^0[H_1]}\Big(d_H\delta_0(\mathrm{d}y) + \mathbb{1}_{(0,\infty)}(y)\Pi_H(y,\infty)\,\mathrm{d}y\Big)$$
$$= \frac{1}{\eta}\Big(d_H\delta_0(\mathrm{d}y) + \mathbb{1}_{(0,\infty)}(y)\Pi_H(y,\infty)\,\mathrm{d}y\Big), \quad y \geq 0,$$

with the second equality being a consequence of our particular scaling of local time. If we assume additionally that

($\mathscr{L}1$) either, $d_H > 0$, or there exists $(a,b) \subset (0,\infty)$ such that $\lambda|_{(a,b)} \ll \Pi_H|_{(a,b)}$,

it follows from Proposition 4.39 that, for any $x \in \mathbb{R}_+$,

$$\lim_{t\to\infty} \|\mathbb{P}^x(\mathcal{O}_t \in \cdot) - \mu\|_{\mathrm{TV}} = 0, \tag{5.27}$$

where $\|\cdot\|_{\mathrm{TV}}$ (as before) denotes the total variation distance. In Proposition 4.39, conditions on the characteristics of the parent process $\boldsymbol{X}$ implying ($\mathscr{L}1$) are given. These underline that most explicit Lévy models fall into the total variation convergence scheme provided that upward regularity is satisfied, since these usually either possess a non-trivial Gaussian component or the Lévy measure is constructed from a Lebesgue density. Finally, assuming

($\mathscr{L}2$) there is $\lambda > 0$ such that $\mathbb{E}^0[\exp(\lambda H_1)] < \infty$,

which is true iff $\Pi|_{[1,\infty)}$ integrates $x \mapsto \exp(\lambda x)$, Proposition 4.41 states that total variation convergence in (5.27) takes place at exponential rate and that $\mathcal{O}$ is exponentially β-mixing whenever the initial distribution integrates $\exp(\lambda\cdot)$. In particular, the stationary overshoot process is exponentially β-mixing, with β-mixing coefficient

$$\beta(t) = \int_{\mathbb{R}_+} \|\mathbb{P}^x(\mathcal{O}_t \in \cdot) - \mu\|_{\mathrm{TV}}\,\mu(\mathrm{d}x) \leq C(\lambda,\delta,\mu)\mathrm{e}^{-t/(2+\delta)}, \quad t \geq 0, \tag{5.28}$$

for some constant $C(\lambda,\delta,\mu) \in (0,\infty)$ and arbitrary $\delta \in (0,1)$. Starting from this general setup, the fundamental observation for our purposes is that, for $\gamma \in \mathcal{C}^2(\mathbb{R})$, we can rewrite (5.24) in terms of an integral wrt. the invariant overshoot distribution μ.

Lemma 5.10. *For any $\gamma \in \mathcal{C}^2(\mathbb{R})$ with bounded derivative we have*

$$\mathcal{A}_H\gamma(x) = \int_{\mathbb{R}_+} \eta\gamma'(x+y)\,\mu(\mathrm{d}y), \quad x \in \mathbb{R}.$$

Proof. Note first that $\mathbb{E}^0[H_1] < \infty$ and boundedness of γ' guarantee that both sides of the equation are well defined. Plugging in and using Fubini we obtain for $x \in \mathbb{R}$,

$$\begin{aligned}\int_{\mathbb{R}_+} \eta\gamma'(x+y)\,\mu(\mathrm{d}y) &= d_H\gamma'(x) + \int_{0+}^{\infty} \gamma'(x+y)\int_{y+}^{\infty}\Pi_H(\mathrm{d}z)\,\mathrm{d}y\\ &= d_H\gamma'(x) + \int_{0+}^{\infty}\int_{(0,z)} \gamma'(x+y)\,\mathrm{d}y\,\Pi_H(\mathrm{d}z)\\ &= d_H\gamma'(x) + \int_{0+}^{\infty} (\gamma(x+z)-\gamma(x))\,\Pi_H(\mathrm{d}z)\\ &= \mathcal{A}_H\gamma(x).\end{aligned}$$

■

Remark 5.11. This formula is valid for any subordinator with finite mean.

It follows from von Neumann's ergodic theorem that, for any $x \geq 0$ and $p \geq 1$,

$$\lim_{S\to\infty} \frac{1}{S}\int_0^S \eta\gamma'(x+\mathcal{O}_t)\,\mathrm{d}t = \mathcal{A}_H\gamma(x), \quad \text{in } L^p(\mathbb{P}^\mu).$$

It is therefore natural to consider as an estimator of $f(x) = \mathcal{A}_H\gamma(x)$, based on overshoot observations $(\mathcal{O}_t)_{0\leq t\leq S}$ up to some *spatial* level $S > 0$, the unbiased (under $\mathbb{P}^\mu$) estimator

$$\widetilde{f}_S(x) = \frac{1}{S}\int_0^S \eta\gamma'(x+\mathcal{O}_t)\,\mathrm{d}t, \quad x \in \mathbb{R},$$

with $\eta = \mathbb{E}^0[X_1] > 0$ assumed to be known (which is not a strict assumption in light of i.i.d. increments of $\boldsymbol{X}$). To establish convergence bounds wrt. to the sup-norm risk, we make use of Theorem 3.7. We apply this result to the function class

$$\mathcal{G} := \{\eta\gamma'(x+\cdot) - \mathcal{A}_H\gamma(x) : x \in \mathbb{Q}\cap D\},$$

to find a convergence rate of $1/\sqrt{S}$ for the sup-norm risk

$$\mathcal{R}^D_\infty(\widetilde{f}_S, f) := \mathbb{E}^0\big[\|\widetilde{f}_S - f\|_{L^\infty(D)}\big],$$

for some bounded open set $D \subset \mathbb{R}$. The choice of evaluating the sup-norm risk wrt. $\mathbb{P}^0$ is somewhat arbitrary and can be replaced by $\mathbb{P}^x$ for any $x \geq 0$ by spatial homogeneity of the Lévy process. We stress however that, although we make heavily use of ergodic arguments, we do not need the process to be started in the stationary overshoot distribution for our results. Similar to the proof of Corollary 5.5, the key for this is Lemma 5.4 in conjunction with exponential ergodicity of $\boldsymbol{\mathcal{O}}$.

Let us assume for the rest of the section that $(\mathcal{L}0)$ – $(\mathcal{L}2)$ are satisfied, if not mentioned otherwise.

PROPOSITION 5.12. *Let $\gamma \in \mathcal{C}^2(\mathbb{R})$ such that γ' is bounded. Then there exists a constant $C_1 > 0$ such that*

$$\mathcal{R}^D_\infty(\widetilde{f}_S, f) \leq C_1\frac{1}{\sqrt{S}}.$$

Proof. By stationarity of $\boldsymbol{\mathcal{O}}$ under $\mathbb{P}^\mu$ and its exponential β-mixing property (5.28), which is guaranteed given our assumptions, it follows easily (see, e.g., the proof of Proposition 2.4 in [67]) for any bounded g and $t > 0$ that

$$\begin{aligned}\|g\|^2_{\mathbb{G},t} = \frac{1}{t}\mathrm{Var}\Big(\int_0^t g(\mathcal{O}_s)\,\mathrm{d}s\Big) &\leq 2\|g\|^2_\infty\int_0^t\int_0^\infty \|\mathbb{P}^x(\mathcal{O}_s\in\cdot) - \mu\|_{\mathrm{TV}}\,\mu(\mathrm{d}x)\,\mathrm{d}s\\ &\leq 2\|g\|^2_\infty\varrho(\lambda,\delta,\mu)(2+\delta),\end{aligned}$$

for some $\delta \in (0,1)$. Hence, there exists a constant $\widetilde{C} > 0$ such that, independently of $t > 0$, for any bounded f, g

$$d_{\mathbb{G},t}(f,g) \leq \widetilde{C}d_\infty(f,g). \tag{5.29}$$

Letting $\mathcal{G} := \{\eta\gamma'(x+\cdot) - \mathcal{A}_H\gamma(x) : x \in \mathbb{Q} \cap D\}$ and using the fact that γ' is Lipschitz on the bounded set D thanks to $\gamma \in \mathcal{C}^2(\mathbb{R})$, it follows with Lemma 3.19 that

$$\mathcal{N}(\varepsilon, \mathcal{G}, d_\infty) \leq \frac{4\eta L\mathrm{diam}(D)}{\varepsilon}, \quad \varepsilon > 0,$$

where L denotes the Lipschitz constant of γ' on D. It therefore follows that the associated entropy integral is finite, i.e.,

$$\int_0^\infty \log \mathcal{N}\big(u, \mathcal{G}, d_\infty\big)\,\mathrm{d}u < \infty,$$

and by (5.29) the same is true for the entropy integral

$$\int_0^\infty \sqrt{\log \mathcal{N}\big(u, \mathcal{G}, d_{\mathbb{G},t}\big)}\,\mathrm{d}u < \bar{C},$$

with a constant $\bar{C}$ independent of t. Since $f(x) = \mathcal{A}_H\gamma(x) = \eta\mu(\gamma'(x+\cdot))$, choosing $m_S = \sqrt{S}$ and plugging into (5.2) therefore reveals that there exists a constant $C_0 > 0$ such that

$$\begin{aligned}\mathbb{E}^\mu\Big[\sup_{x\in D}|\widetilde{f}_S(x) - f(x)|\Big] &= \mathbb{E}^\mu\Big[\sup_{x\in D\cap\mathbb{Q}_+}|\widetilde{f}_S(x) - f(x)|\Big]\\ &= \frac{1}{\sqrt{S}}\mathbb{E}^\mu\Big[\sup_{g\in\mathcal{G}}|\mathbb{G}_S(g)|\Big] \qquad (5.30)\\ &\leq C_0\frac{1}{\sqrt{S}}.\end{aligned}$$

As in the proof of Corollary 5.5, we transfer the sup-norm risk bound from the stationary regime to the case when $\boldsymbol{X}$ is started in 0. This can again be achieved utilizing exponential ergodicity of $\mathcal{O}$. Let

$$g(x,y) = \eta\gamma'(x+y) - \mathcal{A}_H\gamma(x), \quad x,y \geq 0.$$

Then,

$$\|g\|_\infty \leq \mathcal{B} := \|\eta\gamma'\|_\infty + \|\mathcal{A}_H\gamma\|_\infty,$$

which is finite by boundedness of γ'. Using exponential ergodicity of $\mathcal{O}$ as stated in Proposition 4.41 and applying Lemma 5.4 shows that for $\delta \in (0,1)$ and S large enough such that $S \geq (2+\delta)\log S$

$$\begin{aligned}&\Big|\mathbb{E}^0\Big[\sup_{x\in D}|\widetilde{f}_S(x) - f(x)|\Big] - \mathbb{E}^\mu\Big[\sup_{x\in D}|\widetilde{f}_S(x) - f(x)|\Big]\Big|\\ &\leq \Big|\Big|\mathbb{E}^0\Big[\sup_{x\in D}\Big|\frac{1}{T}\int_0^T g(x,\mathcal{O}_s)\Big|\Big] - \Big|\mathbb{E}^\mu\Big[\sup_{x\in D}\Big|\frac{1}{T}\int_0^T g(x,\mathcal{O}_s)\Big|\Big|\Big|\\ &\leq 2(2+\delta)\mathcal{B}\frac{\log S}{S} + c(\delta)\mathcal{B}\mathcal{R}_\lambda\exp(\lambda\cdot)(0)\frac{1}{S}\\ &\lesssim \frac{\log S}{S} + S^{-1}.\end{aligned}$$

Together with (5.30), this implies that

$$\mathcal{R}_\infty^D\big(\widetilde{f}_S, f\big) \leq C_1 S^{-1/2}$$

for some constant $C_1 > 0$, by triangle inequality. ∎

Proposition 5.12 shows that $\widetilde{f}_S$ is not only an elegant but also efficient estimator for $f = \mathcal{A}_H\gamma$, provided that we have an overshoot sample $(\mathcal{O}_t)_{0\leq t\leq S}$ available up to a fixed *level* S. However, we observe the Lévy process up to a fixed *time* T and not up to the random first passage time T_S. Our agenda therefore must be to build an estimator $\widehat{f}_T$ which is $\mathcal{F}_T$-measurable and whose sup-norm convergence properties can be inferred from Proposition 5.12. To this end, we aim to make use of the law of large numbers for Lévy processes. Recalling that $\lim_{T\to\infty} X_T/T = \mathbb{E}^0[X_1] = \eta$ almost surely for any starting distribution of $\boldsymbol{X}$, it follows that, for any $\varepsilon > 0$,

$$\mathbb{P}^0\Big(\Big|\frac{X_T}{T} - \eta\Big| > \varepsilon\Big) \underset{T\to\infty}{\longrightarrow} 0. \tag{5.31}$$

Define

$$\widehat{f}_T(x) := \frac{1}{X_T}\int_0^{X_T} \eta\gamma'(x+\mathcal{O}_t)\,\mathrm{d}t\,\mathbb{1}_{(0,\infty)}(X_T), \quad x\in\mathbb{R}, \tag{5.32}$$

and note that, since $\{X_T > t\} \subset \{T_t \leq T\}$ for any $t \geq 0$, we have

$$\gamma'(x+\mathcal{O}_t)\mathbb{1}_{\{t<X_T\}} = \gamma'(x+\mathcal{O}_t)\mathbb{1}_{\{T_t\leq T\}\cap\{t<X_T\}} \in \mathcal{F}_T,$$

as a consequence of $\gamma'(x+\mathcal{O}_t)\mathbb{1}_{\{T_t\leq T\}} \in \mathcal{F}_T$ thanks to T_t being an $\mathbb{F}$-stopping time. Therefore, $\widehat{f}_T(x) \in \mathcal{F}_T$ for any $x\in\mathbb{R}$ as desired. As a key result, the following preparatory lemma shows that the two essential components involved in an upper bound of the sup-norm risk of $\widehat{f}_T$ are indeed the rate of $\widetilde{f}_{\eta T} = \widetilde{f}_{\mathbb{E}^0[X_T]}$ and the speed of convergence in (5.31).

LEMMA 5.13. *Let $\gamma \in \mathcal{C}^2(\mathbb{R})$ such that γ' is bounded. Then, there exists a constant $C > 0$ such that, for any $\varepsilon \in (0, \eta\wedge 1/2)$ and $T > 0$, we have*

$$\mathcal{R}^D_\infty(\widehat{f}_T, f) \leq C\Big(\frac{1}{\sqrt{\eta T}} + \frac{\varepsilon}{\eta} + \mathbb{P}^0\Big(\Big|\frac{X_T}{T} - \eta\Big| > \varepsilon\Big)\Big). \tag{5.33}$$

Proof. Let again $\mathcal{B} := \eta\|\gamma'\|_\infty + \|\mathcal{A}_H\gamma\|_\infty < \infty$. Then, for $C = 2\max\{1, C_1, \mathcal{B}\}$, it follows by the triangle inequality and Proposition 5.12 that

$$\begin{aligned}
&\mathbb{E}^0\Big[\sup_{x\in D}\Big|\frac{1}{X_T}\int_0^{X_T}\eta\gamma'(x+\mathcal{O}_t)\,\mathrm{d}t - \mathcal{A}_H\gamma(x)\Big|\Big]\\
&\quad\leq \mathbb{E}^0\Big[\frac{\eta T}{X_T}\sup_{x\in D\cap\mathbb{Q}}\Big|\frac{1}{\eta T}\int_0^{\eta T\frac{X_T}{\eta T}}\eta\gamma'(x+\mathcal{O}_t)\,\mathrm{d}t - \mathcal{A}_H\gamma(x)\Big|;\Big\{\Big|\frac{X_T}{T}-\eta\Big|\leq\varepsilon\Big\}\Big] + \mathcal{B}\mathbb{P}^0\Big(\Big|\frac{X_T}{T}-\eta\Big|>\varepsilon\Big)\\
&\quad\leq 2\mathbb{E}^0\Big[\sup_{x\in D\cap\mathbb{Q}}\sup_{|\alpha|\leq\varepsilon/\eta}\Big|\frac{1}{\eta T}\int_0^{\eta T(1+\alpha)}\eta\gamma'(x+\mathcal{O}_t)\,\mathrm{d}t - \mathcal{A}_H\gamma(x)\Big|\Big] + \mathcal{B}\mathbb{P}^0\Big(\Big|\frac{X_T}{T}-\eta\Big|>\varepsilon\Big)\\
&\quad\leq 2\Big(\mathcal{R}^D_\infty(\widetilde{f}_{\eta T}, f) + \mathbb{E}^0\Big[\sup_{x\in D\cap\mathbb{Q}}\sup_{|\alpha|\leq\varepsilon/\eta}\Big|\frac{1}{\eta T}\int_{\eta T}^{\eta T(1+\alpha)}\eta\gamma'(x+\mathcal{O}_t)\,\mathrm{d}t\Big|\Big]\Big) + \mathcal{B}\mathbb{P}^0\Big(\Big|\frac{X_T}{T}-\eta\Big|>\varepsilon\Big)\\
&\quad\leq C\Big(\frac{1}{\sqrt{\eta T}} + \frac{\varepsilon}{\eta} + \mathbb{P}^0\Big(\Big|\frac{X_T}{T}-\eta\Big|>\varepsilon\Big)\Big),
\end{aligned}$$

where for the second inequality we used that, by our choice of $\varepsilon \in (0, \eta\wedge 1/2)$, we have $\eta T/X_T \leq (1-\varepsilon)^{-1} \leq 2$ on $\{|X_T/T - \eta| \leq \varepsilon\}$. ∎

The following result complements results on tail asymptotics of the marginal X_T for *fixed* $T > 0$ of a Lévy process $\boldsymbol{X}$ with bounded jumps, which can be found in Theorem 26.1 of [147], and non-asymptotic tail bounds of a Lévy process for small times $T > 0$, recently discussed in [77]. It is a slight digression from the remainder of this section in the sense that the assumptions ($\mathscr{L}0$)-($\mathscr{L}2$) are dropped in favour of bounded jumps and zero mean of $\boldsymbol{X}$. The statement is of independent interest since it gives nonasymptotic bounds on the speed of convergence of the law of large numbers for Lévy processes with bounded jumps and allows establishing optimal rates for our concrete estimation problem.

THEOREM 5.14. *Suppose that $\boldsymbol{X}$ is a non trivial zero mean Lévy process with bounded jumps and Lévy triplet (a, σ^2, Π). Then, there exists $\beta > 0$ and $T(p) > 0$ for $p > 0$ such that for any $T \geq T(p)$,*

$$\mathbb{P}^0\Big(|X_T| > \sqrt{\beta T \log(T^p)}\Big) \leq 2T^{-p/2}.$$

Proof. Let $\alpha := \inf\{z > 0 : \operatorname{supp}(\Pi) \subset \{x \in \mathbb{R} : |x| \leq z\}\}$ be the maximal jump size of $\boldsymbol{X}$. If $\alpha = 0$, $\boldsymbol{X}$ is a scaled Brownian motion since $\mathbb{E}^0[X_1] = 0$ and $\boldsymbol{X}$ was assumed non-trivial. In this case, the result follows directly from the exponential decay of tails of Brownian motion. Suppose therefore that $\alpha > 0$. We only show $\mathbb{P}^0(X_T > \sqrt{\beta T \log T}) \leq T^{-p/2}$. The statement then follows by performing the same calculations for the dual process $\widehat{\boldsymbol{X}} = -\boldsymbol{X}$, which also is a zero mean Lévy process with jumps bounded in absolute value by α. Since $\boldsymbol{X}$ has bounded jumps and zero mean, its Laplace exponent ψ is well-defined on $(0, \infty)$ and given by

$$\psi(z) := \log \mathbb{E}^0[\exp(zX_1)] = \frac{\sigma^2 z^2}{2} + \int_{-\alpha}^{\alpha} (\mathrm{e}^{zx} - 1 - zx)\, \Pi(\mathrm{d}x), \quad z > 0.$$

Furthermore, observe that ψ is smooth with derivative

$$\psi'(z) = \sigma^2 z + \int_{-\alpha}^{\alpha} (x(\mathrm{e}^{zx} - 1))\, \Pi(\mathrm{d}x), \quad z > 0.$$

By [147, Lemma 26.4], ψ' is invertible on $(0, \infty)$ with strictly increasing inverse denoted by θ. As in the proof of [147, Lemma 26.5] it follows from

$$z = \sigma^2 \theta(z) + \int_{-\alpha}^{\alpha} x(\mathrm{e}^{\theta(z)x} - 1)\, \Pi(\mathrm{d}x), \quad z > 0,$$

that

$$\frac{z}{\theta(z)} \leq \sigma^2 + \mathrm{e}^{\theta(z)\alpha} \int_{-\alpha}^{\alpha} x^2\, \Pi(\mathrm{d}x), \quad z > 0.$$

Since $\theta(0+) = 0$, this yields

$$\limsup_{z \downarrow 0} \frac{z}{\theta(z)} \leq \sigma^2 + \int_{-\alpha}^{\alpha} x^2\, \Pi(\mathrm{d}x).$$

This implies that there exists some $\varepsilon > 0$ and $\delta \geq 1$ such that for all $z \in (0, \varepsilon)$,

$$\theta(z) \geq \frac{z}{\delta\big(\sigma^2 + \int_{-\alpha}^{\alpha} x^2\, \Pi(\mathrm{d}x)\big)}. \tag{5.34}$$

Moreover, it follows from [147, Lemma 26.4] that for any $x > 0$,

$$\mathbb{P}^0(X_T > x) \leq \exp\Big(- \int_0^x \theta(z/T)\,\mathrm{d}z\Big) = \exp\Big(- T \int_0^{x/T} \theta(z)\,\mathrm{d}z\Big). \tag{5.35}$$

Defining $\beta := \delta(\sigma^2 + \int_{-\alpha}^{\alpha} x^2\,\Pi(\mathrm{d}x))$ and letting $T(p) > 0$ be large enough so that $\sqrt{\beta \log T^p/T} \in (0,\varepsilon)$ for all $t \geq T(p)$, it follows from (5.34) and (5.35) that indeed

$$\begin{aligned}\mathbb{P}^0\big(X_T > \sqrt{\beta T \log(T^p)}\big) &\leq \exp\Big(- T \int_0^{\sqrt{\beta \log(T^p)/T}} \theta(z)\,\mathrm{d}z\Big)\\ &\leq \exp\Big(- \frac{T}{\beta} \int_0^{\sqrt{\beta \log(T^p)/T}} z\,\mathrm{d}z\Big)\\ &= T^{-p/2}.\end{aligned}$$

■

With this preparation we can now investigate convergence rates of $\widehat{f}_T$.

THEOREM 5.15. *Let $\gamma \in \mathcal{C}^2(\mathbb{R})$ such that γ' is bounded.*

(i) Suppose that $\mathbb{E}^0[|X_1|^p] < \infty$ for some $p \geq 2$. Then,

$$\mathcal{R}^D_\infty(\widehat{f}_T, f) \in \mathrm{O}\Big(T^{-\frac{1}{2(1+1/p)}}\Big).$$

In particular, if all moments of X_1 exist, then, for any $\varepsilon > 0$,

$$\mathcal{R}^D_\infty(\widehat{f}_T, f) \in \mathrm{O}\Big(T^{-\frac{1}{2+\varepsilon}}\Big).$$

(ii) Suppose that $\boldsymbol{X}$ has bounded jumps. Then, for T large enough, it holds that

$$\mathcal{R}^D_\infty(\widehat{f}_T, f) \lesssim \sqrt{\frac{\log T}{T}}.$$

Proof. (i) Since $\mathbb{E}^0[|X_1|^p] < \infty$, it follows from the Burkholder–Davis–Gundy inequality for the càdlàg martingale $\widetilde{\boldsymbol{X}} := (X_t - \eta t)_{t\geq 0}$ (cf. [65, Theorem VII.92]), that there exists $C_p \in (0,\infty)$ such that

$$\mathbb{E}^0\Big[\Big|\frac{X_T}{T} - \eta\Big|^p\Big] = \frac{1}{T^p}\mathbb{E}^0[|X_T - \eta T|^p] \leq C_p \frac{1}{T^p}\mathbb{E}^0\big[[\widetilde{\boldsymbol{X}}]_T^{p/2}\big] = C_p \mathrm{Var}(X_1)^{p/2} T^{-p/2},$$

Here, $([\widetilde{\boldsymbol{X}}]_t)_{t\geq 0}$ denotes the quadratic variation of $\widetilde{\boldsymbol{X}}$. Hence, by Markov's inequality, it follows that

$$\begin{aligned}\mathbb{P}^0\Big(\Big|\frac{X_T}{T} - \eta\Big| > T^{-1/(2(1+p^{-1}))}\Big) &\leq C_p \mathrm{Var}(X_1)^{p/2} T^{p/(2(1+p^{-1}))} T^{-p/2}\\ &= C_p \mathrm{Var}(X_1)^{p/2} T^{-1/(2(1+p^{-1}))}.\end{aligned} \tag{5.36}$$

Plugging $\varepsilon = T^{-1/(2(1+p^{-1}))}$ into (5.33) and using (5.36), we conclude that

$$\mathcal{R}^D_\infty(\widehat{f}_T, f) \in \mathrm{O}(T^{-1/(2(1+p^{-1}))}).$$

(ii) Since $(X_t - \eta t)_{t\geq 0}$ is a zero mean Lévy process with bounded jumps, it follows from Theorem 5.14 that there exists some constant $\beta > 0$ such that for T large enough

$$\mathbb{P}^0\Big(\Big|\frac{X_T}{T} - \eta\Big| > \sqrt{\frac{\beta \log T}{T}}\Big) = \mathbb{P}^0\big(|X_T - \eta T| > \sqrt{\beta T \log T}\big) \leq \frac{2}{\sqrt{T}}. \tag{5.37}$$

Thus, plugging in $\varepsilon = \sqrt{\log T/T}$ into (5.33) gives the result.

■

Remark 5.16. (i) Since our exponential β-mixing assumption requires flat tails of Π at $+\infty$ and moreover $\mathbb{E}^0[X_1] > 0$, the assumption of exponential moments is quite natural in our modelling framework. When jumps are bounded, $(\mathscr{L}2)$ is always satisfied. Hence, for most Lévy processes falling into our estimation regime, we can expect a convergence rate of approximately $1/\sqrt{T}$.

(ii) One may wonder whether there was anything to gain, if in the definition of $\widehat{f}_T$, we replaced X_T by the running supremum $\overline{X}_T$. In practice, this would be more natural since otherwise—at least intuitively—data $(\mathbb{O}_t)_{X_T < t \leq \overline{X}_T}$ was wasted and moreover the estimator becomes meaningless whenever $X_T \leq 0$ (which, as time progresses becomes increasingly unlikely). The construction of our estimator on the other hand is driven by analytical tractability. However, in terms of the convergence rate of the estimator we cannot expect to gain much by working with the running supremum. This is evident from observing that Doob's maximal inequality for the submartingale $\boldsymbol{X}$ yields that for any $p > 1$ s.t. $X_1 \in L^p(\mathbb{P}^0)$ and $T \geq 1$,

$$\|X_T\|_{L^p(\mathbb{P}^0)} \leq \|\overline{X}_T\|_{L^p(\mathbb{P}^0)} \leq \frac{p}{p-1}\|X_T\|_{L^p(\mathbb{P}^0)}.$$

Let us interpret this result in detail from a nonparametric angle and, as announced at the beginning of this section, compare our estimator $\widehat{f}_T$ to the plug-in estimator given in (5.25) for the restricted setting of subordinators $\boldsymbol{X}$ with strictly positive drift $d_X > 0$ and absolutely continuous Lévy measure Π with bounded density π, for which the latter can be applied.

For the subordinator case, the plug-in estimator has an L^2 convergence rate of $1/\sqrt{T}$. As shown in Theorem 5.15, the overshoot estimator converges at rate $\sqrt{\log T/T}$ with respect to the $\|\cdot\|_\infty$-norm for any given Lévy process with bounded jumps satisfying $(\mathscr{L}0)$ and $(\mathscr{L}1)$ and hence in particular for any subordinator with Lévy measure having bounded support (but not necessarily bounded density since infinite jump activity is allowed). It is well-known from nonparametric invariant density estimation of well-behaved scalar stochastic processes that, within a continuous observation scheme, the invariant density can be estimated with the parametric rate $1/\sqrt{T}$ wrt. the L^2 norm. Estimation wrt. the sup-norm on the other hand introduces an additional log-factor, increasing the optimal rate to $\sqrt{\log T/T}$, as, e.g., in the previously discussed case of scalar ergodic diffusions, see Theorem 5.3.

Thus, in the current nonparametric estimation context we observe the same phenomenon that the common price to be paid is an additional log-factor for optimal estimation with respect to the sup-norm compared to the optimal L^2 rate. This also indicates that in principle, our approach to find an upper bound on the convergence rate of the overshoot estimator via Proposition 5.12 and Lemma 5.13 for a time-dependent observation scheme is tight enough to establish the optimal convergence rate $\sqrt{\log T/T}$ for more general Lévy processes with unbounded jumps.

This is evident from observing that the key result for the proof of Theorem 5.14 is Lemma 26.4 from [147], which relies on a Chernoff bound for the upper tail of a Lévy process at some fixed time T. However, interpreting this bound rigorously requires being able to tightly control the asymptotic behaviour of the inverse of the Laplace exponent's derivative, which for general Lévy processes is not possible. This is why we made use of Markov's inequality with power functions in the proof of part (i) of Theorem 5.15 instead of the generic Chernoff bound. However, for more particular classes of Lévy processes with explicit Laplace exponent, an ansatz similar to Theorem 5.14 may also provide the optimal convergence rate.

5.3.2 *Application*

We now return to the control problem described at the beginning of this section. In the following, we still assume ($\mathscr{L}0$) – ($\mathscr{L}2$) and now present the main tool for our analysis, an auxiliary function f defined via

$$f(x) := \mathcal{A}_H \gamma(x),$$

where $\mathcal{A}_H$ denotes the extended generator of the ladder height process $\boldsymbol{H}$ of $\boldsymbol{X}$ as discussed in Section 5.3.1. Noting that when $\gamma \in \mathcal{C}_0^2(\mathbb{R})$, Dynkin's formula and the fact that the values of $\boldsymbol{X}$ and $\boldsymbol{H}$ coincide at first hitting times almost surely together with our scaling of local time yielding (5.23), imply that

$$f(x) = \lim_{\varepsilon \downarrow 0} \frac{\mathbb{E}^x[\gamma(X_{T_{x+\varepsilon}})] - \gamma(x)}{\mathbb{E}^x[T_{x+\varepsilon}]},$$

this generates an intuition why this function is suitable for the analysis of problem (5.22): using the theory of regenerative processes, see [15], the value $\frac{\mathbb{E}^x[\gamma(X_{T_{x+\varepsilon}})]-\gamma(x)}{\mathbb{E}^x[T_{x+\varepsilon}]}$ coincides with the value of the (s, S) impulse strategy which shifts the process back to $x = s$ whenever the process is above $S = x + \varepsilon$, so that—at least intuitively—$f(x)$ corresponds to the value of the reflection strategy in x. The usefulness of this approach for ergodic impulse control problems is demonstrated in [49], where most of the following results can be found. Some further complementing analysis is carried out in [156]. The main observation is that properties of the function f determine the form of the optimal solution. For our considerations, we assume the following:

($\mathscr{L}3$) The function f has a unique maximum $\theta^* \in \mathbb{R}$, is strictly increasing on $(-\infty, \theta^*]$ and strictly decreasing on $[\theta^*, \infty)$.

Solution for known processes using an auxiliary impulse control problem

In [49], different classes of functions γ are discussed that make ($\mathscr{L}3$) hold for all Lévy processes $\boldsymbol{X}$. The main idea for analysing the problem (5.22) is to introduce artificial fixed costs K for each interaction, so that we are faced with a problem where we expect stationary impulse control strategies of (s, S)-type to be optimal. By considering the solutions for $K \searrow 0$, we then obtain the value and an optimal strategy for the problem without costs. More precisely, for each $K \geq 0$, we define

$$v(K) := \sup_S \liminf_{T \to \infty} \frac{1}{T} \mathbb{E}^x \left[\sum_{n: \tau_n \leq T} \left(\gamma(X^S_{\tau_n,-}) - \gamma(\zeta_n) - K \right) \right],$$

where the supremum is taken over all admissible impulse control strategies $S = (\tau_n, \zeta_n)$. By elementary arguments, it is immediately seen that $v(K)$ is independent of the initial state x. To study the dependence on the fixed costs $K \geq 0$, let us shortly review the key results on long-term average impulse control problems.

LEMMA 5.17 ([156], Theorem 4.3.6). *(i) For all $K \geq 0$*

$$v(K) = \sup_{x^*,\bar{x}\in\mathbb{R},\, x^*<\bar{x}} \frac{\mathbb{E}^{x^*}\left[\gamma\left(X_{T_{\bar{x}}}\right)\right] - \gamma(x^*) - K}{\mathbb{E}^{x^*}[T_{\bar{x}}]}.$$

(ii) If $K > 0$, then an (s, S) strategy of the form

$$\tau_n = \inf\{t \geq \tau_{n-1} : X_t \geq \bar{x}_K\}, \quad \zeta_n = x_K^*,$$

is optimal. The values $\bar{x}_K$ and x_K^* are given as follows: $\bar{x}_K$ is the larger of the two roots of the equation

$$f(x) = v(K).$$

If we denote the lower one by $\underline{x}_K$, then x_K^* is given as the maximizer $x_K^* = y \in [\underline{x}_K, \bar{x}_K]$ of

$$\frac{\mathbb{E}^{y}\left[\gamma\left(X_{T_{\bar{x}_K}}\right)\right] - \gamma(y) - K}{\mathbb{E}^{y}[T_{\bar{x}_K}]}.$$

Under additional assumptions on the Lévy process, it turns out that $x_K^* = \underline{x}_K$ which simplifies the solution of the impulse control problems, but is not needed for our purposes. We now study the dependence of $v(K)$ on K.

THEOREM 5.18 ([156], Theorem 4.3.6, 5.3.3, 5.3.4, and 5.3.5.). *In the singular control problem (5.22), the following holds true:*

(i) $v = f(\theta^*)$ $\;(= \max_{x\in\mathbb{R}} f(x))$;

(ii) $v(K) \nearrow v$ as $K \searrow 0$;

(iii) The (s, S) strategies with upper threshold $\bar{x}_K$ and restarting point x_K^*, $K > 0$, are ε-optimal for (5.22) as $K \searrow 0$;

(iv) $\underline{x}_K \nearrow \theta^*$ and $\bar{x}_K \searrow \theta^*$ as $K \searrow 0$.

The previous results suggest that the reflection strategy at level θ^* is optimal in problem (5.22). However, this strategy does not directly fall into the class of impulse control strategies we consider here, but is of (strictly) singular type. In order not to overburden the presentation with technicalities, we leave out the discussion of extending the strategy space here. Note however that, due to our ergodic problem formulation, extending the control space is even not needed to obtain optimizers for (5.22): the (non-stationary) threshold strategy with time dependent thresholds $x_{K_T}^*$ and $\bar{x}_{K_T}$ (with $K_T \searrow 0$ as $T \nearrow \infty$) is optimal in the class of impulse strategies and converges with arbitrary speed in T by choosing sufficiently small costs K_T, $T \geq 0$. Therefore, the term 'reflection strategy' refers to a suitably fast approximating impulse strategy in the following.

Data-driven singular controls

The results in Section 5.3.1 now directly lead to a method for estimating the optimal reflection boundary θ^*: after having observed the underlying Lévy process for T time units, we define the estimator for the auxiliary function f, using the estimator $\widehat{f}_T$ defined in (5.32), and then choose

$$\widehat{\theta}_T \in \arg\max_{\theta \in \overline{D}} \widehat{f}_T(\theta), \tag{5.38}$$

where D is some arbitrary bounded, open neighborhood of θ^*. The results from Section 5.3.1 now yield that the estimated optimizer gives the optimal value $v = f(\theta^*)$ up to a regret of order $T^{-1/(2(1+1/p))}$ when the p-th moment of $\boldsymbol{X}$ exists and of order $\sqrt{\log T/T}$ when jumps of $\boldsymbol{X}$ are bounded. Indeed:

THEOREM 5.19. *Let D be a bounded open neighborhood of θ^*. Suppose that $\mathbb{E}^0[|X_1|^p] < \infty$ for some $p \geq 2$. Then, it holds that*

$$\mathbb{E}^0\Big[v - f(\widehat{\theta}_T)\Big] \in \mathrm{O}\Big(T^{-\frac{1}{2(1+1/p)}}\Big).$$

If $\boldsymbol{X}$ has bounded jumps, then

$$\mathbb{E}^0\Big[v - f(\widehat{\theta}_T)\Big] \in \mathrm{O}\Big(\sqrt{\tfrac{\log T}{T}}\Big).$$

Proof. As in the proof of Proposition 5.7, we obtain

$$v - f(\widehat{\theta}_T) \leq 2\sup_{x\in D}\Big|f(x) - \widehat{f}_T(x)\Big|,$$

hence

$$\mathbb{E}^0\Big[v - f(\widehat{\theta}_T)\Big] \leq 2\mathcal{R}_\infty^D(\widehat{f}_T, f),$$

yielding the result by Theorem 5.15. ∎

5.4 DISCUSSION

The statistical questions discussed in this chapter have a clear motivation coming from the analysis of data-driven strategies for natural classes of stochastic control problems. For underlying diffusion processes, the solutions to ergodic singular control problems from Section 5.2.2 can be written in terms of the invariant density, such that the key to developing data-driven strategies consists in replacing this quantity by a sample-based analogue. From a statistical perspective, this is advantageous because rate-optimal estimation in this case (as opposed to, e.g., estimation of the drift coefficient) does not require an adaptive choice of the bandwidth. Due to the costs for reflection, the error measure to be used is the sup-norm risk studied in Section 5.2.1. This is an interesting observation as for the—from the stochastic control perspective highly related—impulse control problem investigated in [50], the L^1 risk had to be analysed. The substantially more involved issue of bounding the sup-norm risk of estimators is tackled by means of Theorem 3.7, exploiting mixture properties of the diffusion process. Since the focus of this chapter is on the development of concrete control strategies, we have restricted the presentation in Section 5.2.1 to a concise proof of the required upper bound (see Theorem 5.3).

We have reduced our explicit statistical investigation to the one-dimensional case solely because of the intended application to the stochastic control problem, but optimal convergence rates can be obtained in higher dimensions without much additional effort since exponential ergodicity and heat kernel bounds for the transition density generalize to higher dimensions. Compared to the L^1 risk, the evaluation of the sup-norm risk produces a well-known unavoidable logarithmic factor, which is also reflected in the expected regret per time unit (Theorem 5.9).

While the underlying diffusion processes in Section 5.2 were assumed to have an ergodic behaviour allowing for a statistical analysis, this is not the case for the Lévy-driven problem introduced in 5.3.2. By considering a space-time transformation of the Lévy process $\boldsymbol{X}$ in form of the overshoot process $\mathfrak{O}$, we obtained an ergodic Markov process fitting right into our general modeling framework, which allows to express the quantity of interest for the singular control problem, $f = \mathcal{A}_H\gamma$, as an integral w.r.t. its invariant distribution. Combining a simple mean estimator based on an overshoot sample with classical results on the long-time behaviour of Lévy processes then allowed us to construct an estimator whose performance depends on the tail-behaviour of $\boldsymbol{X}$ and yields an almost parametric sup-norm estimation rate in case of light tails and the optimal nonparametric rate $\sqrt{\log T/T}$ when jumps are bounded.

Based on this estimation procedure, we were then able to identify a data-driven singular control strategy, such that the estimated optimal reflection boundary yields an expected regret of the same order as the nonparametric estimation of the auxiliary function f.

In contrast to the diffusion case, in the Lévy process framework we are not faced with an exploration vs. exploitation problem: due to the spatial homogeneity of Lévy processes, each controlled process carries the same information as the uncontrolled one (if we assume that the decision maker has access to the values $X^S_{\tau_n,-}$) as the decision maker can reconstruct an uncontrolled path by just undoing the controls. Therefore, the following greedy strategy can be applied without additional losses: we use the (approximate) reflection controls with time-dependent boundary

$$\widehat{\theta}_T \in \arg\max_{\theta\in\overline{D}} \widehat{f}_T(\theta)$$

for each time point T.

Finally, let us briefly outline the connection to related research fields. The exploration vs. exploitation trade-off encountered in Section 5.2 is also well-known from the famous multi-armed bandit problem. In this regard, it is interesting to observe that the number of boundaries to be estimated in the control problems in our context does not influence the rate of convergence. Up to the logarithmic factor coming from the sup-norm vs. L^1 risk discussed above, the rates of convergence indeed turn out to be the same for the two-sided problem studied here and the one-sided problem from [50]. This is in strong contrast to the related results for $\mathcal{X}$-armed bandit problems, see [37, 119].

From a more applied point of view, it is furthermore of interest to compare the data-driven procedure proposed here to results obtained by using established methods from (deep) reinforcement learning. These algorithms are very generally applicable, as they usually only require the presence of a Markovian decision process setting. For classical methods such as the regular Q-learning algorithm, very robust convergence results exist; however, the latter is not practicable for problems in which the state space is too large. In the stochastic control setting considered here, a very large state space cannot be avoided, and a natural approach for circumventing this obstacle is to treat the problem based on the Q-learning algorithm with function approximation.

In this respect, the fusion with neural networks has proven to be particularly powerful. A mathematical theory of convergence for the resulting deep reinforcement learning procedures however is still under development. Results from recent contributions such as [80] are very interesting, but there remains a large gap between their theoretical assumptions and the Markov decision process framework that emerges for our concrete control problems. It seems practically impossible to apply their general convergence statements for deep Q-learning to our concrete setting such that one is forced to fall back on purely empirical tests of the algorithms. By way of contrast, our statistically driven method allows for a thorough theoretical analysis and yields rules that are both interpretable and explainable.

Regarding the practical implementation, we do not give a detailed numerical comparison here as this strongly depends on the exact framework, but just mention that in our scenarios both approaches learn the optimal rule reasonably well, where the statistical approach is (not surprisingly) faster and for a longer time horizon very accurate.

5.A PROOF OF LEMMA 5.8

Before we start with the proof we need some preparatory remarks. The length of the first exploration period is given by

$$\tau_1 = \inf\{t \geq \sigma_1 : X_t = 0\} = \sigma_1 + T_0 \circ \theta_{\sigma_1},$$

with $\sigma_1 = \inf\{t \geq 0 : X_t = -B\} \vee \inf\{t \geq 0 : X_t = B\}$, $T_a = \inf\{t \geq 0 : X_t = a\}$ for $a \in \mathbb{R}$ and $(\theta_t)_{t\geq 0}$ denote the transition operators of the Markov process $\boldsymbol{X}$. We will need that $\mathbb{E}_b^0[\tau_1^3] < \infty$. By the strong Markov property and the fact that $\mathbb{P}_b^0(\sigma_1 < \infty)$ by point recurrence of the ergodic diffusion $\boldsymbol{X}$ we obtain

$$\mathbb{E}_b^0[\tau_1^3] \leq 4\big(\mathbb{E}_b^0[T_B^3] + \mathbb{E}_b^0[T_{-B}^3]\big) + 4\big(\mathbb{E}_b^B[T_0^3] + \mathbb{E}_b^{-B}[T_0^3]\big),$$

and thus finiteness of the third moment of τ_1 boils down to finiteness of the third moment of first hitting times of the diffusion. From [120, Section 5], see also [19], it is known that if the diffusion coefficient σ is bounded and there exist $r, M_0 > 0$ such that

$$-\frac{xb(x)}{\sigma^2(x)} \geq r, \quad \forall |x| > M_0, \tag{5.39}$$

then $\mathbb{E}^x[T_a^n] < \infty$ for all $n < r + 1/2$. In our setting boundedness of σ is satisfied with $0 < \underline{\nu} \leq \sigma(x) \leq \bar{\nu} < \infty$ and we have

$$\operatorname{sgn}(x)\frac{b(x)}{\sigma^2(x)} \leq -\gamma, \quad \forall |x| > A,$$

for some constants $\gamma, A > 0$. Thus, for $|x| > A \vee r/\gamma = M_0$, (5.39) is fulfilled, which implies that for any $n \in \mathbb{N}$ and $a, x \in \mathbb{R}$, $\mathbb{E}^x[T_a^n] < \infty$. It is worth noting that [120] demonstrate how the existence of moments of hitting times is initimately connected to polynomial ergodicity of a diffusion and hence the existence of arbitrarily large hitting time moments can be regarded as a consequence of exponential ergodicity of $\boldsymbol{X}$ under the given assumptions. In particular $\mathbb{E}_b^0[\tau_1^3] < \infty$ as desired.

We will also make use of the following simple observation.

LEMMA 5.20. *Let X be a random variable taking values in $[1,\infty)$ almost surely and suppose that*

$$(0,2) \ni \beta \mapsto \mathbb{E}[X^\beta]$$

is differentiable with $\frac{\partial}{\partial\beta}\mathbb{E}[X^\beta] = \mathbb{E}[\frac{\partial}{\partial\beta}X^\beta]$. Then, the function

$$[0,1] \ni \alpha \mapsto \mathrm{Var}(X^\alpha)$$

is increasing.

Proof. By the smoothness assumptions we have

$$\frac{\partial}{\partial\alpha}\mathrm{Var}(X^\alpha) = 2\alpha\mathbb{E}[X^\alpha \log X] - 2\alpha\mathbb{E}[X^\alpha]\mathbb{E}[X^{2\alpha}\log X] = 2\alpha\mathrm{Cov}(X^\alpha, f(X^\alpha)),$$

with $[1,\infty) \ni x \mapsto f(x) = \alpha^{-1}x\log x$. Since f is increasing it follows that $\mathrm{Cov}(X^\alpha, f(X^\alpha)) \geq 0$, proving the assertion. ∎

We are now ready to carry out the proof of Lemma 5.8.

Proof of Lemma 5.8. We start with some necessary notation. Let η_n^j, $n \in \mathbb{N}$, be the length of the n-th exploration period for $j = 0$ and the length of the n-th exploitation period for $j = 1$. In particular, $\eta_1^0 = \tau_1$ and $(\eta_n^0)_{n=2,3,\ldots}$ is an i.i.d. family of random variables under $\mathbb{P}_b^0$. Define also $\tau_n^j := \sum_{i=1}^n \eta_i^j$ as the length of the first n exploration/exploitation periods, thus

$$\tau_n = \tau^0_{n-\sum_{i=1}^n c_i} + \tau^1_{\sum_{i=1}^n c_i},$$

where $\tau_0^0 = \tau_0^1 = 0$. Finally, denote $N_t := \inf\{n \in \mathbb{N} : \tau_n > t\}$ and the number of exploration/exploitation periods starting before time $t \geq 0$, $N_t^j = K^j_{N_t}$, where

$$K_n^j := \#\{i \leq n : c_i = j\}, \quad n \in \mathbb{N}, j \in \{0,1\}.$$

Clearly, $\tau^0_{N_t^0-1} \leq S_t \leq \tau^0_{N_t^0}$ for any $t \geq 0$, and thus

$$\frac{\tau^0_{N_t^0-1}}{N_t^0} \leq \frac{S_t}{N_t^0} \leq \frac{\tau^0_{N_t^0}}{N_t^0}.$$

By construction and the strong law of large numbers, both the left-hand and the right-hand side $\mathbb{P}_b^0$-a.s. converge to $\mathbb{E}_b^0[\eta_2^0]$ and hence

$$\frac{S_T}{N_T^0} \xrightarrow[T\to\infty]{} \mathbb{E}_b^0[\eta_2^0], \quad \mathbb{P}_b^0\text{-a.s.}. \tag{5.40}$$

Let now $\underline{\eta}^j, \bar{\eta}^j$ be such that

$$\underline{\eta}^j \leq_{\mathrm{st}} \eta_n^j \leq_{\mathrm{st}} \bar{\eta}^j, \quad n \in \mathbb{N}, j \in \{1,2\},$$

where $\leq_{\mathrm{st}}$ denotes stochastic ordering. Existence of such random variables is clear in case $j = 0$ and for $j = 1$ we can take $\underline{\eta}^1$ to be a cycle length when reflecting in $\pm 1/B$ and $\bar{\eta}_1$ be a cycle length when reflecting in $\pm B$. Choosing $\underline{\eta} = \underline{\eta}^0 \wedge \underline{\eta}^1$ and $\bar{\eta} = \bar{\eta}^0 \vee \bar{\eta}^1$ we have

$$\underline{\eta} \leq_{\mathrm{st}} \eta_n^j \leq_{\mathrm{st}} \bar{\eta}, \quad n \in \mathbb{N}, j \in \{1, 2\},$$

Let now $(\underline{\eta}_n)_{n\in\mathbb{N}}$ and $(\bar{\eta}_n)_{n\in\mathbb{N}}$ be i.i.d. copies of $\underline{\eta}$ and $\bar{\eta}$, resp., where by resorting to a coupling argument if needed, we may assume wlog that $\underline{\eta}_n \leq \eta_n^j \leq \bar{\eta}_n$, $\mathbb{P}_b^0$-a.s. for all $n \in \mathbb{N}$, $j \in \{0, 1\}$ (see, e.g., [162, Section 3]). Defining

$$\underline{N}_t = \inf\Big\{n \in \mathbb{N} : \sum_{i=1}^n \underline{\eta}_i > t\Big\}, \quad \overline{N}_t = \inf\Big\{n \in \mathbb{N} : \sum_{i=1}^n \bar{\eta}_i > t\Big\},$$

it follows that

$$\overline{N}_t \leq N_t \leq \underline{N}_t, \quad \mathbb{P}_b^0\text{-a.s.},$$

With the standard renewal theorem we have

$$\lim_{t\to\infty} \frac{\underline{N}_t}{t} = \frac{1}{\mathbb{E}_b^0\big[\underline{\eta}\big]}, \text{ and } \lim_{t\to\infty} \frac{\overline{N}_t}{t} = \frac{1}{\mathbb{E}_b^0\big[\bar{\eta}\big]}, \quad \mathbb{P}_b^0\text{-a.s. and in } L^1(\mathbb{P}_b^0).$$

By construction, we have

$$\frac{N_t^0}{t^{2/3}} \leq \overline{M}\Big(\frac{N_t}{t}\Big)^{2/3} \leq \overline{M}\Big(\frac{\underline{N}_t}{t}\Big)^{2/3}$$

and since by Jensen's inequality and the above

$$\mathbb{E}_b^0\Big[\Big(\frac{\underline{N}_t}{t}\Big)^{2/3}\Big] \leq \mathbb{E}_b^0\Big[\frac{\underline{N}_t}{t}\Big]^{2/3} \xrightarrow[t\to\infty]{} \mathbb{E}_b^0\big[\underline{\eta}\big]^{-2/3},$$

it follows that

$$\limsup_{T\to\infty} T^{-2/3}\mathbb{E}_b^0[N_T^0] \leq \overline{M}\mathbb{E}_b^0\big[\underline{\eta}\big]^{-2/3} =: M,$$

which establishes the second part of the assertion. For the first part, consider the uncontrolled diffusion process $\boldsymbol{X}$ and let

$$\breve{\tau}_n = \begin{cases} 0, & \text{if } n = 0 \\ \inf\{t \geq \sigma_n : X_t = 0\}, & \text{if } n \in \mathbb{N}, \end{cases}$$

where

$$\sigma_n = \inf\{t \geq \breve{\tau}_{n-1} : X_t = B\} \vee \inf\{t \geq \breve{\tau}_{n-1} : X_t = -B\}, \quad n \in \mathbb{N}.$$

By the strong Markov property, $(\breve{\tau}_n)_{n\in\mathbb{N}}$ are i.i.d. and if we denote $\breve{N}_t = \inf\{n \in \mathbb{N} : \breve{\tau}_n > t\}$ for $t \geq 0$, then $(\breve{N}_t)_{t\geq 0}$ is a renewal process with increment distribution $\breve{\tau}_1$. Furthermore, we define in analogy to the controlled case above $\breve{\eta}_n^i := \breve{\tau}_{C_n^i} - \breve{\tau}_{C_n^i - 1}$ for $i = 0, 1$, where $C_n^0 := \min\{m \in \mathbb{N} : \sum_{i=1}^m (1 - c_i) \geq n\}$ and $C_n^1 := \min\{m \in \mathbb{N} : \sum_{i=1}^m c_i \geq n\}$, and $\breve{\tau}_n^j := \sum_{i=1}^n \breve{\eta}_i^j$ for $j = 0, 1$. Finally, let $\breve{N}_t^i = K_{\breve{N}_t}^i$ for $t \geq 0$. Clearly,

$$\breve{\eta}_n^0 \overset{\mathrm{d}}{=} \eta_n^0 \quad \text{and} \quad \breve{\eta}_n^1 \geq_{\mathrm{st}} \eta_n^1$$

for any $n \in \mathbb{N}$, which yields

$$\breve{S}_t \leq_{\mathrm{st}} S_t, \quad t \geq 0, \tag{5.41}$$

where

$$\breve{S}_t = \int_0^t \sum_{n=1}^{\breve{N}_t^0} \mathbb{1}_{[\breve{\tau}_{c_n^0-1}, \breve{\tau}_{c_n^0}]}(s)\,\mathrm{d}s \in \left[\breve{\tau}^0_{\breve{N}_t^0-1}, \breve{\tau}^0_{\breve{N}_t^0}\right], \quad t \geq 0. \tag{5.42}$$

By the standard renewal theorem and the properties of $(c_n)_{n\in\mathbb{N}}$ we have

$$\frac{\breve{N}_t^0}{t^{2/3}} \geq \left(\frac{\breve{N}_t}{t}\right)^{2/3} \underset{t\to\infty}{\longrightarrow} \mathbb{E}_b^0[\breve{\tau}_1]^{-2/3}, \quad \mathbb{P}_b^0\text{-a.s.}$$

which, on account of the fact that by combining the strong law of large numbers and (5.42) we have

$$\frac{\breve{S}_T}{\breve{N}_T^0} \underset{T\to\infty}{\longrightarrow} \mathbb{E}_b^0[\breve{\tau}_1], \quad \mathbb{P}_b^0\text{-a.s.},$$

shows that

$$\liminf_{t\to\infty} \frac{\breve{S}_t}{t^{2/3}} \geq \mathbb{E}_b^0[\breve{\tau}_1]^{1/3} =: \widetilde{m}, \quad \mathbb{P}_b^0\text{-a.s..}$$

Consequently, an application of Fatou's lemma provides

$$\liminf_{t\to\infty} \mathbb{E}_b^0\left[\frac{\breve{S}_t}{t^{2/3}}\right] \geq \widetilde{m}.$$

Thus, choosing $m = \widetilde{m}/2$ there exists $\varepsilon \in (0, \widetilde{m}/2)$ such that for any $T > 0$ large enough we have

$$\mathbb{E}_b^0[\breve{S}_T] \geq T^{2/3}(m+\varepsilon),$$

which, together with Markov's inequality and (5.41), yields

$$\begin{aligned}
\mathbb{P}_b^0(S_T \leq mT^{2/3}) \leq \mathbb{P}_b^0(\breve{S}_T \leq mT^{2/3}) &\leq \mathbb{P}_b^0(\mathbb{E}_b^0[\breve{S}_T] - \breve{S}_T \geq \varepsilon T^{2/3}) \\
&\leq \mathbb{P}_b^0\big(\mathbb{E}_b^0\big[\breve{\tau}^0_{\breve{N}_T^0-1}\big] - \breve{\tau}^0_{\breve{N}_T^0} \geq \varepsilon T^{2/3}\big) \\
&\leq \varepsilon^{-2} T^{-4/3} \mathbb{E}_b^0\Big[\big(\breve{\tau}^0_{\breve{N}_T^0} - \mathbb{E}_b^0[\breve{\tau}^0_{\breve{N}_T^0-1}]\big)^2\Big] \\
&\leq 2\varepsilon^{-2} T^{-4/3}\big(\mathrm{Var}_{\mathbb{P}_b^0}(\breve{\tau}^0_{\breve{N}_T^0}) + \mathbb{E}_b^0\big[(\breve{\tau}_1)^2\big]\big).
\end{aligned}$$

Thus, to show that $\mathbb{P}_b^0(S_T \leq mT^{2/3}) \lesssim T^{-1/3}$ it is enough to establish that

$$\mathrm{Var}_{\mathbb{P}_b^0}(\breve{\tau}^0_{\breve{N}_T^0}) \lesssim T. \tag{5.43}$$

To this end, note first that for any $n \geq 2$ and $T \geq 0$ we have

$$\begin{aligned}
\{\breve{N}_T^0 \leq n\} &= \bigcup_{k=1}^n \bigcup_{l\geq k} \big(\{\breve{N}_T^0 = k\} \cap \{\breve{N}_T = l\}\big) \\
&= \bigcup_{k=1}^n \bigcup_{l\geq k} \Big\{\breve{\tau}_k^0 + \breve{\tau}^1_{l-k} > T, \breve{\tau}_{l-1} \leq T, \sum_{i=1}^{l-1}(1-c_i) \in \{k-1, k\}\Big\}
\end{aligned}$$

$$=: \bigcup_{k=1}^{n} \bigcup_{l \geq k} A_{k,l}.$$

By construction, $(\breve{\eta}_n^0)_{k \geq n+1} \perp\!\!\!\perp \mathcal{G}_n$, where

$$\mathcal{G}_n := \sigma\big(\{\breve{\eta}_1^0, \dots, \breve{\eta}_n^0\} \cup \{\breve{\eta}_i^1 : i < C_n^0\}\big)$$

is the σ-algebra generated by the lengths of the first "exploration/exploitation" periods up to the n-th "exploration" period of the uncontrolled process $\boldsymbol{X}$. Clearly, $A_{k,l} \in \mathcal{G}_n$ for any $k = 1, \dots, n$ and $l \geq k$, which shows that $\breve{N}_T^0$ is a $(\mathcal{G}_n)_{n \in \mathbb{N}}$ stopping time. Hence, we can use Wald's second moment identity, see [15, Propostion A10.2], to obtain

$$\mathrm{Var}_{\mathbb{P}_b^0}(\breve{\tau}_{\breve{N}_T^0}^0) = \mathrm{Var}_{\mathbb{P}_b^0}\Big(\sum_{i=1}^{\breve{N}_T^0} \breve{\eta}_i^0\Big) = \mathbb{E}_b^0[\breve{\tau}_1]^2 \cdot \mathrm{Var}_{\mathbb{P}_b^0}(\breve{N}_T^0) + \mathrm{Var}_{\mathbb{P}_b^0}(\breve{\tau}_1) \cdot \mathbb{E}_b^0[\breve{N}_T^0].$$

Since by Jensen's inequality and the renewal theorem

$$\limsup_{T \to \infty} T^{-2/3} \mathbb{E}_b^0[\breve{N}_T^0] \leq \limsup_{T \to \infty} \overline{M} T^{-2/3} \mathbb{E}^0\big[\breve{N}_T^{2/3}\big] \leq \overline{M} \limsup_{T \to \infty} \big(\mathbb{E}^0[\breve{N}_T]/T\big)^{2/3} < \infty,$$

(5.43) will follow if we can prove that $\mathrm{Var}_{\mathbb{P}_b^0}(\breve{N}_T^0) \lesssim T$. By classical renewal arguments, cf. [15, Proposition 6.3] we have

$$\lim_{T \to \infty} \frac{\mathrm{Var}_{\mathbb{P}_b^0}(\breve{N}_T)}{T} = \frac{\mathrm{Var}_{\mathbb{P}_b^0}(\breve{\tau}_1)}{\mathbb{E}_b^0[\breve{\tau}_1]^3}, \tag{5.44}$$

and by construction it follows that

$$\begin{aligned} \mathrm{Var}_{\mathbb{P}_b^0}(\breve{N}_T^0) &= \mathbb{E}^0[(\breve{N}_T^0)^2] - \mathbb{E}^0[\breve{N}_T^0]^2 \\ &\leq \mathbb{E}^0\big[\big((\breve{N}_T)^{2/3} + \mathfrak{d}\big)^2\big] - \mathbb{E}^0\big[(\breve{N}_T)^{2/3}\big]^2 \\ &\lesssim \mathrm{Var}_{\mathbb{P}_b^0}\big((\breve{N}_T)^{2/3}\big) + \mathbb{E}_b^0\big[(\breve{N}_T)^{2/3}\big]. \end{aligned}$$

By Jensen's inequality we have $\mathbb{E}_b^0\big[(\breve{N}_T)^{2/3}\big] \lesssim T$ and (5.44) combined with Lemma 5.20 yields $\mathrm{Var}_{\mathbb{P}_b^0}\big((\breve{N}_T)^{2/3}\big) \lesssim T$ (note here that $\mathbb{E}_b^0[\frac{\partial}{\partial \beta}(\breve{N}_T)^\beta] \leq 2\mathbb{E}_b^0[(\breve{N}_T)^3] < \infty$ for $\beta \in (0,2)$ and thus we may differentiate under the integral to obtain $\frac{\partial}{\partial \beta}\mathbb{E}_b^0[(\breve{N}_T)^\beta] = \mathbb{E}_b^0[\frac{\partial}{\partial \beta}(\breve{N}_T)^\beta]$ as needed). Thus, $\mathrm{Var}_{\mathbb{P}_b^0}(\breve{N}_T^0) \lesssim T$, which shows that indeed $\mathbb{P}_b^0(T^{-2/3} S_T \leq m) \lesssim T^{-1/3}$. ■

Bibliography

[1] R. Adamczak. "A tail inequality for suprema of unbounded empirical processes with applications to Markov chains". In: *Electron. J. Probab.* 13 (2008), no. 34, 1000–1034. DOI: 10.1214/EJP.v13-521.

[2] R. Adamczak and W. Bednorz. "Exponential concentration inequalities for additive functionals of Markov chains". In: *ESAIM Probab. Stat.* 19 (2015), pp. 440–481. DOI: 10.1051/ps/2014032.

[3] C. Aeckerle-Willems and C. Strauch. "Concentration of scalar ergodic diffusions and some statistical implications". In: *Ann. Inst. Henri Poincaré Probab. Stat.* (2021), to appear. arXiv: 1807.11331.

[4] Y. Aït-Sahalia and L. P. Hansen, eds. *Handbook of Financial Econometrics: Tools and Techniques*. 1st ed. Vol. 1. Handbooks in Finance. North Holland, 2009, pp. xxviii+780.

[5] L. Alili, L. Chaumont, P. Graczyk, and T. Żak. "Inversion, duality and Doob h-transforms for self-similar Markov processes". In: *Electron. J. Probab.* 22 (2017), Paper No. 20, 18. DOI: 10.1214/17-EJP33.

[6] G. Alsmeyer. "On the Markov renewal theorem". In: *Stochastic Process. Appl.* 50.1 (1994), pp. 37–56. DOI: 10.1016/0304-4149(94)90146-5.

[7] G. Alsmeyer and F. Buckmann. "Fluctuation theory for Markov random walks". In: *J. Theoret. Probab.* 31.4 (2018), pp. 2266–2342. DOI: 10.1007/s10959-017-0778-9.

[8] L. H. R. Alvarez. "Stochastic forest stand value and optimal timber harvesting". In: *SIAM J. Control Optim.* 42.6 (2004), pp. 1972–1993. DOI: 10.1137/S0363012901393456.

[9] L. H. R. Alvarez and L. A. Shepp. "Optimal harvesting of stochastically fluctuating populations". In: *J. Math. Biol.* 37.2 (1998), pp. 155–177. DOI: 10.1007/s002850050124.

[10] L. H. R. Alvarez and J. Virtanen. "A class of solvable stochastic dividend optimization problems: on the general impact of flexibility on valuation". In: *Econom. Theory* 28.2 (2006), pp. 373–398. DOI: 10.1007/s00199-005-0627-4.

[11] L. H. Alvarez. *A Class of Solvable Stationary Singular Stochastic Control Problems*. 2018. arXiv: 1803.03464.

[12] C. Amorino and A. Gloter. "Invariant density adaptive estimation for ergodic jump-diffusion processes over anisotropic classes". In: *J. Statist. Plann. Inference* 213 (2021), pp. 106–129. DOI: 10.1016/j.jspi.2020.11.006.

[13] C. Amorino and E. Nualart. *Optimal convergence rates for the invariant density estimation of jump-diffusion processes*. 2021. arXiv: 2101.08548.

[14] D. Applebaum. *Lévy Processes and Stochastic Calculus*. 2nd ed. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2009.

[15] S. Asmussen. *Applied probability and queues*. 2nd ed. Vol. 51. Applications of Mathematics (New York). Stochastic Modelling and Applied Probability. Springer-Verlag, New York, 2003, pp. xii+438.

[16] S. Asmussen and M. Taksar. “Controlled diffusion models for optimal dividend pay-out”. In: *Insurance Math. Econom.* 20.1 (1997), pp. 1–15. DOI: 10.1016/S0167-6687(96)00017-0.

[17] J. Azéma, M. Duflo, and D. Revuz. “Mesure invariante des processus de Markov récurrents”. In: *Séminaire de Probabilités, III (Univ. Strasbourg, 1967/68)*. Lecture Notes in Mathematics, Vol. 88. Springer, Berlin, 1969, pp. 24–33.

[18] D. Bakry, P. Cattiaux, and A. Guillin. “Rate of convergence for ergodic continuous Markov processes: Lyapunov versus Poincaré”. In: *J. Funct. Anal.* 254.3 (2008), pp. 727–759. DOI: 10.1016/j.jfa.2007.11.002.

[19] S. Balaji and S. Ramasubramanian. “Passage time moments for multidimensional diffusions”. In: *J. Appl. Probab.* 37.1 (2000), pp. 246–251. DOI: 10.1017/s0021900200015382.

[20] Y. Baraud. “A Bernstein-type inequality for suprema of random processes with applications to model selection in non-Gaussian regression”. In: *Bernoulli* 16.4 (2010), pp. 1064–1085. DOI: 10.3150/09-BEJ245.

[21] A. Bátkai, M. Kramar Fijavž, and A. Rhandi. *Positive operator semigroups*. Vol. 257. Operator Theory: Advances and Applications. Birkhäuser/Springer, Cham, 2017, pp. xvii+364. DOI: 10.1007/978-3-319-42813-0.

[22] D. Belomestny, F. Comte, V. Genon-Catalot, H. Masuda, and M. Reiß. *Lévy matters IV - Estimation for discretely observed Lévy processes*. Lecture Notes in Mathematics, Vol. 2128. Springer, Cham, 2015.

[23] P. Bertail and G. Ciolek. “New Bernstein and Hoeffding type inequalities for regenerative Markov chains”. In: *ALEA Lat. Am. J. Probab. Math. Stat.* 16.1 (2019), pp. 259–277. DOI: 10.30757/alea.v16-09.

[24] J. Bertoin, K. van Harn, and F. W. Steutel. “Renewal theory and level passage by subordinators”. In: *Statist. Probab. Lett.* 45.1 (1999), pp. 65–69. DOI: 10.1016/S0167-7152(99)00043-7.

[25] J. Bertoin. *Lévy processes*. Vol. 121. Cambridge Tracts in Mathematics. Cambridge University Press, Cambridge, 1996, pp. x+265.

[26] J. Bertoin. *Subordinators, Lévy processes with no negative jumps and branching processes*. MaPhySto Lecture Notes Series No. 8. Aarhus, 2000.

[27] J. Bertoin, T. Budd, N. Curien, and I. Kortchemski. “Martingales in self-similar growth-fragmentations and their connections with random planar maps”. In: *Probab. Theory Related Fields* 172.3-4 (2018), pp. 663–724. DOI: 10.1007/s00440-017-0818-5.

[28] J. Bertoin and M. Savov. “Some applications of duality for Lévy processes in a half-line”. In: *Bull. Lond. Math. Soc.* 43.1 (2011), pp. 97–110. DOI: 10.1112/blms/bdq084.

[29] J. Bierkens, G. O. Roberts, and P.-A. Zitt. “Ergodicity of the zigzag process”. In: *Ann. Appl. Probab.* 29.4 (2019), pp. 2266–2301. DOI: 10.1214/18-AAP1453.

[30] R. M. Blumenthal and R. K. Getoor. “Local times for Markov processes”. In: *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* 3 (1964), pp. 50–74. DOI: 10.1007/BF00531683.

[31] R. M. Blumenthal and R. K. Getoor. *Markov processes and potential theory*. Pure and Applied Mathematics, Vol. 29. Academic Press, New York-London, 1968, pp. x+313.

[32] R. M. Blumenthal. *Excursions of Markov processes*. Probability and its Applications. Birkhäuser Boston Inc., Boston, MA, 1992, pp. xii+275. DOI: 10.1007/978-1-4684-9412-9.

[33] D. Bosq. *Nonparametric statistics for stochastic processes*. 2nd ed. Vol. 110. Lecture Notes in Statistics. Estimation and prediction. Springer-Verlag, New York, 1998, pp. xvi+210. DOI: 10.1007/978-1-4612-1718-3.

[34] D. Bosq. "Parametric rates of nonparametric estimators and predictors for continuous time processes". In: *Ann. Statist.* 25.3 (1997), pp. 982–1000. DOI: 10.1214/aos/1069362734.

[35] S. Boucheron, G. Lugosi, and P. Massart. *Concentration inequalities: A nonasymptotic theory of independence*. Oxford University Press, Oxford, 2013, pp. x+481. DOI: 10.1093/acprof:oso/9780199535255.001.0001.

[36] N. Brosse, A. Durmus, and E. Moulines. "The promises and pitfalls of Stochastic Gradient Langevin Dynamics". In: *Advances in Neural Information Processing Systems 31*. Ed. by S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett. Curran Associates, Inc., 2018, pp. 8268–8278.

[37] S. Bubeck, R. Munos, G. Stoltz, and C. Szepesvári. "$\mathcal{X}$-armed bandits". In: *J. Mach. Learn. Res.* 12 (2011), pp. 1655–1695.

[38] A. Cadenillas, S. Sarkar, and F. Zapatero. "Optimal dividend policy with mean-reverting cash reservoir". In: *Math. Finance* 17.1 (2007), pp. 81–109. DOI: 10.1111/j.1467-9965.2007.00295.x.

[39] J. V. Castellana and M. R. Leadbetter. "On smoothed probability density estimation for stationary processes". In: *Stochastic Process. Appl.* 21.2 (1986), pp. 179–193. DOI: 10.1016/0304-4149(86)90095-5.

[40] P. Cattiaux and A. Guillin. "Deviation bounds for additive functionals of Markov processes". In: *ESAIM Probab. Stat.* 12 (2008), pp. 12–29. DOI: 10.1051/ps:2007032.

[41] E. Çinlar. "Lévy systems of Markov additive processes". In: *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* 31 (1974/75), pp. 175–185. DOI: 10.1007/BF00536006.

[42] E. Çinlar. "Markov additive processes. I, II". In: *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* 24 (1972), 85–93; ibid. 24 (1972), 95–121. DOI: 10.1007/BF00532536.

[43] E. Çinlar. "On semi-Markov processes on arbitrary spaces". In: *Proc. Cambridge Philos. Soc.* 66 (1969), pp. 381–392. DOI: 10.1017/s0305004100045096.

[44] L. Chaumont, A. Kyprianou, J. C. Pardo, and V. Rivero. "Fluctuation theory and exit systems for positive self-similar Markov processes". In: *Ann. Probab.* 40.1 (2012), pp. 245–279. DOI: 10.1214/10-AOP612.

[45] L. Chaumont, H. Pantí, and V. Rivero. "The Lamperti representation of real-valued self-similar Markov processes". In: *Bernoulli* 19.5B (2013), pp. 2494–2523. DOI: 10.3150/12-BEJ460.

[46] M.-F. Chen. *Eigenvalues, inequalities, and ergodic theory*. Probability and its Applications (New York). Springer-Verlag London, Ltd., London, 2005, pp. xiv+228.

[47] Z.-Q. Chen, E. Hu, L. Xie, and X. Zhang. "Heat kernels for non-symmetric diffusion operators with jumps". In: *Journal of Differential Equations* 263.10 (2017), pp. 6576 –6634. DOI: https://doi.org/10.1016/j.jde.2017.07.023.

[48] Y. S. Chow. "On moments of ladder height variables". In: *Adv. in Appl. Math.* 7.1 (1986), pp. 46–54. DOI: 10.1016/0196-8858(86)90005-9.

[49] S. Christensen and T. Sohr. "A solution technique for Lévy driven long term average impulse control problems". In: *Stochastic Process. Appl.* (2020). DOI: https://doi.org/10.1016/j.spa.2020.07.016.

[50] S. Christensen and C. Strauch. *Nonparametric learning for impulse control problems*. 2020. arXiv: 1909.09528.

[51] S. Christensen, C. Strauch, and L. Trottner. *Learning to reflect: A unifying approach for data-driven stochastic control strategies*. 2021. arXiv: 2104.11496.

[52] K. L. Chung and J. B. Walsh. *Markov processes, Brownian motion, and time symmetry*. 2nd ed. Grundlehren der Mathematischen Wissenschaften, Vol. 249. Springer, New York, 2005.

[53] S. J. M. Clémençon. "Moment and probability inequalities for sums of bounded additive functionals of regular Markov chains via the Nummelin splitting technique". In: *Statist. Probab. Lett.* 55.3 (2001), pp. 227–238. DOI: 10.1016/S0167-7152(00)00236-4.

[54] F. Comte and F. Merlevède. "Adaptive estimation of the stationary density of discrete and continuous time mixing processes". In: vol. 6. New directions in time series analysis (Luminy, 2001). 2002, pp. 211–238. DOI: 10.1051/ps:2002012.

[55] G. Da Prato, K. D. Elworthy, and J. Zabczyk. "Strong Feller property for stochastic semilinear equations". In: *Stochastic Analysis and Applications* 13.1 (1995), pp. 35–45.

[56] G. Da Prato and J. Zabczyk. "Smoothing properties of transition semigroups in Hilbert spaces". In: *Stochastics and Stochastic Reports* 35.2 (1991), pp. 63–77.

[57] A. S. Dalalyan and Y. A. Kutoyants. "Asymptotically efficient trend coefficient estimation for ergodic diffusion". In: *Math. Methods Statist.* 11.4 (2002), 402–427 (2003).

[58] A. Dalalyan. "Sharp adaptive estimation of the drift function for ergodic diffusions". In: *Ann. Statist.* 33.6 (2005), pp. 2507–2528.

[59] A. Dalalyan and M. Reiß. "Asymptotic statistical equivalence for ergodic diffusions: the multidimensional case". In: *Probab. Theory Related Fields* 137.1-2 (2007), pp. 25–47. DOI: 10.1007/s00440-006-0502-7.

[60] A. Dalalyan and M. Reiß. "Asymptotic statistical equivalence for scalar ergodic diffusions". In: *Probab. Theory Relat. Fields* 134.2 (2006), pp. 248–282.

[61] M. H. A. Davis. *Markov models and optimization*. Vol. 49. Monographs on Statistics and Applied Probability. Chapman & Hall, London, 1993, pp. xiv+295. DOI: 10.1007/978-1-4899-4483-2.

[62] J. A. Davydov. "Mixing conditions for Markov chains". In: *Teor. Verojatnost. i Primenen.* 18 (1973), pp. 321–338.

[63] J. Dedecker, P. Doukhan, G. Lang, J. R. León R., S. Louhichi, and C. Prieur. *Weak dependence: with examples and applications*. Vol. 190. Lecture Notes in Statistics. Springer, New York, 2007, pp. xiv+318.

[64] J. Dedecker and S. Gouëzel. "Subgaussian concentration inequalities for geometrically ergodic Markov chains". In: *Electron. Commun. Probab.* 20 (2015), no. 64, 12. DOI: 10.1214/ECP.v20-3966.

[65] C. Dellacherie and P.-A. Meyer. *Probabilités et potentiel. Chapitres V à VIII*. Revised. Vol. 1385. Actualités Scientifiques et Industrielles. Théorie des martingales. Hermann, Paris, 1980, pp. xviii+476.

[66] S. Dereich, L. Döring, and A. E. Kyprianou. "Real self-similar processes started from the origin". In: *Ann. Probab.* 45.3 (2017), pp. 1952–2003. DOI: 10.1214/16-AOP1105.

[67] N. Dexheimer, C. Strauch, and L. Trottner. *Mixing it up: A general framework for Markovian statistics*. 2020. arXiv: 2011.00308.

[68] S. Dirksen. "Tail bounds via generic chaining". In: *Electron. J. Probab.* 20 (2015), no. 53, 29. DOI: 10.1214/EJP.v20-3760.

[69] W. Doeblin. *Œuvres Complètes—Collected Works*. Ed. by M. Yor and B. Bru. Springer International Publishing, 2020.

[70] R. A. Doney and R. A. Maller. "Stability of the overshoot for Lévy processes". In: *Ann. Probab.* 30.1 (2002), pp. 188–212. DOI: 10.1214/aop/1020107765.

[71] L. Döring and L. Trottner. *Stability of overshoots of Markov additive processes*. 2020.

[72] R. Douc, G. Fort, and A. Guillin. "Subgeometric rates of convergence of f-ergodic strong Markov processes". In: *Stochastic Process. Appl.* 119.3 (2009), pp. 897–923. DOI: 10.1016/j.spa.2008.03.007.

[73] D. Down and S. P. Meyn. "Piecewise linear test functions for stability and instability of queueing networks". In: *Queueing Systems Theory Appl.* 27.3-4 (1997), 205–226 (1998). DOI: 10.1023/A:1019166115653.

[74] D. Down, S. P. Meyn, and R. L. Tweedie. "Exponential and uniform ergodicity of Markov processes". In: *Ann. Probab.* 23.4 (1995), pp. 1671–1691.

[75] N. H. Du, N. T. Dieu, and N. N. Nhu. "Conditions for permanence and ergodicity of certain SIR epidemic models". In: *Acta Appl. Math.* 160 (2019), pp. 81–99. DOI: 10.1007/s10440-018-0196-8.

[76] A. Durmus and E. Moulines. "High-dimensional Bayesian inference via the unadjusted Langevin algorithm". In: *Bernoulli* 25.4A (2019), pp. 2854–2882. DOI: 10.3150/18-BEJ1073.

[77] C. Duval and E. Mariucci. *Non-asymptotic control of the cumulative distribution function of Lévy processes*. 2020. arXiv: 2003.09281.

[78] J.-P. Eckmann and M. Hairer. "Uniqueness of the invariant measure for a stochastic PDE driven by degenerate noise". In: *Comm. Math. Phys.* 219.3 (2001), pp. 523–565. DOI: 10.1007/s002200100424.

[79] S. N. Ethier and T. G. Kurtz. *Markov processes - Characterization and convergence*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, 1986.

[80] J. Fan, Z. Wang, Y. Xie, and Z. Yang. *A Theoretical Analysis of Deep Q-Learning*. 2020. arXiv: 1901.00137v3.

[81] C. Fuchs. *Inference for diffusion processes – With applications in life sciences*. Springer, Heidelberg, 2013, pp. xx+430. DOI: 10.1007/978-3-642-25969-2.

[82] F. Gao, A. Guillin, and L. Wu. "Bernstein type's concentration inequalities for symmetric Markov processes". In: *Teor. Veroyatnost. i Primenen* 58.3 (2013), pp. 521–549.

[83] R. K. Getoor. "Excursions of a Markov process". In: *Ann. Probab.* 7.2 (1979), pp. 244–266.

[84] I. I. Gihman and A. V. Skorohod. *Stochastic differential equations*. Translated from the Russian by Kenneth Wickwire, Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 72. Springer-Verlag, New York-Heidelberg, 1972, pp. viii+354.

[85] E. Giné and R. Nickl. "An exponential inequality for the distribution function of the kernel density estimator, with applications to adaptive estimation". In: *Probab. Theory Relat. Fields* 143.3-4 (2009), pp. 569–596.

[86] E. Giné and R. Nickl. *Mathematical foundations of infinite-dimensional statistical models*. Cambridge Series in Statistical and Probabilistic Mathematics, [40]. Cambridge University Press, New York, 2016, pp. xiv+690. DOI: 10.1017/CBO9781107337862.

[87] M. Giordano and K. Ray. *Nonparametric Bayesian inference for reversible multi-dimensional diffusions*. 2020. arXiv: 2012.12083.

[88] U. Grenander and M. I. Miller. "Representations of knowledge in complex systems". In: *J. Roy. Statist. Soc. Ser. B* 56.4 (1994). With discussion and a reply by the authors, pp. 549–603.

[89] P. S. Griffin. "Sample path behavior of a Lévy insurance risk process approaching ruin, under the Cramér-Lundberg and convolution equivalent conditions". In: *Ann. Appl. Probab.* 26.1 (2016), pp. 360–401. DOI: 10.1214/14-AAP1094.

[90] M. Hairer and J. Mattingly. "The strong Feller property for singular stochastic PDEs". In: *Ann. Inst. Henri Poincaré Probab. Stat.* 54.3 (2018), pp. 1314–1340. DOI: 10.1214/17-AIHP840.

[91] J. Hawkes. "Potential theory of Lévy processes". In: *Proc. London Math. Soc. (3)* 38.2 (1979), pp. 335–352. DOI: 10.1112/plms/s3-38.2.335.

[92] A. Hening, D. H. Nguyen, S. C. Ungureanu, and T. K. Wong. "Asymptotic harvesting of populations in random environments". In: *J. Math. Biol.* 78.1-2 (2019), pp. 293–329. DOI: 10.1007/s00285-018-1275-1.

[93] K. Ichihara and H. Kunita. "A classification of the second order degenerate elliptic operators and its probabilistic characterization". In: *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* 30 (1974), pp. 235–254. DOI: 10.1007/BF00533476.

[94] K. Itô. *Poisson point processes and their application to Markov processes*. SpringerBriefs in Probability and Mathematical Statistics. Springer, Singapore, 2015, pp. xi+43. DOI: 10.1007/978-981-10-0272-4.

[95] J. Ivanovs. "One-sided Markov Additive Processes and Related Exit Problems". PhD thesis. University of Amsterdam, 2007.

[96] J. Ivanovs, O. Boxma, and M. Mandjes. "Singularities of the matrix exponent of a Markov additive process with one-sided jumps". In: *Stochastic Process. Appl.* 120.9 (2010), pp. 1776–1794. DOI: 10.1016/j.spa.2010.05.007.

[97] J. Jacod. "Théorème de renouvellement et classification pour les chaînes semi-markoviennes". In: *Ann. Inst. H. Poincaré Sect. B (N.S.)* 7 (1971), pp. 83–129.

[98] J. Jacod and A. N. Shiryaev. *Limit theorems for stochastic processes*. 2nd ed. Vol. 288. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Springer-Verlag, Berlin, 2003, pp. xx+661. DOI: 10.1007/978-3-662-05265-5.

[99] G. Jongbloed, F. H. van der Meulen, and A. W. van der Vaart. "Nonparametric inference for Lévy-driven Ornstein-Uhlenbeck processes". In: *Bernoulli* 11.5 (2005), pp. 759–791. DOI: 10.3150/bj/1130077593.

[100] O. Kallenberg. *Foundations of modern probability*. 2nd ed. Probability and its Applications (New York). Springer-Verlag, New York, 2002, pp. xx+638. DOI: 10.1007/978-1-4757-4015-8.

[101] H. Kaspi and A. Mandelbaum. "On Harris recurrence in continuous time". In: *Math. Oper. Res.* 19.1 (1994), pp. 211–222. DOI: 10.1287/moor.19.1.211.

[102] H. Kesten. "Renewal theory for functionals of a Markov chain with general state space". In: *Ann. Probability* 2 (1974), pp. 355–386. DOI: 10.1214/aop/1176996654.

[103] P. Kevei. "Ergodic properties of generalized Ornstein-Uhlenbeck processes". In: *Stochastic Process. Appl.* 128.1 (2018), pp. 156–181. DOI: 10.1016/j.spa.2017.04.010.

[104] S. W. Kiu. "Semistable Markov processes in $\mathbf{R}^n$". In: *Stochastic Process. Appl.* 10.2 (1980), pp. 183–191. DOI: 10.1016/0304-4149(80)90020-4.

[105] C. Klüppelberg, A. E. Kyprianou, and R. A. Maller. "Ruin probabilities and overshoots for general Lévy insurance risk processes". In: *Ann. Appl. Probab.* 14.4 (2004), pp. 1766–1801. DOI: 10.1214/105051604000000927.

[106] Y. A. Kutoyants. *Statistical Inference for Ergodic Diffusion Processes*. Springer Series in Statistics. New York: Springer, 2004.

[107] A. Kuznetsov and J. C. Pardo. "Fluctuations of stable processes and exponential functionals of hypergeometric Lévy processes". In: *Acta Appl. Math.* 123 (2013), pp. 113–139. DOI: 10.1007/s10440-012-9718-y.

[108] A. E. Kyprianou. "Deep factorisation of the stable process". In: *Electron. J. Probab.* 21 (2016), Paper No. 23, 28. DOI: 10.1214/16-EJP4506.

[109] A. E. Kyprianou. *Fluctuations of Lévy processes with applications*. 2nd ed. Universitext. Introductory lectures. Springer, Heidelberg, 2014, pp. xviii+455. DOI: 10.1007/978-3-642-37632-0.

[110] A. E. Kyprianou, V. Rivero, B. Şengül, and T. Yang. "Entrance laws at the origin of self-similar Markov processes in high dimensions". In: *Trans. Amer. Math. Soc.* 373.9 (2020), pp. 6227–6299. DOI: 10.1090/tran/8086.

[111] S. P. Lalley. "Conditional Markov renewal theory. I. Finite and denumerable state space". In: *Ann. Probab.* 12.4 (1984), pp. 1113–1148.

[112] R. Lande, S. Engen, and B.-E. Saether. "Optimal harvesting, economic discounting and extinction risk in fluctuating populations". In: *Nature* 372.6501 (1994), pp. 88–90.

[113] F. Leblanc. "Density estimation for a class of continuous time processes". In: *Math. Methods Statist.* 6.2 (1997), pp. 171–199.

[114] M. Lemanczyk. "General Bernstein-Like Inequality for Additive Functionals of Markov Chains". In: *Journal of Theoretical Probability* (2020). DOI: 10.1007/s10959-020-01006-z.

[115] O. Lepski. "Multivariate density estimation under sup-norm loss: oracle approach, adaptation and independence structure". In: *Ann. Statist.* 41.2 (2013), pp. 1005–1034. DOI: 10.1214/13-AOS1109.

[116] P. Lezaud. "Chernoff and Berry–Esséen inequalities for Markov processes". In: *ESAIM: Probability and Statistics* 5 (2001), pp. 183–201.

[117] G. Liang and M. Zervos. *Ergodic singular stochastic control motivated by the optimal sustainable exploitation of an ecosystem*. 2020. arXiv: 2008.05576.

[118] T. M. Liggett. *Continuous time Markov processes*. Vol. 113. Graduate Studies in Mathematics. An introduction. American Mathematical Society, Providence, RI, 2010, pp. xii+271. DOI: 10.1090/gsm/113.

[119] A. Locatelli and A. Carpentier. "Adaptivity to Smoothness in $\mathcal{X}$-armed bandits". In: *Proceedings of the 31st Conference On Learning Theory*. Ed. by S. Bubeck, V. Perchet, and P. Rigollet. Vol. 75. Proceedings of Machine Learning Research. PMLR, 2018, pp. 1463–1492.

[120] E. Löcherbach, D. Loukianova, and O. Loukianov. "Polynomial bounds in the ergodic theorem for one-dimensional diffusions and integrability of hitting times". In: *Ann. Inst. Henri Poincaré Probab. Stat.* 47.2 (2011), pp. 425–449. DOI: 10.1214/10-AIHP359.

[121] M. N. Malyshkin. "Subexponential estimates for the rate of convergence to the invariant measure for stochastic differential equations". In: *Teor. Veroyatnost. i Primenen.* 45.3 (2000), pp. 489–504. DOI: 10.1137/S0040585X97978403.

[122] B. Maslowski. "Strong Feller property for semilinear stochastic evolution equations and applications". In: *Stochastic systems and optimization (Warsaw, 1988)*. Vol. 136. Lect. Notes Control Inf. Sci. Springer, Berlin, 1989, pp. 210–224. DOI: 10.1007/BFb0002683.

[123] H. Masuda. "Ergodicity and exponential β-mixing bounds for multidimensional diffusions with jumps". In: *Stochastic Processes and their Applications* 117.1 (2007), pp. 35 –56. DOI: https://doi.org/10.1016/j.spa.2006.04.010.

[124] H. Masuda. "Erratum to: "Ergodicity and exponential β -mixing bound for multidimensional diffusions with jumps" [Stochastic Process. Appl. 117 (2007) 35–56]". In: *Stochastic Processes and their Applications* 119.2 (2009), pp. 676 –678. DOI: https://doi.org/10.1016/j.spa.2008.02.010.

[125] H. Masuda. "On multidimensional Ornstein-Uhlenbeck processes driven by a general Lévy process". In: *Bernoulli* 10.1 (2004), pp. 97–120. DOI: 10.3150/bj/1077544605.

[126] F. Merlevède, M. Peligrad, and E. Rio. "Bernstein inequality and moderate deviations under strong mixing conditions". In: *High dimensional probability V: the Luminy volume*. Vol. 5. Inst. Math. Stat. (IMS) Collect. Inst. Math. Statist., Beachwood, OH, 2009, pp. 273–292. DOI: 10.1214/09-IMSCOLL518.

[127] S. P. Meyn and R. L. Tweedie. "Generalized resolvents and Harris recurrence of Markov processes". In: *Doeblin and modern probability (Blaubeuren, 1991)*. Vol. 149. Contemp. Math. Amer. Math. Soc., Providence, RI, 1993, pp. 227–250. DOI: 10.1090/conm/149/01273.

[128] S. Meyn and R. L. Tweedie. *Markov chains and stochastic stability*. Second. With a prologue by Peter W. Glynn. Cambridge University Press, Cambridge, 2009, pp. xxviii+594. DOI: 10.1017/CBO9780511626630.

[129] S. P. Meyn and R. L. Tweedie. "Stability of Markovian processes. I. Criteria for discrete-time chains". In: *Adv. in Appl. Probab.* 24.3 (1992), pp. 542–574. DOI: 10.2307/1427479.

[130] S. P. Meyn and R. L. Tweedie. "Stability of Markovian processes. II. Continuous-time processes and sampled chains". In: *Adv. in Appl. Probab.* 25.3 (1993), pp. 487–517. DOI: 10.2307/1427521.

[131] S. P. Meyn and R. L. Tweedie. "Stability of Markovian processes. III. Foster-Lyapunov criteria for continuous-time processes". In: *Adv. in Appl. Probab.* 25.3 (1993), pp. 518–548. DOI: 10.2307/1427522.

[132] S. P. Meyn and R. L. Tweedie. "Stability of Markovian processes. III. Foster-Lyapunov criteria for continuous-time processes". In: *Adv. in Appl. Probab.* 25.3 (1993), pp. 518–548. DOI: 10.2307/1427522.

[133] R. Nickl and K. Ray. "Nonparametric statistical inference for drift vector fields of multi-dimensional diffusions". In: *Ann. Statist.* 48.3 (2020), pp. 1383–1408. DOI: 10.1214/19-AOS1851.

[134] E. Nummelin. *General irreducible Markov chains and nonnegative operators*. Vol. 83. Cambridge Tracts in Mathematics. Cambridge University Press, Cambridge, 1984, pp. xi+156. DOI: 10.1017/CBO9780511526237.

[135] B. Øksendal and A. Sulem. *Applied stochastic control of jump diffusions*. Universitext. 3: Springer, Cham, 2019, pp. xvi+436. DOI: 10.1007/978-3-030-02781-0.

[136] H. S. Park and R. Maller. "Moment and MGF convergence of overshoots and undershoots for Lévy insurance risk processes". In: *Adv. in Appl. Probab.* 40.3 (2008), pp. 716–733. DOI: 10.1239/aap/1222868183.

[137] J. Paulsen. "Sharp conditions for certain ruin in a risk process with stochastic return on investments". In: *Stochastic Process. Appl.* 75.1 (1998), pp. 135–148. DOI: 10.1016/S0304-4149(98)00012-X.

[138] V. H. de la Peña and E. Giné. *Decoupling: From dependence to independence*. Probability and its Applications (New York). Springer-Verlag, New York, 1999, pp. xvi+392. DOI: 10.1007/978-1-4612-0537-1.

[139] S. Peszat and J. Zabczyk. "Strong Feller property and irreducibility for diffusions on Hilbert spaces". In: *Ann. Probab.* 23.1 (1995), pp. 157–172.

[140] Z. Qian and W. Zheng. "A representation formula for transition probability densities of diffusions and applications". In: *Stochastic Process. Appl.* 111.1 (2004), pp. 57–76. DOI: 10.1016/j.spa.2003.12.004.

[141] D. Revuz and M. Yor. *Continuous Martingales and Brownian Motion*. 3rd. Vol. 293. Grundlehren der mathematischen Wissenschaften. Berlin: Springer, 1999.

[142] E. Rio. *Asymptotic theory of weakly dependent random processes*. Vol. 80. Probability Theory and Stochastic Modelling. Translated from the 2000 French edition [MR2117923]. Springer, Berlin, 2017, pp. xviii+204. DOI: 10.1007/978-3-662-54323-8.

[143] E. Rio. *Théorie asymptotique des processus aléatoires faiblement dépendants*. Mathématiques & Applications 31. Berlin: Springer, 2000.

[144] M. Rosenbaum and P. Tankov. "Asymptotic results for time-changed Lévy processes sampled at hitting times". In: *Stochastic Process. Appl.* 121.7 (2011), pp. 1607–1632. DOI: 10.1016/j.spa.2011.03.013.

[145] P.-M. Samson. "Concentration of measure inequalities for Markov chains and Φ-mixing processes". In: *Ann. Probab.* 28.1 (2000), pp. 416–461. DOI: 10.1214/aop/1019160125.

[146] N. Sandrić. "A note on the Birkhoff ergodic theorem". In: *Results Math.* 72.1-2 (2017), pp. 715–730. DOI: 10.1007/s00025-017-0681-9.

[147] K. Sato. *Lévy processes and infinitely divisible distributions*. 2nd ed. Vol. 68. Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge, 2013, pp. xiv+521.

[148] K. Sato and M. Yamazato. "Operator-self-decomposable distributions as limit distributions of processes of Ornstein-Uhlenbeck type". In: *Stochastic Process. Appl.* 17.1 (1984), pp. 73–100. DOI: 10.1016/0304-4149(84)90312-0.

[149] R. L. Schilling. "Conservativeness and extensions of Feller semigroups". In: *Positivity* 2.3 (1998), pp. 239–256. DOI: 10.1023/A:1009748105208.

[150] R. L. Schilling and L. Partzsch. *Brownian motion*. 2nd ed. De Gruyter Graduate. An introduction to stochastic processes. De Gruyter, Berlin, 2014, pp. xvi+408. DOI: 10.1515/9783110307306.

[151] R. L. Schilling and J. Wang. "Strong Feller continuity of Feller processes and semigroups". In: *Infin. Dimens. Anal. Quantum Probab. Relat. Top.* 15.2 (2012), pp. 1250010, 28. DOI: 10.1142/S0219025712500105.

[152] M. Sharpe. *General theory of Markov processes*. Vol. 133. Pure and Applied Mathematics. Academic Press Inc., Boston, MA, 1988, pp. xii+419.

[153] S. J. Sheu. "Some estimates of the transition density of a nondegenerate diffusion Markov process". In: *Ann. Probab.* 19.2 (1991), pp. 538–561.

[154] Y. Shimizu. "Density estimation of Lévy measures for discretely observed diffusion processes with jumps". In: *J. Japan Statist. Soc.* 36.1 (2006), pp. 37–62. DOI: 10.14490/jjss.36.37.

[155] Y. Shimizu. "Functional estimation for Lévy measures of semimartingales with Poissonian jumps". In: *J. Multivariate Anal.* 100.6 (2009), pp. 1073–1092. DOI: 10.1016/j.jmva.2008.10.006.

[156] T. Sohr. "Contributions to Optimal Stopping and Long-Term Average Impulse Control". PhD thesis. https://macau.uni-kiel.de/receive/macau_mods_00000762: CAU Kiel, 2020.

[157] R. Stephenson. "On the exponential functional of Markov additive processes, and applications to multi-type self-similar fragmentation processes and trees". In: *ALEA Lat. Am. J. Probab. Math. Stat.* 15.2 (2018), pp. 1257–1292. DOI: 10.30757/alea.v15-47.

[158] C. Strauch. "Adaptive invariant density estimation for ergodic diffusions over anisotropic classes". In: *Ann. Statist.* 46.6B (2018), pp. 3451–3480. DOI: 10.1214/17-AOS1664.

[159] C. Strauch. "Exact adaptive pointwise drift estimation for multidimensional ergodic diffusions". In: *Probab. Theory Related Fields* 164.1-2 (2016), pp. 361–400. DOI: 10.1007/s00440-014-0614-4.

[160] C. Strauch. "Sharp adaptive drift estimation for ergodic diffusions: The multivariate case". In: *Stochastic Process. Appl.* 125.7 (2015), pp. 2562 –2602.

[161] Y. Tamura and S. Yamada. "Reliability Analysis Based on a Jump Diffusion Model with Two Wiener Processes for Cloud Computing with Big Data". In: *Entropy* 17.7 (2015), pp. 4533–4546. DOI: 10.3390/e17074533.

[162] H. Thorisson. "Coupling methods in probability theory". In: *Scandinavian journal of statistics* (1995), pp. 159–182.

[163] A. B. Tsybakov. *Introduction to nonparametric estimation*. Springer Series in Statistics. Revised and extended from the 2004 French original, Translated by Vladimir Zaiats. Springer, New York, 2009, pp. xii+214. DOI: 10.1007/b13794.

[164] R. L. Tweedie. "Topological conditions enabling use of Harris methods in discrete and continuous time". In: *Acta Appl. Math.* 34.1-2 (1994), pp. 175–188. DOI: 10.1007/BF00994264.

[165] B. Tzen and M. Raginsky. *Neural Stochastic Differential Equations: Deep Latent Gaussian Models in the Diffusion Limit*. 2019. arXiv: 1905.09883.

[166] A. Y. Veretennikov. "On Castellana-Leadbetter's condition for diffusion density estimation". In: *Stat. Inference Stoch. Process.* 2.1 (1999), 1–9 (2000). DOI: 10.1023/A:1009996608986.

[167] A. Y. Veretennikov. "The mixing rate and the averaging principle for hypoelliptic stochastic differential equations". In: *Izv. Akad. Nauk SSSR Ser. Mat.* 52.5 (1988), pp. 899–908, 1118. DOI: 10.1070/IM1989v033n02ABEH000824.

[168] A. Y. Veretennikov and S. A. Klokov. "On the subexponential rate of mixing for Markov processes". In: *Teor. Veroyatn. Primen.* 49.1 (2004), pp. 21–35.

[169] A. Y. Veretennikov. "On polynomial mixing bounds for stochastic differential equations". In: *Stoch. Processes Appl.* 70 (1997), pp. 115–127.

[170] G. Viennet. "Inequalities for absolutely regular sequences: application to density estimation". In: *Probab. Theory Related Fields* 107.4 (1997), pp. 467–492. DOI: 10.1007/s004400050094.

[171] V. Vigon. "Simplifiez vos Lévy en titillant la factorisation de Wiener–Hopf". PhD thesis. INSA de Rouen, 2002.

[172] V. Vigon. "Votre Lévy rampe-t-il?" In: *J. London Math. Soc. (2)* 65.1 (2002), pp. 243–256. DOI: 10.1112/S0024610701002885.

[173] V. A. Volkonskiĭ and J. A. Rozanov. "Some limit theorems for random functions. II". In: *Teor. Verojatnost. i Primenen.* 6 (1961), pp. 202–215.

[174] F. Xi and G. Yin. "The strong Feller property of switching jump-diffusion processes". In: *Statist. Probab. Lett.* 83.3 (2013), pp. 761–767. DOI: 10.1016/j.spl.2012.11.021.

[175] F. Xi and C. Zhu. "Jump type stochastic differential equations with non-Lipschitz coefficients: non-confluence, Feller and strong Feller properties, and exponential ergodicity". In: *J. Differential Equations* 266.8 (2019), pp. 4668–4711. DOI: 10.1016/j.jde.2018.10.006.

[176] C. Zhu and G. Yin. "On strong Feller, recurrence, and weak stabilization of regime-switching diffusions". In: *SIAM J. Control Optim.* 48.3 (2009), pp. 2003–2031. DOI: 10.1137/080712532.

[177] J. D. Zund. "George David Birkhoff and John von Neumann: a question of priority and the ergodic theorems, 1931–1932". In: *Historia Math.* 29.2 (2002), pp. 138–156. DOI: 10.1006/hmat.2001.2338.

Index of Notation

$\mathbb{R}_+$	$[0,\infty)$
$\boldsymbol{\lambda}$	Lebesgue measure on $\mathbb{R}^d$
$\boldsymbol{\lambda}_+$	Lebesgue measure on $\mathbb{R}_+$
$\langle x,y\rangle$	scalar product of elements x,y from a Hilbert space $\mathcal{H}$
$\mathcal{H}'$	dual space of a Hilbert space $\mathcal{H}$
$(\mathcal{X},\mathcal{B}(\mathcal{X}))$	topological space $\mathcal{X}$ with Borel σ-algebra $\mathcal{B}(\mathcal{X})$
$f\in\mathcal{B}(\mathcal{X})$	function $f\colon\mathcal{X}\to\mathbb{R}$ is $\mathcal{B}(\mathcal{X})$-measurable
$\mathcal{B}_b(\mathcal{X})$	space of bounded, $\mathcal{B}(\mathcal{X})$-measurable functions $f\colon\mathcal{X}\to\mathbb{R}$
$\mathcal{B}_+(\mathcal{X})$	space of non-negative, $\mathcal{B}(\mathcal{X})$-measurable functions $f\colon\mathcal{X}\to\mathbb{R}$
$\mathcal{B}^+(\mathcal{X})$	family of accessible sets for a ψ-irreducible Markov process $\boldsymbol{X}$ with state space $\mathcal{X}$
$\mathcal{C}(\mathcal{X})$	space of continuous functions $f\colon\mathcal{X}\to\mathbb{R}$
$\mathcal{C}_b(\mathcal{X})$	space of bounded, continuous functions $f\colon\mathcal{X}\to\mathbb{R}$
$\mathcal{C}_0(\mathcal{X})$	space of continuous functions $f\colon\mathcal{X}\to\mathbb{R}$ vanishing at infinity
$\mathcal{C}^k(\mathbb{R}^d)$	space of k-times continuously differentiable functions $f\colon\mathbb{R}^d\to\mathbb{R}$
$\mathcal{C}^k_b(\mathbb{R}^d)$	space of k-times continuously differentiable functions $f\colon\mathbb{R}^d\to\mathbb{R}$ such that $\lVert f^{(l)}\rVert_\infty<\infty$ for all $l=0,\dots,k$
$\mathcal{S}(\mathbb{R})$	Schwartz space of rapidly decreasing functions $f\colon\mathbb{R}\to\mathbb{C}$
$\mathcal{S}'(\mathbb{R})$	space of tempered distributions
$L^p(\mathcal{X},\mu)$	L^p-space of functions $f\in\mathcal{B}(\mathcal{X})$ s.t. $\int_{\mathcal{X}}\lvert f(x)\rvert^p\,\mu(\mathrm{d}x)<\infty$ for some measure μ on a given measurable space $(\mathcal{X},\mathcal{B}(\mathcal{X}))$ and $p\geq 1$
$\mu(f)$	$\int_{\mathcal{X}} f(x)\,\mu(\mathrm{d}x)$ for $f\in L^1(\mathcal{X},\mu)\cup\mathcal{B}_+(\mathcal{X})$
$\bar{\mu}$	positive tail of a measure μ on $(\mathbb{R},\mathcal{B}(\mathbb{R}))$, defined by $\bar{\mu}(x)=\mu((x,\infty))$ for $x\geq 0$
$(\mathcal{A},\mathcal{D}(\mathcal{A}))$	extended generator $\mathcal{A}$ with domain $\mathcal{D}(\mathcal{A})$ for Borel right process $\boldsymbol{X}$
$(\widetilde{\mathcal{A}},\mathcal{D}(\widetilde{\mathcal{A}}))$	infinitesimal generator $\widetilde{\mathcal{A}}$ with domain $\mathcal{D}(\widetilde{\mathcal{A}})$ for Feller process $\boldsymbol{X}$
$(U_\lambda)_{\lambda>0}$	resolvent of a Markov process
$\int_0^t H_s\,\mathrm{d}X_s$	stochastic integral up to time $t\geq 0$ wrt. semimartingale $(X_s)_{s\geq 0}$ for appropriate predictable process $(H_s)_{s\geq 0}$
$\lVert\nu\rVert_{\mathrm{TV}}$	total variation norm of a signed measure ν
$\lVert f\rVert_\infty$	sup-norm of a function $f\colon\mathcal{X}\to\mathbb{R}$ on normed space $\mathcal{X}$
$f(T)\in\mathrm{O}(g(T))$	$\limsup_{t\to\infty}\left\lvert\frac{f(T)}{g(T)}\right\rvert<\infty$ for functions $f,g\colon\mathbb{R}_+\to\mathbb{R}$
$f(T)\in\mathrm{o}(g(T))$	$\limsup_{t\to\infty}\left\lvert\frac{f(T)}{g(T)}\right\rvert=0$ for functions $f,g\colon\mathbb{R}_+\to\mathbb{R}$
$x\lesssim y$	$\exists C>0\colon\ x\leq Cy$