Electronic Journal of Statistics
Vol. 18 (2024) 191–246
ISSN: 1935-7524
https://doi.org/10.1214/23-EJS2199

Strong invariance principles for ergodic Markov processes

Ardjen Pengel and Joris Bierkens

Delft Institute of Applied Mathematics, Delft University of Technology
Mekelweg 4, 2628 CD Delft, the Netherlands
e-mail: a.l.pengel@tudelft.nl; joris.bierkens@tudelft.nl

Abstract: Strong invariance principles describe the error term of a Brownian approximation to the partial sums of a stochastic process. While these strong approximation results have many applications, results for continuous-time settings have been limited. In this paper, we obtain strong invariance principles for a broad class of ergodic Markov processes. Strong invariance principles provide a unified framework for analysing commonly used estimators of the asymptotic variance in settings with a dependence structure. We demonstrate how this can be used to analyse the batch means method for simulation output of Piecewise Deterministic Monte Carlo samplers. We also derive a fluctuation result for additive functionals of ergodic diffusions using our strong approximation results.

MSC2020 subject classifications: Primary 65C05; secondary 60J25.
Keywords and phrases: Strong invariance principle, piecewise deterministic Markov processes, asymptotic variance estimation.

Received June 2022.

Contents

1. Introduction

Let $X = (X_k)_{k\in\mathbb{N}}$ be a stochastic sequence defined on a common probability space and consider the partial sum process S_n, given by $S_n = \sum_{k=1}^n X_k$. We say that a strong invariance principle holds for X if there exist a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ on which we can construct a sequence of random variables $X' = (X'_k)_{k\in\mathbb{N}}$ and a Brownian motion $W = (W(t))_{t\geq 0}$, such that X and X' are equal in law and

$$|S'_n - \mu n - \sigma W(n)| = \mathscr{O}(\psi_n) \quad \text{a.s.},$$

where S'_n denotes the partial sum process of X', μ and σ are finite constants determined by the law of the process, $\mathscr{O}$ is a placeholder for the asymptotic regime, and ψ_n the corresponding approximation error. More specifically, if $S = (S_t)_{t\geq 0}$ denotes a stochastic process and $\psi = (\psi_t)_{t\geq 0}$ is some positive sequence, we write

$$S_T = o(\psi_T) \text{ a.s.} \quad \text{and} \quad S_T = O(\psi_T) \text{ a.s.}$$

to denote

$$\mathbb{P}\left(\lim_{T\to\infty} S_T/\psi_T = 0\right) = 1 \quad \text{and} \quad \mathbb{P}\left(\limsup_{T\to\infty} |S_T|/\psi_T < \infty\right) = 1$$

respectively. For technical convenience, we will usually make no distinction between X and X'.

For a sequence of independent, identically distributed random variables with mean zero and unit variance, the Komlós-Major-Tusnády approximation [51, 52] asserts that if $E|X_1|^p < \infty$ for some $p > 2$, then on a suitably enriched probability space, we can construct a Brownian motion $W = \{W(t), t \geq 0\}$ such that

$$S_n = W(n) + o(n^{1/p}) \quad \text{a.s.} \tag{1.1}$$

If we additionally assume that the moment-generating function exists in a neighbourhood of zero, i.e., $\mathbb{E}e^{t|X|} < \infty$ for some $t > 0$, then one can construct a Brownian motion W such that

$$S_n = W(n) + O(\log n) \quad \text{a.s.} \tag{1.2}$$

Furthermore, if only existence of the the second moment is assumed, [60] showed that there exists a sequence $t_n \sim n$ such that

$$S_n = W(t_n) + o(n^{1/2}) \quad \text{a.s.} \tag{1.3}$$

The error terms appearing in the strong invariance principles (1.1), (1.2), and (1.3) are optimal. The approximation error appearing in the strong invariance principle also quantifies the convergence rate in the functional central limit theorem, as shown in [20, Theorem 1.16 and Theorem 1.17]. These strong approximation results are powerful tools used to obtain numerous results in both probability and statistics, as seen in e.g., [23], [22], [72], and [81].

Naturally, it is of great interest to extend these results beyond the i.i.d. setting. An extensive overview of invariance principles for dependent sequences is given in [6]. In Markovian settings, strong approximation results were obtained by [24], [19], [89], and [61], among others. The strong invariance principle of [61] attains the Komlós-Major-Tusnády bound given in (1.2). The results of [19] and [61] are established through an application of Nummelin splitting, introduced in the seminal papers of [3] and [67]. Provided that the transition operator of the chain satisfies a one-step minorisation condition, a bivariate process can be constructed such that this process possesses a recurrent atom and the first coordinate of the constructed process is equal in law to the original Markov chain. Consequently, the chain inherits a regenerative structure and can thus be divided into independent identically distributed cycles. By application of the Komlós-Major-Tusnády approximations strong invariance principles can be obtained. Strong approximation results for Markov chains are useful tools for analysing estimators of the asymptotic variance of Markov Chain Monte Carlo (MCMC) sampling algorithms. The results of [25, 26], [38], and [89] show strong consistency of the batch means and spectral variance estimators for MCMC simulation output using the appropriate strong invariance principles.

Recently, there has been growing interest in Monte Carlo algorithms based on Piecewise Deterministic Markov Processes (PDMPs). The main appeal of these processes is their non-reversible nature. It is well known that non-reversibility can significantly improve performance of sampling methods, in terms of both convergence rate to equilibrium and asymptotic variance, see for example, the results of [48] and [55] regarding convergence to stationarity and [33] and [76] regarding the asymptotic variance. Furthermore, PDMPs have piecewise deterministic paths and can therefore be simulated without discretisation error, in contrast to for example Langevin and Hamiltonian dynamics. The primary sampling algorithms belonging to this class are the Zig-Zag Sampler and the Bouncy Particle Sampler, introduced in [10] and [12] respectively. Moreover, since these processes maintain the correct target distribution if sub-sampling is employed, they enjoy advantageous scaling properties to large datasets, as seen in [9].

For many useful results regarding the estimation of the asymptotic variance of Markov chain simulation output to carry over to PDMP-based methods, it is required that a strong invariance principle holds for the underlying continuous-time process. In this paper, we obtain strong approximation results for a broad class of (continuous-time) ergodic Markov processes. Firstly, we show that the strong invariance principle given in Theorem 4.1 can be obtained directly through ergodicity and moment conditions. However, the resulting error rate is not explicit and therefore less convenient to work with.

A natural approach for obtaining a more refined strong invariance principle would be through regenerative properties of the process. However, it is in general not possible to show that the transition semigroup satisfies a minorisation condition such that a regenerative structure can be obtained. The resolvent chain, on the other hand, does satisfy a one-step minorisation condition. Utilising this result, [58] extends the concept of Nummelin splitting to Harris recurrent Markov processes. Hence we can redefine the process such that it is embedded in a richer process which is endowed with a recurrent atom. Although the resulting cycles are not independent and we therefore do not have regeneration in the classic sense, we do obtain short-range dependence. Therefore we can utilise the approximation results of [5] to obtain a strong invariance principle attaining a convergence rate of order $O(T^{1/4}\log T)$. This result is formulated in Theorem 4.5 and covers a wide range of Markov processes including ergodic diffusions. Although the nearly optimal bound $O(T^{1/p}\log(T)^2)$ of [5] does not carry over, to the best of our knowledge, there are currently no approaches established that lead to superior rates for the class of processes considered in Theorem 4.5.

For PDMPs we are able to give a strong invariance principle with an improved approximation error. We show that the univariate Zig-zag process has regenerative cycles. This allows us to follow the approach of [61] such that the optimal strong approximation error of $O(T^{1/p})$ can be obtained. Moreover, if the target distribution factorises into a product of independent densities, the optimal approximation bound carries over to the multivariate settings. Furthermore, we also show that the results of [61] can be extended under less restrictive conditions such that the optimal approximation error (1.2) is still attained. Finally, we discuss some applications of our obtained strong invariance principles. We demonstrate how the obtained strong approximation results can be utilised for analysing the batch means estimator of the asymptotic variance of continuous-time Monte Carlo samplers. Theorem 5.2 weakens the existing regularity conditions guaranteeing strong convergence of the batch means estimator in an MCMC setting. This is a direct consequence of the fact that Theorems 4.6 and 4.7 obtain the optimal approximation rate of $O(T^{1/p})$ whereas previous work on estimation of the MCMC standard error is based on strong invariance principles with limited accuracy, which we further explain in Remark 5.3. Furthermore, we demonstrate the applicability of our results to diffusion processes and show that the magnitude of increments can be described with our obtained approximation results.

This article is organised as follows. In Section 2, we give a brief introduction of Piecewise Deterministic Markov processes and state our motivational example. In Section 3, we review Nummelin splitting in continuous time as introduced in [58] and discuss other relevant results. In Section 4, the main results of the paper are given. In Section 5, we discuss the estimation of the asymptotic variance for

PDMC simulation output. Section 6 illustrates the applicability of our results to diffusion processes. In Section 7, the proofs of the main results are given.

2. Motivating example: estimation of the piecewise deterministic Monte Carlo standard error

Suppose our goal is to sample from a probability distribution $\pi(dx)$ on $E = \mathbb{R}^k$, which admits Lebesgue density

$$\pi(x) = \frac{e^{-U(x)}}{\int_{\mathbb{R}^k} e^{-U(x)}\,dx}, \tag{2.1}$$

where U is referred to as the associated potential of the target π. We will assume that U is twice continuously differentiable and can be evaluated pointwise. Typically, the objective is to compute expectations with respect to this distribution, in other words, we are interested interested in $\pi(f) = \int f(x)\pi(dx)$, for some appropriately integrable function f.

Piecewise Deterministic Monte Carlo (PDMC) samplers consist of a position and a velocity component. We will consider processes $Z = (Z_t)_{t\geq 0}$ with $Z_t = (X_t, V_t)$, where X_t and V_t denote the position and velocity component respectively. Our process takes values in $E = \mathfrak{X} \times \mathcal{V}$, where $\mathfrak{X}$ denotes the state-space of the position component and $\mathcal{V}$ denotes the space of attainable velocities. Piecewise Deterministic Markov processes are characterised by their deterministic dynamics between random event times along with a Markov kernel that describes the transitions at events. More specifically, their deterministic dynamics are described by some ordinary differential equation. Both the Zig-Zag process and the Bouncy Particle sampler have piecewise linear trajectories characterised by

$$\frac{dX_t}{dt} = V_t \quad \text{and} \quad \frac{dV_t}{dt} = 0.$$

Thus the rate of change of the position is described by the velocity, whereas the velocity does not change along the deterministic dynamics. Changes in the velocity occur according to some inhomogeneous Poisson process of rate $\lambda(Z_t)$. The Poisson events consist of changes in the velocity component of our process. The fundamental idea behind these sampling methods is to choose the event rate and the changes in velocity such that the position component explores the state-space according to the target distribution π. The event rate should increase in an appropriate manner as the position moves towards regions of lower probability mass.

For the Zig-Zag process the set of possible velocities is given by $\mathcal{V} = \{-1, +1\}^d$. We distinguish d types of events for the Zig-Zag Sampler. For every dimension i of our position component, an event will consist of flipping component i of the velocity, while keeping the other $(d-1)$ components unchanged. More specifically, our transition at events can be described by $F_i : \mathcal{V} \to \mathcal{V}$, which is the mapping that flips the i-th component of the velocity, i.e., for $v \in \mathcal{V}$ we have that the k-th entry of $F_i(v)$ is given by

$$(F_i(v))_k = \begin{cases} -v_k & \text{for } k = i \\ v_k & \text{for } k \neq i, \end{cases}$$

where v_k denotes the k-th entry of the velocity v for $k = 1, \ldots, d$. A change in the i-th component of the velocity will be governed by an inhomogeneous Poisson process of rate λ_i. For the (canonical) Zig-Zag Sampler these rates are given by

$$\lambda_i(x, v) = \left(v_i \partial_{x_i} U(x)\right)^+, \tag{2.2}$$

where $(x)^+ := \max\{x, 0\}$. Hence for the Zig-Zag process events occur with rate

$$\lambda_Z(x, v) = \sum_{i=1}^{d} \lambda_i(x, v) = \sum_{i=1}^{d} \left(v_i \partial_{x_i} U(x)\right)^+. \tag{2.3}$$

The simulation scheme for Zig-Zag is given in Algorithm 1 below.

Algorithm 1 Zig-Zag Sampler

1: Initialise $(X_0, V_0) \leftarrow (x, v)$ and $T_0 \leftarrow 0$.
2: For $k = 1, 2, \ldots$ simulate $\tau_k^1, \ldots, \tau_k^d$ according to

$$\Pr(\tau_k^i \geq t) = \exp\left(-\int_0^t \lambda_i(X_{\tau_{k-1}} + sV_{\tau_{k-1}}, V_{\tau_{k-1}})\right) ds,$$

for $i = 1, \ldots, d$.
3: For $s \in (0, \tau_k)$ set $(X_{\tau_{k-1}+s}, V_{\tau_{k-1}+s}) \leftarrow (X_{\tau_{k-1}} + sV_{\tau_{k-1}}, V_{\tau_{k-1}})$.
4: The time of the k-th event is given by $T_k = T_{k-1} + \tau_k^{i_0}$, with $i_0 = \min_i \{\tau_k^i\}_{i=1}^d$.
5: Update velocity of component i_0 at the event time
$V_{T_k} = F_{i_0}(V_{T_{k-1}})$.

In [9] it is shown that if we have

$$\lambda_i(x, v) - \lambda_i(x, F_i(v)) = v_i \partial_{x_i} U(x), \text{ for } i = 1, \ldots, d,$$

then the Zig-Zag process has the desired invariant distribution given by $\pi(dx)\nu(dv)$, where the target distribution π is the marginal distribution of the position component and ν is a uniform distribution over the set of velocities $\mathcal{V}$. Consider the case when the target π is of product form, namely $\pi(x) = \prod_{i=1}^d \pi_i(x_i)$, where each π_i is a one-dimensional probability density. Then the Zig-Zag process with stationary distribution π can be defined through d independent one-dimensional Zig-Zag processes. The potential of the product form target is given by $U(x) = -\sum_{i=1}^d \log \pi_i(x_i)$, and therefore the corresponding Poisson event rates are given by

$$\lambda_i(x, v) = \left(-v_i \frac{\partial_{x_i} \pi_i(x_i)}{\pi_i(x_i)}\right)^+ = \left(v_i \partial_{x_i} U_i(x)\right)^+, \tag{2.4}$$

where $U_i(x) = -\log \pi_i(x_i)$. Because the switching intensity of every coordinate only depends on its own position and velocity, we see that the corresponding

Poisson processes are independent. Therefore it follows that the d-dimensional Zig-Zag process Z_t with target distribution π can be decomposed into d independent one-dimensional Zig-Zag processes $(Z_t^i)_{i=1}^d$, where every coordinate i moves according to Z_t^i which has target distribution π_i for $i = 1, \dots, d$.

For the simulation scheme of the Bouncy Particle Sampler we refer to [12]. In the one-dimensional case the canonical BPS and ZZS are described by the same PDMP. For a more detailed introduction to PDMP-based samplers we refer to [37]. It can be shown that under very mild regularity conditions both sampling processes admit a stationary distribution given by

$$\mu(dx, dv) = \pi(dx)\upsilon(dv), \tag{2.5}$$

where the target distribution π is the marginal distribution of the position component and ν is the marginal distribution of the velocity component. Moreover, an ergodic law of large numbers holds, i.e.,

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T g(X_s, V_s)\, ds = \int_E g(x, v)\mu(dx, dv) =: \mu(g),$$

for all μ-integrable g. Let f be a function such that $\pi(|f|) < \infty$, then from the independence of position and velocity at equilibrium, we see that $\frac{1}{T}\int_0^T f(X_s)ds$, the time average of the position component, is a natural estimator for $\pi(f)$, the expectation with respect to π. In order to assess the accuracy of our sampling method, we require a central limit theorem to hold;

$$\sqrt{T}\left(\frac{1}{T}\int_0^T f(X_s)ds - \pi(f)\right) \xrightarrow{d} \mathcal{N}(0, \Sigma_f) \quad \text{as } T \to \infty, \tag{2.6}$$

and estimate the corresponding asymptotic variance Σ_f. Moreover, the asymptotic variance is also useful for determining the efficiency of the sampling algorithm via measures such as the effective sample size, see for example [44] or [88]. The estimation of the asymptotic variance is also required for the implementation of stopping rules, which consists of justifiable criteria for termination of the simulation. In order to validate stopping rules that guarantee a desired level of precision, in [42] it is shown that the estimator of the asymptotic covariance matrix must be strongly consistent. Strong invariance principles play a central role in the analysis of estimators of the asymptotic variance of Markov Chain Monte Carlo (MCMC) sampling algorithms, see for example [25, 26], [38], and [89]. In this paper, we obtain strong approximation results for broad classes of ergodic Markov processes. We show that for PDMPs many results regarding estimation of the asymptotic variance immediately carry over.

3. Nummelin splitting in continuous time

Let $X = (X_t)_{t\geq 0}$ be a stochastic process defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P}_x)$, with Polish state space $(E, \mathscr{E})$ and initial value $X_0 = x$. We consider the case where X is a positive Harris recurrent strong Markov process with transition semigroup given by $(P_t)_{t\geq 0}$ with finite invariant measure π. By definition of positive Harris recurrence, π can be normalised to be a probability measure and we have that

$$\pi(A) > 0 \implies \mathbb{P}_x\left(\int_0^\infty \mathbb{1}_{\{X_s\in A\}}ds = \infty\right) = 1, \quad x \in E. \tag{3.1}$$

Throughout this paper we will additionally require ergodicity of the considered processes. We say that a Markov process X is *ergodic* with convergence rate Ψ if

$$\|P_t(x,\cdot) - \pi\|_{TV} \leq V(x)\Psi(t), \quad \text{for all } x \in E \text{ and } t \geq 0, \tag{3.2}$$

where V is some positive π-integrable function and Ψ some positive function that tends to zero as $t \to \infty$. Furthermore, a process is called polynomially or exponentially ergodic if Ψ decays at a polynomial rate $(1+t)^{-\beta}$ or exponential rate $e^{-\gamma t}$ respectively for some $\beta, \gamma > 0$. For a more thorough discussion of these definitions we refer to [62].

The resolvent chain $\bar{X} = (\bar{X}_n)_{n\geq 0}$ is obtained by observing the process at independent exponential times, i.e., $\bar{X}_n := X_{T_n}$ for $n \geq 0$. Here $(T_n)_{n\geq 0}$ denote the sampling times at which we observe the process X, which are defined as $T_0 := 0$ and $T_n := \sum_{k=1}^n \sigma_k$, where $(\sigma_k)_{k\geq 1}$ denote a sequence of i.i.d. standard exponential random variables with mean equal to one. The resolvent chain will inherit positive Harris recurrence from the original process, see for example [47, Thereom 1.4]. The transition kernel of the process $\bar{X} = (\bar{X}_n)_{n\in\mathbb{N}}$ is given by

$$U(x, A) = \int_0^\infty P_t(x, A)e^{-t}dt, \tag{3.3}$$

and satisfies the one-step minorisation condition, see for example [47] or [75],

$$U(x, A) \geq h \otimes \nu(x, A), \tag{3.4}$$

where $h \otimes \nu(x, A) = h(x)\nu(A)$, with $h(x) = \alpha\mathbb{1}_C(x)$ for some $\alpha \in (0, 1)$, a measurable set C with $\pi(C) > 0$, and $\nu(\cdot)$ a probability measure equivalent to $\pi(\cdot \cap C)$.

The minorisation condition of the resolvent chain motivates the introduction of the kernel $K((x, u), dy) : E \times [0, 1] \to E$ given by

$$K((x,u),dy) = \begin{cases} \nu(dy) & \text{for } (x,u) \in C \times [0,\alpha] \\ W(x,dy) & \text{for } (x,u) \in C \times (\alpha,1] \\ U(x,dy) & \text{for } x \notin C, \end{cases} \tag{3.5}$$

where the residual kernel $W(x, dy)$ is defined as

$$W(x, dy) = \frac{U(x, dy) - \alpha\nu(dy)}{1-\alpha}. \tag{3.6}$$

Since the resolvent chain is also positive Harris recurrent, it will hit C infinitely often. Given that the resolvent chain has hit C, with probability α the chain will move independently of its past according to the small measure ν and with probability $(1-\alpha)$ it will move according to the residual kernel W. By the Borel-Cantelli lemma the residual chain will move according to ν infinitely often. Let R_k denote the k-th time that the resolvent chain moves according to ν. The randomised stopping times $(R_k)_k$ serve as regeneration epochs for the resolvent; for every k, $\bar{X}_{R_k}$ has law ν and is independent of both its past and of R_k. The implied regenerative properties that the process X obtains through its resolvent are made explicit with the approach of [58]. Their framework requires the following regularity conditions on the transition semigroup of the process X:

Assumption 1.

(i) The semigroup $(P_t)_{t\geq 0}$ is Feller, i.e., for every bounded and continuous function f, the mapping $x \longmapsto P_t f(x) = \int_E P_t(x,dy)f(y)$ is bounded and continuous.

(ii) There exists a σ-finite measure Λ on $(E,\mathscr{E})$ such that for every $t>0$, $P_t(x,dy) = p_t(x,dy)\Lambda(dy)$, with $(t,x,y) \longmapsto p_t(x,y)$ jointly measurable.

Note that by Assumption 1 it follows that $U(x,dy)$, the transition kernel of the resolvent chain, also has a density with respect to $\Lambda(dy)$, which we will denote by $u(x,y)$. At the so-called sampling times of the process X, we can apply the Nummelin splitting technique to the resolvent chain. We then fill in the original process between the sampling times. Following this procedure, [58] construct on an extended probability space a process Z with state space $E\times[0,1]\times E$, that admits a recurrent atom. The first coordinate of Z has the same law as the original process X, the second coordinate denotes the auxiliary variables employed in order to generate draws from the resolvent chain via the splitting procedure, and the third coordinate corresponds to the subsequent values of the resolvent chain.

The process $Z=(Z_t^1,Z_t^2,Z_t^3)_{t\geq 0}$ can be constructed according to the following procedure. Firstly, let $Z_0^1 = X_0 = x$. Independently of Z_1 generate $Z_0^2 \sim U[0,1]$, where $U[0,1]$ denotes the uniform distribution on the unit interval. Given $\{Z_0^2=u\}$, draw Z_0^3 according to $K((x,u),dx')$. Then inductively for $n\geq 1$, on $Z_n=(x,u,x')$:

I. Choose σ_{n+1} according to

$$\left(\frac{p_t(x,x')}{u(x,x')}\mathbb{1}_{\{0<u(x,x')<\infty\}} + \mathbb{1}_{\{u(x,x')\in\{0,\infty\}\}}\right)e^{-t}dt \quad \text{on} \quad \mathbb{R}_+. \tag{3.7}$$

The next sampling time T_{n+1} is given by $T_n+\sigma_{n+1}$.

II. On $\{\sigma_{n+1}=t\}$, put $Z^2_{T_n+s} := u$ and $Z^3_{T_n+s} := x'$ for all $0\leq s<t$.

III. Draw a bridge of Z^1 conditioned on its starting point $Z^1_{T_n}$ and end point $Z^3_{T_n}$, so that for every $0<s<t$ we obtain

$$Z^1_{T_n+s} \sim \frac{p_s(x,y)p_{t-s}(y,x')}{p_t(x,x')}\mathbb{1}_{\{p_t(x,x')>0\}}\Lambda(dy). \tag{3.8}$$

Let $Z^1_{T_n+s} := x_0$ for some fixed $x_0\in E$ on $\{p_t(x,x')=0\}$. Moreover, given

$Z^1_{T_n+s} = y$ on $s + u < t$ we have that

$$Z^1_{T_n+s+u} \sim \frac{p_u(y,y')p_{t-s-u}(y',x')}{p_{t-s}(y,x')} \mathbb{1}_{\{p_{t-s}(y,x')>0\}} \Lambda(dy'). \tag{3.9}$$

Again, on $\{p_{t-s}(y,x') = 0\}$, let $Z^1_{T_n+s} = x_0$.

IV. At jump time T_{n+1} we have $Z^1_{T_{n+1}} := Z^3_{T_n} = x'$. Draw $Z^2_{T_{n+1}}$ independently of $Z_s, s < T_{n+1}$, uniformly on the unit interval. Given $\{Z^2_{T_{n+1}} = u'\}$, generate

$$Z^3_{T_{n+1}} \sim K((x',u'), dx''). \tag{3.10}$$

Note that in the construction of Z the inter-sampling times $(\sigma_n)_{n\geq 1}$ are drawn according to (3.7), their conditional distribution given the starting and endpoint of the sampled chain. Equation (3.8) and (3.9), describe the distributions of points in a bridge of the process X. The first coordinate of Z consists of bridges drawn according to the law of the original process X, between realisations of the resolvent chain. The results of [58, 59] that we work with are given in the following propositions. Firstly, the first coordinate of Z has the desired distribution.

Proposition 3.1 ([58, Proposition 2.8])**.** *The constructed process Z from the simulation scheme given in (3.7)–(3.10) is a Markov process with respect to its natural filtration $\mathbb{F}$. Moreover, the first coordinate Z^1 is equal in law to our process X, namely,*

$$\mathcal{L}((X_t)_{t\geq 0}|X_0 = x) = \mathcal{L}((Z^1_t)_{t\geq 0}|Z^1_0 = x).$$

Moreover, $(T_n - T_{n-1})_{n\geq 1}$ are i.i.d exponential random variables and are independent of Z^1; therefore, we also have that

$$\mathcal{L}((X_{T_n})_{n\geq 0}|X_0 = x) = \mathcal{L}((Z^1_{T_n})_{n\geq 0}|Z^1_0 = x).$$

Moreover, the process X is embedded in a richer process Z, which admits a recurrent atom $A := C \times [0,\alpha] \times E$ in the sense of the following proposition.

Proposition 3.2 ([59, Proposition 4.2])**.** *Let (S_n, R_n) be a sequence of stopping times defined as $S_0 = R_0 := 0$ and*

$$S_{n+1} := \inf\{T_m > R_n : Z_{T_m} \in A\} \quad \textit{and} \quad R_{n+1} := \inf\{T_m : T_m > S_{n+1}\}.$$

Then Z_{R_n} is independent of $\mathcal{F}_{R_{n-1}}$ for all $n \geq 1$ and $(Z_{R_n})_{n\geq 1}$ is an i.i.d sequence with

$$Z_{R_n} \sim \nu(dx)\lambda(du)K((x,u),dx') \quad \textit{for all } n \geq 1.$$

The stopping times $\{S_n\}_n$ thus denote the hitting times of the recurrent atom A for the jump process $(Z_{T_n})_n$, and $\{R_n\}_n$ denote the implied regeneration epochs of the process Z. As a direct consequence, we obtain the following regenerative structure for the original process.

Proposition 3.3 ([59, Proposition 4.4])**.** *Let f be a measurable π-integrable function, then we can construct a sequence of increasing stopping times $\{R_n\}_n$ with $R_0 = 0$ and*

$$\xi_n := \int_{R_{n-1}}^{R_n} f(X_s)\, ds, \quad n \geq 1,$$

such that the sequence $\{\xi_n\}_n$ is a stationary sequence under $\mathbb{P}_\nu$. Moreover, for $n \geq 2$, ξ_n is independent of $\mathcal{F}_{R_{n-2}}$.

The regenerative structure given in Proposition 3.3 was also noted by [82]. They define a process X to be *one-dependent regenerative* if there exists, on a possibly enlarged probability space, a sequence of randomised stopping times R_n with corresponding cycle lengths $\rho_n = R_{n+1} - R_n$ such that $\{(X_{R_n+t})_{t\geq 0}, (\rho_{n+k})_{k\geq 0}\}$ has the same distribution for each $n \geq 1$ and are independent of $\{(\rho_n)_{n=1}^{k-1}, (X_t)_{t<R_{n-1}}\}$ for $n \geq 2$. Note that according to this definition the initial cycle is allowed to have a different distribution. In [58] a constructive approach towards this result is given, in which they explicitly define the corresponding stopping times and the recurrent atom. By the implied regenerative structure of X, we obtain the following characterisation of the stationary measure.

Proposition 3.4 ([82, Theorem 2])**.** *Let X be a positive recurrent one-dependent regenerative process, then we can characterise its stationary measure as follows*

$$\pi(A) = \frac{1}{\varrho}\ \mathbb{E}_\nu \int_0^{R_1} \mathbb{1}_{\{X_s \in A\}}\, ds, \tag{3.11}$$

where ϱ is defined as $\mathbb{E}_\nu R_1$. Moreover, we have the following erdogic law of large numbers

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T f(X_s) ds = \frac{1}{\varrho}\ \mathbb{E}_\nu \int_0^{R_1} f(X_s) ds \quad a.s., \tag{3.12}$$

for all $f : E \to \mathbb{R}^d$ with $\pi(\|f\|) < \infty$.

Note that the normalisation constant $1/\varrho$ given in Proposition 3.4 is finite and non-zero due to the positive Harris recurrence of the process.

Remark 3.5. The framework of [58, 59] does not require ergodicity. Moreover, it is important to note that contrary to the classically regenerative case, Proposition 3.4 does not imply convergence in total variation to the stationary measure. For a counterexample see [82, Remark 3.2]. □

For our applications, we will require ergodicity and hence we must additionally impose this as stated in (3.2). These ergodicity requirements are usually established through Foster–Lyapunov drift conditions; see [32] and [39] for exponential and polynomial ergodicity respectively. These results have been applied to several classes of diffusion processes, see for example [15, Theorem 8.3 and 8.4] and [85, Theorem 3.1 and 4.1].

For PDMPs, [11] show aperiodicity, positive Harris recurrence, and exponential ergodicity of the Zig-Zag process for target distributions that have a non-degenerate local maximum and appropriately decaying tails. In [30] and [34] conditions for exponential ergodicity of the Bouncy Particle Sampler are given. Utilising hypocoercivity techniques, [2] establish polynomial rates of convergence for PDMPs with heavy-tailed stationary distributions. When we are concerned with PDMPs we will require the following regularity conditions on the stationary density:

Assumption 2. *Assume that the density of π is twice continuously differentiable, strictly positive, has a non-degenerate local maximum and* $\lim_{\|x\|\to\infty} \pi(x) = 0$*. Moreover, assume that π has a finite set of local extrema.*

These regularity conditions are often imposed in order to analyse the ergodic behaviour of PDMPs. Assumption 2 with accompanying conditions on the decay of the tails of the target distribution are used to show various rates of ergodicity.

4. Main theorems

The most straightforward approach for obtaining a strong approximation result for Markov processes would be through ergodicity requirements. In [53] it is shown that a multivariate strong invariance principle holds for sums of random vectors satisfying a strong mixing condition; see also Theorem 7.1. This mixing condition is satisfied when one has an appropriate rate of ergodicity of the process. All proofs are provided in Section 7.

Theorem 4.1. *Let $X = (X_t)_{t\geq 0}$ be polynomially ergodic of order $\beta \geq (1+\varepsilon)(1+2/\delta)$ for some $\varepsilon, \delta > 0$. Then for every initial distribution and for all $f : E \to \mathbb{R}^d$ with $\pi(\|f\|^{2+\delta}) < \infty$, we can construct a process that is equal in law to X together with a standard d-dimensional Brownian motion $W = (W(t))_{t\geq 0}$ on some probability space such that*

$$\left\| \int_0^T f(X_t)\,dt - T\pi(f) - \Sigma_f^{1/2} W(T) \right\| = O(\psi_T) \quad a.s. \tag{4.1}$$

with

$$\psi_T = T^{1/2-\min(\delta/(2\delta+4),\ \lambda)} \text{ for some } \lambda \in (0, 1/2), \tag{4.2}$$

and positive semi-definite $d \times d$ covariance matrix Σ_f given by

$$\Sigma_f = \int_0^\infty \mathrm{Cov}_\pi\left(f(X_0), f(X_s)\right) ds + \int_0^\infty \mathrm{Cov}_\pi(f(X_s), f(X_0))\, ds, \tag{4.3}$$

with all entries converging absolutely and integration of matrices defined element-wise.

Remark 4.2. The asymptotic covariance matrix Σ_f given in Theorem 4.1 cannot be simplified. Only for the univariate case ($p = 1$) and for reversible processes do we obtain that

$$\Sigma_f = 2\int_0^\infty \mathrm{Cov}_\pi(f(X_0), f(X_s))\, ds. \tag{4.4}$$

As a result of the reversibility, the cross-covariance matrices in (4.3) will be symmetric and thus the asymptotic covariance can be expressed as (4.4). □

The rate ψ_T appearing in Theorem 4.1 will depend on the dependence and moment structure of the considered process. For processes admitting higher order moments and having faster decaying levels of dependence the approximation bound ψ_T will tend to infinity at a slower rate. This can be interpreted as the magnitude of the difference between the centred additive functional of the process and the approximating Brownian motion being smaller. Although result (4.1) has useful applications for arbitrary $\lambda \in (0, 1/2)$, many refined limit theorems require an explicit remainder term, where more insight is given regarding the impact of the moment and dependence structure on the approximation error. In order to derive a more refined strong invariance principle we will make us of splitting arguments. Following the continuous time Nummelin splitting technique, as introduced in [58] and described in Section 3, it follows that the process can be embedded in a richer process, which admits a recurrent atom. Hence the process can be redefined such that it can be split in identically distributed blocks of random variables, which are one-dependent. Therefore we can utilise the approximation results for weakly m-dependent sequences of [5] to obtain a strong invariance principle; see also Theorem 7.5.

Proposition 4.3. *Let $X = (X_t)_{t\geq 0}$ be an aperiodic, positive Harris recurrent Markov process for which Assumption 1 is satisfied. Let $f : E \to \mathbb{R}$, be a given π-integrable function. Define the sequence of random times $\{R_n\}_{n=1}^{\infty}$ and $\{\xi_n\}_{n=1}^{\infty}$ as in Propositions 3.2 and 3.3. Moreover, assume that*

$$\mathbb{E}_\nu[R_1^q] < \infty \quad \text{for some} \quad q > 2, \tag{4.5}$$

$$\mathbb{E}_\nu \left| \int_0^{R_1} f(X_s) ds \right|^p < \infty \quad \text{for some} \quad p > 2. \tag{4.6}$$

Then for every initial distribution we can construct a process, on an enriched probability space, that is equal in law to X together with two standard Brownian motions W_1 and W_2 such that

$$\left| \int_0^T f(X_s) ds - T\pi(f) - W_1(\sigma_T^2) - W_2(\tau_T^2) \right| = O(\psi_T) \ \text{a.s.}, \tag{4.7}$$

where $\{\sigma_T^2\}$ and $\{\tau_T^2\}$ are non-decreasing sequences with $\sigma_T^2 = \frac{\sigma_\xi^2}{\varrho} T + O(\frac{T}{\log T})$, $\tau_T^2 = O(\frac{T}{\log T})$ as $T \to \infty$, and $\psi_T, \pi(f), \varrho$, and σ_ξ are defined in equations (4.9) to (4.12) below.

In Proposition 4.3 we obtain an explicit approximation error. In alignment with expectations, we see that the existence of higher-order moments will result in an improved approximation error. However, the required moment conditions for Proposition 4.3 stated in (4.5) and (4.6) are impractical and would be burdensome, if not impossible, to verify directly for most applications. For classically regenerative Markov chains this problem also arises, see the analogous requirements of regenerative simulation given in [66] and the strong approximation result of [19]. The results of [46] were the first to simplify moment conditions of this form and give practical sufficient conditions for regenerative simulation. More specifically, in their main result they show that polynomial or geometric ergodicity and moment conditions with respect to the stationary measure are sufficient to guarantee finiteness of the second moment of a cycle. This result was generalised to higher order cycle moments by [49] and [4]; hence simplifying the required conditions of [19]. However, the aforementioned approaches are all for Markov chains satisfying a one-step minorisation condition, i.e., for the classically regenerative setting. Since our setting involves a more complicated reconstruction of the process of interest, the results do not immediately carry over. In Theorem 4.3, we show that the cycle moment conditions (4.5) and (4.6) required for Proposition 4.3 can also be guaranteed with more easily verifiable ergodicity and moment conditions.

Theorem 4.4. *Let $X = (X_t)_{t\geq 0}$ be an aperiodic, positive Harris recurrent Markov process for which Assumption 1 is satisfied. Moreover, let X be polynomially ergodic of order $\beta > 1 + p(p+\varepsilon)/\varepsilon$, for some $\varepsilon > 0$ then*

$$\mathbb{E}_\nu\left[(R_1)^{\beta-1}\right] < \infty.$$

Moreover, for all measurable $f : E \to \mathbb{R}$ with $\pi(|f|^{p+\varepsilon}) < \infty$ with $p \geq 1$ we have that

$$\mathbb{E}_\nu\left|\int_0^{R_1} f(X_s)ds\right|^p < \infty.$$

By combining Proposition 4.3 and Theorem 4.4 we obtain the desired strong invariance principle.

Theorem 4.5. *Let $X = (X_t)_{t\geq 0}$ be an aperiodic, positive Harris recurrent Markov process for which Assumption 1 is satisfied. Moreover, let X be polynomially ergodic of order $\beta > 1 + p(p+\varepsilon)/\varepsilon$, for given $p > 2$ and some $\varepsilon > 0$. Then for every initial distribution and for all measurable $f : E \to \mathbb{R}$ with $\pi(|f|^{p+\epsilon}) < \infty$ we can, on an enriched probability space, define a process that is equal in law to X and two standard Brownian motions W_1 and W_2 such that*

$$\left|\int_0^T f(X_s)ds - T\pi(f) - W_1(\sigma_T^2) - W_2(\tau_T^2)\right| = O(\psi_T)\ a.s., \tag{4.8}$$

where $\{\sigma_T^2\}$ and $\{\tau_T^2\}$ are non-decreasing sequences with $\sigma_T^2 = \frac{\sigma_\xi^2}{\varrho}T + O(\frac{T}{\log T})$, $\tau_T^2 = O(\frac{T}{\log T})$, and

$$\psi_T = \max\left\{T^{1/4}\log T, T^{1/p}\log^2(T)\right\}, \tag{4.9}$$

$$\pi(f) = \frac{1}{\varrho}\,\mathbb{E}_\nu\int_0^{R_1} f(X_s)\,ds, \tag{4.10}$$

$$\varrho = \mathbb{E}_\nu[R_1], \;\; and \tag{4.11}$$

$$\sigma_\xi = \sqrt{\mathrm{Var}_\nu(\xi_1) + 2\,\mathrm{Cov}_\nu(\xi_1,\xi_2)}\;. \tag{4.12}$$

Proof. The assertion follows immediately from Proposition 4.3 and Theorem 4.4. □

The appearance of the second Brownian motion in Theorem 4.5 is inherited from the strong invariance principle of [5]. Although we obtain different time perturbations of the Brownian motions, all desired properties carry over. The second Brownian motion appearing in (4.8) is of a smaller magnitude, and will therefore be asymptotically negligible in typical applications. Furthermore, even though the two Brownian motions are not independent, their correlation decays over time

$$\mathrm{Corr}\left(W_1(\sigma_t^2), W_2(\tau_s^2)\right) \to 0, \quad \text{as } t, s \to \infty. \tag{4.13}$$

Note that the nearly optimal convergence rate $O(T^{1/p}\log^2 T))$ obtained by [5] does not carry over. Instead, we obtain an approximation error that cannot be improved beyond $O(T^{1/4}\log T)$. Obtaining a superior approximation error remains an open problem for the class of processes considered in Theorem 4.5. A possible approach for attaining a better convergence rate would be to extend to results of [5] to a multivariate setting and then follow the approach of [61].

The univariate Zig-zag process passes every point in its state-space, in particular also the local optima of its target density, an infinite amount of times. This allows us the define regenerative cycles of the process. Therefore we can adapt the approach of [61] and obtain the optimal bound of $O(T^{1/p})$ for the strong approximation of the one-dimensional Zig-Zag process.

Theorem 4.6. *Let $Z = (X_t, V_t)_{t\geq 0}$ be an aperiodic, positive Harris recurrent one-dimensional Zig-zag process with an invariant distribution $\pi \otimes v$, where π satisfies Assumption 2. Moreover, let Z be polynomially ergodic of order $\beta > 1 + p(p+\varepsilon)/\varepsilon$, for given $p > 2$ and some $\varepsilon \in (0,1)$. Then for every initial distribution and for all measurable $f : E \to \mathbb{R}$ with $\pi(|f|^{p+\epsilon}) < \infty$ there exists a Brownian motion W such that*

$$\left|\int_0^T f(X_s)ds - T\pi(f) - \sigma_f^2 W(T)\right| = O(T^{1/p}) \;\; a.s., \tag{4.14}$$

where σ_f^2 can be characterised as (4.16).

In [61] a strong invariance principle is obtained for one-dimensional Markov chains satisfying a one-step minorization condition by making use of the implied regenerative properties. Note that their approach carries over for any regenerative process. However, they assume that the chain is exponentially ergodic and that the test function f is bounded. The boundedness of f is very restrictive for applications in MCMC, since it excludes many interesting examples such as the posterior mean and variance. Theorem 4.6 extends their results by only imposing polynomial ergodicity and only a necessary moment condition for the test function.

Furthermore, we see that if the target distribution is of product form, i.e., satisfies the factorisation $\pi(x) = \prod_{i=1}^{d} \pi_i(x_i)$, then the optimal bound carries over to the multivariate setting.

Theorem 4.7. *Let $Z = (X_t, V_t)_{t\geq 0}$ be an aperiodic, positive Harris recurrent d–dimensional Zig-zag process with an invariant distribution $\pi \otimes \upsilon$, where π is of product form and every π_i satisfies Assumption 2. Moreover, let Z be polynomially ergodic of order $\beta > 1 + p(p+\varepsilon)/\varepsilon$, for given $p > 2$ and some $\varepsilon \in (0,1)$. Then for every initial distribution and for all $f : E \to \mathbb{R}^d$ that can be decomposed as $\prod_i f_i(x_i)$ with $\pi(\|f\|^p) < \infty$, there exists a standard d-dimensional Brownian motion W such that*

$$\left\| \int_0^T f(X_t)\, dt - T\pi(f) - \Sigma_f^{1/2} W(T) \right\| = O(T^{1/p}) \quad a.s. \tag{4.15}$$

and covariance matrix $\Sigma_f = \operatorname{diag}\{\sigma^2_{f_1}, \dots, \sigma^2_{f_d}\}$ with

$$\sigma^2_{f_i} = \int_0^\infty \operatorname{Cov}_\pi(f_i(X_0^i), f_i(X_s^i))\, ds + \int_0^\infty \operatorname{Cov}_\pi(f_i(X_s^i), f_i(X_0^i))\, ds. \tag{4.16}$$

Note that although the proof of Theorem 4.7 relies on the fact that the d-dimensional Zig-Zag process Z can be decomposed into d one-dimensional independent Zig-Zag processes, the multivariate invariance principle does not directly follow from an application of Theorem 4.6, since even though the individual coordinates have regenerative cycles, the multivariate process Z does not possess regeneration times. Moreover, it must be guaranteed that the approximating Brownian motions for the individual components are defined on the same probability space.

Remark 4.8. From Theorem 4.4 we see that polynomial ergodicity of a sufficiently high order and moments with respect to the stationary distribution guarantee the existence of the p–th order cycle moments, which in turn determines the approximation error in our strong invariance results. In general, if we assume polynomial ergodicity of order $\beta > 1$, then from Remark 7.7 and (7.32), we see that the approximation error of Theorem 4.6 can in general taken to be of order $O(T^\alpha)$ with

$$\alpha = \max\{1/p', 1/(\beta - 1)\},$$

where $p' < \frac{1}{2}(\sqrt{\varepsilon(\varepsilon + 4(\beta-1))}) - \varepsilon)$ if $p > \frac{1}{2}(\sqrt{\varepsilon(\varepsilon + 4(\beta-1))}) - \varepsilon)$ and $p' = p$ otherwise. Therefore we see that a faster polynomial rate of convergence to the stationary measure improves the approximation error, up to the point where the approximation error from the moment conditions dominates. The same conclusion can be seen to hold for Theorem 4.5 and 4.7. Furthermore, from Remark 7.8, we see that under the assumption of exponential ergodicity, the conclusions of Theorem 4.4 and all aforementioned strong invariance principles hold with their stated approximation error. □

Remark 4.9. Note that in Theorem 4.6, the rate function $\lambda(x, v) = (vU'(x))^+$ guarantees the existence of regenerative cycles of the process. Namely, for every stationary point of π we can take an appropriate velocity, such that they form a regeneration epoch for the process. Hence for any PDMP with deterministic dynamics such that the process remains aperiodic, positive Harris recurrent, and polynomially ergodic, the strong invariance principles of Theorem 4.6 and 4.7 will hold. □

5. Analysis of batch means for Piecewise Deterministic Monte Carlo

In order to assess the accuracy of our PDMC sampler, we require a central limit theorem to hold and estimate the corresponding asymptotic variance. In [8] several conditions are given to obtain a CLT for the univariate Zig-Zag process. In [34], [30], and [11] a CLT is obtained for the Bouncy Particle sampler and Zig-Zag process respectively through geometric drift conditions, which in turn also imply exponential ergodicity. The strong invariance principles we obtained in Theorems 4.1, 4.5, 4.6, and 4.7 immediately imply the following central limit theorems for polynomially ergodic Markov processes.

Corollary 5.1. *Let $(Z_t)_{t\geq 0}$ with $Z_t = (X_t, V_t)$ be polynomially ergodic of order $\beta \geq (1+\varepsilon)(1+2/\delta)$ for some $\varepsilon, \delta > 0$. Then we have that for all $f : E \to \mathbb{R}^d$ with $\mu(\|f\|^{2+\delta}) < \infty$, a central limit theorem holds:*

$$\frac{1}{\sqrt{T}} \int_0^T (f(X_s, V_s) - \mu(f))\, ds \xrightarrow{d} \mathcal{N}_p(0, \Sigma_f). \tag{5.1}$$

Additionally, also a functional central limit theorem holds:

$$\left(\frac{1}{\sqrt{n}} \int_0^{nt} (f(X_s, V_s) - \mu(f))\, ds \right)_{t\geq 0} \xrightarrow{d} \Sigma_f^{1/2} W \text{ as } n \to \infty, \tag{5.2}$$

where

$$\Sigma_f = \int_0^\infty \mathrm{Cov}_\mu(f(X_0, V_0), f(X_s, V_s))\, ds + \int_0^\infty \mathrm{Cov}_\mu(f(X_s, V_s), f(X_0, V_0))\, ds, \tag{5.3}$$

$W = (W_t)_{t\geq 0}$ denotes a standard d-dimensional Brownian motion and the weak convergence is with respect to the Skorohod topology on $D[0, \infty)$, the space of real-valued càdlàg functions with domain $[0, \infty)$.

Proof. By [20, Theorem 1.17], the FCLT immediately follows from the strong invariance principle formulated in Theorem 4.1. Similarly, by [26, Proposition 2.1] the CLT follows. □

By the same argument the CLT follows for the processes considered in Theorems 4.5, 4.6, and 4.7. For simplicity, we will mainly consider the one-dimensional case, i.e. our quantity of interest is given by $\pi(f)$, with $f : E \to \mathbb{R}$ a given π-integrable function. Let the simulation output, which in our case consists of the position component of a PDMP, be given by $(X_t)_{t\in[0,T]}$. Note that from Corollary 5.1 also a (functional) central limit theorem follows for the position component of the process. We are interested in estimating the asymptotic variance (5.3); which we will denote by σ_f^2, when we are not considering the multivariate setting.

The batch means method divides the obtained sample trajectory of our process into non-overlapping parts. The sample variance of the means of the obtained batches gives rise to a natural estimator for the asymptotic variance. More specifically, we divide our simulation output in k_T batches of length ℓ_T such that $k_T = \lfloor T/\ell_T \rfloor$. We proceed by computing the sample average of each obtained batch;

$$\bar{Z}_i(\ell_T) := \frac{1}{\ell_T}\int_{(i-1)\ell_T}^{i\ell_T} f(X_s)ds, \quad i = 1, \dots, k_T. \tag{5.4}$$

If a functional central limit theorem holds for our process, it follows that the computed means $\bar{Z}_i(\ell_T)$ are asymptotically independent and identically distributed for each fixed amount of batches. Hence, we can heuristically reason that the sample variance of $(\bar{Z}_i(\ell_T)_{i=1}^{k_T}$ will be close to $\operatorname{Var}(\bar{Z}_i(\ell_T))$. Moreover, since each $\bar{Z}_i(\ell_T)$ is also an empirical mean, it is reasonable to expect their variance to be approximately σ_f^2/ℓ_T. The batch means estimator of the asymptotic variance is defined by correcting the sample variance of the batch means $(\bar{Z}_i(\ell_T))_{i=1}^{k_T}$ by a factor ℓ_T, namely

$$\hat{\sigma}_T^2 = \frac{\ell_T}{k_T - 1}\sum_{i=1}^{k_T}\left(\bar{Z}_i(\ell_T) - \frac{1}{k_T}\sum_{i=1}^{k_T}\bar{Z}_i(\ell_T)\right)^2. \tag{5.5}$$

Following the framework of [26], we impose the following conditions on the amount of batches and their length.

Assumption 3. *Let the amount of batches k_T and their lengths ℓ_T be such that*

i. $k_T \to \infty$, $\ell_T \to \infty$, *and* $\ell_T/T \to 0$ *as* $T \to \infty$,
ii. ℓ_T *and* T/ℓ_T *are both monotonically increasing,*
iii. there exists a constant $c \geq 1$ *such that* $\sum_{n=1}^{\infty} k_n^{-c} < \infty$.

The first requirement of Assumption 3 is a necessary condition for consistency as seen from the results of [41]. The second requirement is solely for technical reasons and the third requirement ensures that the amount of the batches grows fast enough; if we choose $\ell_T = T^{\alpha}$ the requirement holds for all $\alpha \in (0,1)$, since we can choose $c > 1/(1-\alpha)$.

Theorem 5.2. *Let Z be polynomially ergodic of order $\beta > 1 + p(p+\varepsilon)/\varepsilon$, for given $p > 2$ and some $\varepsilon \in (0,1)$ with stationary measure μ with $\mu(|g|^p) < \infty$. Assume that Assumption 3 holds and that*

$$\frac{T^{2/p}}{\ell_T}\log(T) \to 0, \ \text{as } T \to \infty, \tag{5.6}$$

then for every initial distribution $\hat{\sigma}_T^2 \to \sigma_f^2$ as $T \to \infty$ with probability 1.

Proof. The result follows from Theorem 4.1, [49, Proposition 3], and [26, Theorem 3.3]. □

Remark 5.3. Note that Theorem 5.2 weakens the currently available regularity conditions guaranteeing strong convergence of the batch means estimator in an MCMC setting. This is a direct consequence of the fact that Theorems 4.6 and 4.7 obtain the optimal approximation rate of $O(T^{1/p})$ whereas the results of [49] are based upon the strong invariance principle of [19] which attains the rate $O(T^\gamma \log T)$, with $\gamma = \max(1/p, 1/4)$. More specifically, for f with $\pi(|f|^p) < \infty$ [49] requires $T^\gamma \log^3(T)/\ell_T \to 0$ as $T \to \infty$. In particular for the case where $p > 4$, Theorem 5.2 is able to significantly weaken the conditions on the required batch length ℓ_T. As a direct result of the smaller batch lengths, we are able to use a higher number of batches k_T, which results in a smaller variance for the batch means estimator, as seen in Theorem 5.4. Note that a similar conclusion holds for the overlapping batch means and spectral variance estimators considered in [38]. □

We see from the required assumption (5.6) that a larger approximation error in the strong invariance principle, which corresponds to higher orders of dependence, results in a larger required batch size ℓ_T. This is in agreement with the idea behind batching methods; every batch should give a proper representation of the dependence structure of the process. Otherwise, a structural bias will be introduced in the estimation procedure. On the other hand, choosing the batch size larger than necessary will result in a lower amount of batches k_T leading to a higher variance for the estimator. Strong approximations can also be used to characterise the mean squared error and obtain a central limit theorem for the batch means estimator.

Theorem 5.4. *Let Z be polynomially ergodic of order $\beta > 1 + p(p+\varepsilon)/\varepsilon$, for given $p > 2$ and some $\varepsilon \in (0,1)$ with stationary measure μ with $\mu(|f|^p) < \infty$. Let the initial distribution be given by μ and assume that Assumption 3 holds and $\mathbb{E}_\mu C^2 < \infty$, where C is defined in (7.68) below. Then we have that*

$$\mathbb{E}_\mu \left|\hat{\sigma}_T^2 - \sigma_f^2\right|^2 = 2\sigma_f^4 \frac{\ell_T}{T} + O\left(\frac{T^{1/p}}{\sqrt{T}} \log^{\frac{1}{2}} T\right) + O\left(\ell_T^{-1} T^{2/p} \log T\right). \tag{5.7}$$

Moreover, if $\ell_T^{-1} T^{1/p} (T \log T)^{1/2} \to 0$ as $T \to \infty$, then we obtain a CLT for the batch means estimator

$$\sqrt{k_T}(\hat{\sigma}_T^2 - \sigma_f^2) \xrightarrow{d} \mathcal{N}(0, 2\sigma_f^4) \text{ as } T \to \infty. \tag{5.8}$$

Proof. By the imposed conditions of the process, the strong invariance principle formulated in Theorem 4.1 holds. The first claim then follows by [27, Theorem 1 and Lemma 3] and the second by [27, Proposition 2]. □

The first and second term in (5.7) describe the variance, whereas the third term represents the bias. Note that the second term does not depend on ℓ_T and tends to zero. The obtained bounds for the variance are sharp, whereas, the bounds for bias have room for improvement.

In the multivariate setting, where our quantity of interest is given by $\pi(f)$, with $f : E \to \mathbb{R}^d$ a given π-integrable function, the batch means estimator is given by

$$\hat{\Sigma}_T = \frac{\ell_T}{k_T - 1} \sum_{i=1}^{k_T} \left(\bar{Z}_i(\ell_T) - \frac{1}{k_T} \sum_{i=1}^{k_T} \bar{Z}_i(\ell_T) \right) \left(\bar{Z}_i(\ell_T) - \frac{1}{k_T} \sum_{i=1}^{k_T} \bar{Z}_i(\ell_T) \right)^T, \tag{5.9}$$

where $\bar{Z}_i(\ell_T)$ is defined in (5.4). Given the strong invariance principle of Theorem 4.1, the results of [88] for the multivariate batch means estimator immediately carry over.

Theorem 5.5. *Let Z be polynomially ergodic of order $\beta \geq (1+\varepsilon)(1+2/\delta)$ for some $\varepsilon, \delta > 0$. Let $f : E \to \mathbb{R}^d$ with $\mu(\|f\|^{2+\delta}) < \infty$. Assume that Assumption 3 holds and that*

$$\frac{\psi_T^2}{\ell_T} \log(T) \to 0, \text{ as } T \to \infty, \tag{5.10}$$

with ψ_T defined in (4.2), then for every initial distribution we have that $\hat{\Sigma}_T \to \Sigma_f$ as $T \to \infty$ with probability 1.

Proof. The claim follows from Theorem 4.1 and [88, Theorem 2]. $\square$

Furthermore, if the target distribution is of product form and we consider the Zig-Zag Sampler, then Theorem 4.7 gives a strong invariance principle with an explicit approximation error. Therefore, we can replace condition (5.10) of Theorem 5.5 with (5.6) for every component of the Zig-Zag process. This results in a condition that can more easily be verified.

5.1. Discussion

5.1.1. Batch size selection for PDMC

In [41] it is shown that there exists no consistent estimator of σ_f^2 with fixed amounts of batches. Hence the amount of batches should explicitly depend on the length of the simulation T. For the standard choice $\ell_T = T^\alpha$ we see that for $\alpha > 1/2p$ we obtain both strong consistency and L^2-convergence of the batch means estimator. Theorem 5.4 suggests that $\alpha^* = (2+p)/2p$ would be optimal in the mean squared error sense. The well-known results of [17], [43], and [83] obtain a bound for the bias of order $O(\ell_T^{-1})$, which implies an optimal (in the MSE sense) batch size of $\ell_T^\diamond \asymp T^{1/3}$. However, the aforementioned results require the sampling process to be stationary, uniformly ergodic, and satisfy moment condition $\pi(f^{12}) < \infty$. Obtaining the bias term of order $O(\ell_T^{-1})$ for batch means under milder conditions remains an unaddressed problem. Theorem 5.2 and 5.4 only require a strong invariance principle, which we have shown holds under polynomial ergodicity; a very reasonable assumption for simulation output. Moreover, these results do not require stationarity and thus hold for every initial distribution. Theorem 5.4 imposes more demanding conditions on ℓ_T

than aforementioned frameworks, however, it is quite reasonable to let the batch size depend on the dependence structure of the process through ψ_T, instead of only the auto-covariance function $(\gamma(s))_{s\geq 0}$ through the constant $\int s\gamma(s)ds$, as is the case in the aforementioned results. Moreover, in practice, the performance of batch means methods with batch size $\ell_T^\diamond$ are often found to be sub-optimal whereas larger batch sizes see better finite sample performance, as noted by for example [38]. We see that for exponentially and polynomially ergodic sampling algorithms the batch size choice $\ell_T^* = T^{\alpha^*}\log(T)$ gives almost sure convergence, convergence in mean square, and guarantees asymptotic normality of the BM estimator. However, the optimal tuning parameter does depend on the number of moments of the target distribution. If no theoretical guarantees can be obtained, we can in practice also assess the level of tail decay of our target distribution by examining the simulation output. For a survey of statistical methods for the detection of heavy tails, estimation of the tail index, and the number of finite moments, see for example [1] and all their given references. For uniformly ergodic sampling algorithms, the aforementioned results imply an optimal batch size of order $T^{1/3}$.

An alternative approach for determining the optimal batch size was given by [18], which obtains an optimal batch size of $\tilde{\ell}_T \asymp T^{1/2}$ by minimising the distance between the cumulants of the studentised ergodic average and a standard Gaussian, which suggest that the resulting confidence intervals enjoy improved finite-sample properties.

5.1.2. Asymptotic normality of the batch means estimator

We see that given polynomial ergodicity, also the central limit theorem for the batch-means estimator carries over to the PDMC setting. The results of [80] require uniform ergodicity and the moment condition $\pi(f^{12}) < \infty$, in order to obtain asymptotic normality of the batch, means estimator. Since uniform ergodicity is not attainable for most practical problems, less stringent conditions on the rate of ergodicity are desired. Theorem 5.4 places more restrictive conditions on the batch size and excludes the choice $\ell_T^\diamond \asymp T^{1/3}$. In [16] a CLT for the batch means estimator is obtained assuming reversibility, stationarity, geometric ergodicity, and moment condition $\pi(f^8) < \infty$. Moreover, the required batch size must be such that $k_T = o(\ell_T^2)$. Hence their result is also unable to guarantee asymptotic normality for batch size $\ell_T^\diamond$. We see that Theorem 5.4 gives more practical conditions for guaranteeing asymptotic normality of the batch-means estimator, in particular, the results are applicable to non-reversible processes.

5.1.3. Spectral variance and overlapping batch means estimators for the PDMC standard error

Analogous to the batch means method, given the strong invariance principle formulated in Theorem 4.1, many results for other estimators of the asymptotic variance also carry over. In [38] more convenient alternatives are given for some of the requirements of the framework given in [25]. The results of [38] regarding spectral variance and overlapping batch means estimators for MCMC output are thus also applicable for PDMC, with minor adjustments to their assumptions. Note that the assumed minorisation condition and geometric ergodicity of the Markov chain in [38] are only imposed such that the strong invariance principle of [19] holds. Although implementation of spectral variance estimators for continuous-time output might be impractical, these estimators are still of theoretical interest. Numerous estimation methods, such as overlapping batch means and certain standardised time series methods, with feasible implementation for PDMC output, can be shown to be (asymptotically) equivalent to spectral estimators. Furthermore, we expect the results of [89] and [57] regarding spectral variance and generalised overlapping batch means estimators respectively to remain valid in the continuous-time setting. Hence also the implications for the optimal values of the tuning parameters of these estimation methods for the asymptotic variance remain valid. Lastly, note that our results hold for all sampling algorithms that produce continuous-time output, and are not restricted to the PDMP setting.

5.1.4. Regenerative simulation

From the proof of Theorem 4.6, we see that the univariate Zig-Zag sampler possesses a recurrent atom. Moreover, any local optimum with an appropriate velocity can be taken as a recurrent atom. Hence, regenerative simulation can also be considered for the estimation of the asymptotic variance. Let $(R_k)_{k\in\mathbb{N}}$ denote the hitting times of the chosen regeneration epoch of the process, then we can define the contribution of cycle k to the time-average as

$$\xi_k := \int_{R_{k-1}}^{R_k} f(X_s)\,ds, \quad k \geq 1,$$

and the corresponding cycle lengths as $\tau_k = R_k - R_{k-1}$. From the strong law of large numbers, it follows that $\hat{\pi}_{RS}(f) = \sum_{j=1}^{n} \xi_j / R_n$ is a consistent estimator of $\pi(f)$. Moreover, the corresponding asymptotic variance can be estimated by

$$\hat{\sigma}^2_{RS} = \frac{\sum_{j=1}^{n}(\xi_j - \hat{\pi}_{RS}(f)\tau_j)^2}{\frac{1}{n}R_n^2}.$$

For a more detailed description of regenerative simulation, we refer to for example [14] or [46]. Note that $\hat{\sigma}^2_{RS}$ is a ratio estimator and hence can be biased for an insufficient number of tours. Although this bias is small when the coefficient of determination of R_n is small, as explained in for example [14], there are other caveats to this approach that also need to be taken into account. Firstly, the practicality of the regeneration-based estimator will depend on the length of the regenerative cycles. As mentioned in the discussions of [38] and [40], it can take the chain a lot of time to reach its regeneration epoch even in moderately large finite state-spaces or as the dimension of the Markov chain increases.

The expected time for the Zig-zag sampler to move between modes increases proportionally to the ratio of their density value, as seen from the results of [65]. Thus if the chosen regeneration epoch is a local maximum of π that has a substantially lower density value compared to the global maximum, the tours required for regenerative simulation are expected to be long.

Moreover, regenerative simulation requires the identification of the regenerative states. For the case with the Zig-zag sampler, this requires that the location of a local extremum of the target distribution is known. Note that even though our results assume the existence of at least one local maximum, we do not require to know its location or even that the sampler has to visit all local optima often. Therefore, in case an appropriate maximum of the target density is known a priori or can be obtained with low computational cost, regenerative simulation can be considered. In general, the batch means or overlapping batch means methods are more widely applicable.

6. Increments of additive functionals of ergodic Markov processes

Strong approximation results enable various asymptotic properties of Brownian motion to carry over to other stochastic processes. In this section, we show that the strong invariance principle given in Theorem 4.5 can be used to show that the increments of additive functionals of Markov processes are of the same magnitude as Brownian increments, provided we have sufficient decay of the approximation error. The following theorem describes the magnitude of the fluctuations of Brownian increments over subintervals of length a_T.

Theorem 6.1 ([21, Theorem 1])**.** *Let $W = (W_t)_{t\geq 0}$ denote a Brownian motion, and let a_T be a positive non-decreasing function of T such that $0 < a_T \leq T$ and T/a_T is non-decreasing. Then*

$$\limsup_{T\to\infty} \sup_{0\leq t\leq T-a_T} \sup_{0\leq u\leq a_T} \beta_T |W_{t+u} - W_t| = 1 \quad a.s., \tag{6.1}$$

where

$$\beta_T = \left(2a_T \left[\log \frac{T}{a_T} + \log\log T\right]\right)^{-1/2}.$$

Taking $a_T = T$ gives the law of iterated logarithm, and for $a_T = c\log T$ with $c > 0$, the Erdös-Rényi law of large numbers for Brownian motion is obtained, as seen in for example [22, Theorem 2.4.3]. This fluctuation result has been extended to other processes such as integrated Brownian motion, fractional Brownian motion, and non-stationary Gaussian processes, see [56], [36] and [71] respectively. While these fluctuation results are of independent interest, they are also used as building blocks in applications, such as proving convergence properties of kernel density estimators, see for example [74] and [29]. These fluctuation results are also used for proving almost sure convergence of various estimators of the asymptotic variance in simulation output settings, see the references given in Section 4.1. By the Komlós-Major-Tusnády approximation the fluctuation result immediately carries over for i.i.d. sequences satisfying appropriate moment conditions, as seen in [22, Theorem 3.1.1 and 3.2.1].

In order to describe the fluctuations of additive functionals over an interval of a specified length a_T, we require an explicit remainder term for the Brownian approximation, as given in Theorem 4.5. However, due to the appearance of the second Brownian motion in this invariance principle and the perturbed time sequences, it is not immediate that the Brownian fluctuation result carries over. In [5] it is shown that the magnitude of the increments of partial sums of weakly m-dependent sequences are indeed given by Theorem 6.1, due to the smaller scaling of the second Brownian motion. However, in our case the perturbed time sequences are random since they depend on the amount of one-dependent regenerative cycles of the process, hence the desired result does not follow directly from [5, Theorem 4].

Theorem 6.2. *Let $X = (X_t)_{t\geq 0}$ be an aperiodic, positive Harris recurrent Markov process for which Assumption 1 is satisfied. Moreover, let X be polynomially ergodic of order $\beta > 3 + p/\varepsilon$, for given $p > 2$ and some $\varepsilon > 0$. Consider a function $f : E \to \mathbb{R}$ with $\pi(f) = 0$ and $\pi(|f|^{p+\varepsilon}) < \infty$. Let a_T be a given positive non-decreasing function of T such that*

i. $0 < a_T \leq T$,
ii. T/a_T *is non-decreasing,*
iii. a_T *is regularly varying at* ∞ *with index* $\zeta \in (0, 1]$.

Suppose that $\beta_T \psi_T = o(1)$, where

$$\beta_T = \left(2a_T \left[\log \frac{T}{a_T} + \log\log T\right]\right)^{-1/2},$$

and

$$\psi_T = \max\left\{T^{1/4}\log T, T^{1/p}\log^2(T)\right\}.$$

Then we have that

$$\limsup_{T\to\infty} \sup_{0\leq t\leq T-a_T} \sup_{0\leq u\leq a_T} \beta_T \left|\int_t^{t+u} f(X_s)ds\right| \leq \frac{\sigma_\xi^2}{\varrho} \quad a.s. \tag{6.2}$$

As noted by [5], the split invariance principle also implies the distributional version of Theorem 6.2; with similar adaptations to their argument this would also hold in our case. Since the approximation error ψ_T of Theorem 4.5 cannot be guaranteed to be smaller than $O(T^{1/4}\log T)$, the fluctuation result given in Theorem 6.2 cannot describe the magnitude of increments over slowly growing time intervals a_T.

6.1. Application to diffusion processes

Diffusions are an important class of processes for which the strong approximation given in Theorem 4.5 and the related fluctuation result given in Theorem 6.2

are applicable. Let $X = (X_t)_{t\geq 0}$ denote a one-dimensional diffusion process that is defined as the solution of the following time-homogeneous stochastic differential equation (SDE)

$$\begin{cases} dX_t = b(X_t)dt + \sigma(X_t)dW_t \\ X_0 \sim \mu, \end{cases} \tag{6.3}$$

where μ is the initial distribution of the process, $\mathfrak{X} \subseteq \mathbb{R}$ denotes the state-space, $b : \mathfrak{X} \to \mathbb{R}$ and $\sigma : \mathfrak{X} \to \mathbb{R}$ denote the drift and volatility function respectively, and the process W is a Brownian motion. We assume that all required regularity conditions hold such that the existence and uniqueness of a strong solution of the SDE is guaranteed. For example, we can impose Lipschitz conditions on the drift and volatility of the SDE. For a more detailed explanation, we refer to [79].

For diffusion processes to admit the desired ergodic properties we must impose additional regularity conditions. Let x_0 denote the initial value of our process, then the scale function of a one-dimensional diffusion is given by

$$s(u) = \int_{x_0}^{u} \exp\left[-2\int_{x_0}^{z} \frac{b(y)}{\sigma^2(y)}dy\right] dz \text{ and must satisfy } \lim_{u\to\pm\infty} s(u) = \pm\infty. \tag{6.4}$$

If condition (6.4) holds it follows that the diffusion is recurrent, that is, the time for the process to return to any bounded subset of its state space is a.s. finite. The speed density of the diffusion process $m : \mathfrak{X} \to \mathbb{R}^+$, given by $m(u) = \big(s(u)\sigma^2(u)\big)^{-1}$, must be Lebesque integrable for the diffusion to be positive Harris recurrent. For higher-dimensional diffusion processes [7] gives conditions that guarantee positive Harris recurrence. The results of [54, Theorem 2.3] show that diffusions are aperiodic if the drift and diffusion coefficients are Hölder continuous and the diffusion coefficient is uniformly elliptic on an open ball. Alternatively, from [84, Remark 4.3; Theorem 2.6] we see that aperiodicity can also be obtained under linear growth conditions on the drift, uniform ellipticity of the diffusion coefficient, and requiring that the transition probability is positive for any set with positive Lebesgue measure. In order for the obtained strong invariance principles given in Theorem 4.1 and Theorem 4.5 to hold, we require polynomial or exponential convergence to stationarity. These assumptions are usually obtained by verifying drift conditions for the diffusion processes, see for example [15, Theorem 8.3 and 8.4] and [85, Theorem 3.1 and 4.1].

In order for the strong approximation result in Theorem 4.5 and the related fluctuation result of Theorem 6.2 to hold, the Nummelin splitting scheme of [58] must be applicable. Therefore we must impose regularity conditions such that Assumption 1 is satisfied, i.e., the transition semigroup of the diffusion must be Feller and admit densities with respect to some dominating measure. Under appropriate growth and continuity conditions on the drift and volatility, diffusion processes are Feller, see for example [90, Theorem 2.2]. Moreover, if the volatility function σ is strictly positive (positive-definite in the multivariate case), the diffusion is elliptic and therefore admits transition densities; [86, Theorem 3.2.1]. Hence, Assumption 1 is satisfied. Alternatively, for multivariate diffusions, we can impose the parabolic Hörmander condition which ensures that the propagation of the noise through the different coordinates is sufficient, such that the transition density exists, see for example [79, Theorem 38.16].

6.2. Discussion and suggestions for further research

We see that Theorem 4.1 and Theorem 4.5 are applicable to a broad class of diffusions and extend the current results on strong approximations for diffusion processes. In [45] and [64] strong invariance principles are obtained for diffusions and a complementary fluctuation result and change point test respectively. The results of [64] yield an explicit approximation error comparable to that of Theorem 4.5, but are only applicable to stochastic integrals with respect to Brownian motion, i.e., diffusion processes with no drift. The results of [45] give an implicit approximation error and hold for singular diffusions. The strong invariance principle of [45] is not covered by our results since singular diffusions generally do not satisfy the mixing properties required for our framework.

The obtained strong invariance principles offer numerous applications for diffusion processes, see for example [23] and their given references. Following the approach of [5, Proposition 2], Theorem 4.5 can be used to obtain a change-point test for diffusions. If the diffusion process we consider has a drift that enforces mean-reversion, we could construct a test for the existence of a deterministic linear trend over specified time periods. This approach would require continuous-time output of a diffusion process, and is therefore more of theoretical interest. However, it is plausible that the asymptotic behaviour of the change-point test should carry over to the high-frequency setting, where the diffusion is observed discretely and it is assumed the inter-observation times tend to zero.

7. Proofs

7.1. Theorem 4.1

In [53] a strong invariance principle is given for random variables that satisfy certain mixing conditions. In order to state their result, we first briefly introduce mixing coefficients. Let $\mathscr{A}$ and $\mathscr{B}$ denote two sub σ-algebras of our probability space. The α-mixing coefficients of two σ-algebras quantify their dependence as follows

$$\alpha(\mathscr{A},\mathscr{B}) = \sup\{\Pr(F\cap G) - \Pr(F)\Pr(G) : F \in \mathscr{A},\ G \in \mathscr{B}\}.$$

The mixing coefficients of a stochastic process X, endowed with its natural filtration, are defined as $\alpha_X(s) := \sup_t \alpha(\mathcal{F}^t_{-\infty}, \mathcal{F}^{\infty}_{t+s})$ for $s > 0$, with $\mathcal{F}^t_{-\infty} = \sigma(X_u : u \leq t)$ and $\mathcal{F}^{\infty}_{t+s} = \sigma(X_u : u \geq t+s)$. The mixing coefficients of a process measure the dependence between events in terms of units of time that they are apart. For a stationary Markov process the mixing coefficients simplify to $\alpha(s) = \alpha(\sigma(X_0), \sigma(X_s))$, as shown in for example [13, page 118].

Theorem 7.1 ([53, Theorem 4]). *Let $\xi = (\xi_k)_{k=1}^{\infty}$ be a stationary sequence taking values in $\mathbb{R}^d$ with mean zero and $\sup_k \mathbb{E}\|\xi_k\|^p \leq 1$, for some $\delta \in (0,1]$. Moreover, let α_ξ the α-mixing coefficients of ξ decay polynomially with rate $n^{-(1+\varepsilon)(1+2/\delta)}$ for some $\varepsilon > 0$. Then we can redefine ξ on a new probability space on which we can also construct a d-dimensional Brownian motion W with covariance matrix Σ_ξ, with absolutely converging entries*

$$(\Sigma_\xi)_{ij} = \mathbb{E}[\xi_{i1}\xi_{j1}] + \sum_{k=2}^{\infty} \mathbb{E}[\xi_{i1}\xi_{jk}] + \sum_{k=2}^{\infty} \mathbb{E}[\xi_{ik}\xi_{j1}], \text{ for } 1 \leq i,j \leq p,$$

such that

$$\left\| \sum_{k=1}^{n} \xi_k - W(n) \right\| = O(n^{1/2-\lambda_\xi}) \text{ a.s.}$$

for some $\lambda_\xi \in (0,1/2)$ depending only on ε, δ and d.

The following lemmata are useful in the proof of Theorem 4.1.

Lemma 7.2 ([31, Theorem F.3.3]). *Let X be an ergodic Markov process with initial distribution μ and rate of convergence to stationarity given by Ψ, then $\alpha_X(s)$, the α-mixing coefficients of the process X, decay according to Ψ, i.e., for all $s \geq 0$ we have that*

$$\alpha_X(s) \leq \mu(V)\Psi(s),$$

where Ψ and V are as stated in (3.2).

Lemma 7.3 ([28] and [77]). *Let $(\Omega, \mathscr{F}, \Pr)$ be a probability space and $\mathscr{A}$ and $\mathscr{B}$ be two sub σ-algebras and consider random variables X and Y that are measurable with respect to these σ-algebras respectively. Moreover, assume that $X \in L^p(\Pr)$ and $Y \in L^q(\Pr)$, for some $p, q \geq 1$. Then we can bound their covariance in terms of the α-mixing coefficients as follows*

$$|\mathrm{Cov}(X,Y)| \leq 8\alpha\left(\mathscr{A},\mathscr{B}\right)^{1/r} \|X\|_p \|Y\|_q, \text{ with } p,q,r \in [1,\infty] \text{ and } \frac{1}{p}+\frac{1}{q}+\frac{1}{r} = 1.$$

Lemma 7.4 ([73, Corollary 1]). *Let W denote a d-dimensional Brownian motion and let $\|W_t\|$ denote the corresponding Bessel process, then we have that*

$$\mathbb{P}\left(\max_{T\in[0,1]} \|W(T)\| > u\right) = \frac{\pi^{(d-1)/2}}{2^{d/2-1}\Gamma(d/2)} u^{d-2} e^{-u^2/2}(1+o(1)),$$

as $u \to \infty$.

Following a traditional blocking argument it is now straightforward to show that the result of [53] also holds for continuous-time ergodic processes.

7.1.1. *Proof of Theorem 4.1*

Proof. Firstly, assume that we have a stationary process, i.e., our initial distribution is equal to π. For technical convenience introduce $Y = (Y_t)_{t\geq 0}$, where

$Y_t := f(X_t) - \pi(f)$ for $t \geq 0$, and $\xi = (\xi_k)_{k=1}^n$, with $n := n_T := \lfloor T \rfloor$ and $\xi_k := \int_{k-1}^k Y_t dt$ for $k = 1, \ldots, n$. Note that Y_t is a d-dimensional vector, i.e., $Y_t = (Y_{1t}, \ldots, Y_{dt})^\top$ and therefore also each ξ_k is a d-dimensional vector, $\xi_k = (\xi_{1k}, \ldots, \xi_{dk})^\top$. Furthermore, by definition n is a function of the sample size T, however, for technical convenience we suppress this. Since we are in the setting of Lemma 7.2, X has polynomially decaying α-mixing coefficients, which we will denote with $(\alpha_X(s))_{s\geq 0}$. Consequently, we have that Y and ξ are both stationary processes with polynomially decaying α-mixing coefficients $(\alpha_Y(s))_{s\geq 0}$ and $(\alpha_\xi(h))_{h\in\mathbb{N}}$ respectively. This can easily be seen by observing that $\sigma(f(X_t)) \subseteq \sigma(X_t)$ and $\sigma(\xi_k) \subseteq \sigma(X_s : k-1 \leq s \leq k)$. In order to show that a strong invariance principle holds for Y, we will show that it holds for ξ and determine the growth rate of the corresponding remainder terms. Moment conditions for ξ are directly inherited by the assumed moment conditions for X. By an application of Jensen's inequality we see that for $p = 2 + \delta$ we have that

$$\pi\left(\|\xi_k\|^p\right) = \mathbb{E}_\pi \left\| \int_{k-1}^k Y_s ds \right\|^p \leq \mathbb{E}_\pi \int_{k-1}^k \|Y_s\|^p ds = \pi\left(\|f - \pi(f)\|^p\right) < \infty.$$

Therefore, by Theorem 7.1, we can redefine ξ on a new probability space on which we can also construct a d-dimensional Brownian motion W with covariance matrix Σ_ξ, with absolutely converging entries

$$(\Sigma_\xi)_{ij} = \mathbb{E}[\xi_{i1}\xi_{j1}] + \sum_{k=2}^\infty \mathbb{E}[\xi_{i1}\xi_{jk}] + \sum_{k=2}^\infty \mathbb{E}[\xi_{ik}\xi_{j1}], \text{ for } 1 \leq i, j \leq d,$$

such that

$$\left\| \sum_{k=1}^n \xi_k - W(n) \right\| = O(n^{1/2-\lambda_\xi}) \text{ a.s.}$$

for some $\lambda_\xi \in (0, 1/2)$ depending only on ε, δ and d. The claim follows if we show that for any $\varepsilon > 0$ we have that

$$\left\| \sum_{k=1}^n \xi_k - \int_0^T Y_t dt \right\| = O(T^{1/p+\varepsilon}) \text{ a.s. for } T \to \infty, \tag{7.1}$$

$$\|W_T - W_n\| = o(T^{1/p+\varepsilon}) \text{ a.s. for } T \to \infty, \text{ and} \tag{7.2}$$

$$\Sigma_f = \Sigma_\xi. \tag{7.3}$$

In order to show that (7.1) holds, we note that

$$\left\| \int_0^T Y_s ds - \sum_{k=1}^n \xi_k \right\| = \left\| \int_n^T Y_s ds \right\| \leq \int_n^{n+1} \|Y_s\| ds. \tag{7.4}$$

By a Borel-Cantelli argument, it will follow that

$$\int_n^{n+1} \|Y_s\|\, ds = O(n^{1/p+\varepsilon}) = O(T^{1/p+\varepsilon}) \quad \text{a.s. for } T \to \infty. \tag{7.5}$$

Indeed, let $\varepsilon > 0$ be given and introduce the event

$$A_{n,\varepsilon} = \left\{ \int_n^{n+1} \|Y_s\| ds > n^{(1+\varepsilon)/p} \right\}.$$

By Markov's inequality it follows that the introduced sequence of events satisfies

$$\begin{aligned}\sum_{n=1}^{\infty} \mathbb{P}_\pi\left(A_{n,\varepsilon}\right) &\leq \sum_{n=1}^{\infty} \mathbb{P}_\pi\left(\int_n^{n+1} \|Y_s\|^p ds > n^{1+\varepsilon}\right)\\ &\leq \pi(\|f-\pi(f)\|)^p \sum_{n=1}^{\infty} \frac{1}{n^{1+\varepsilon}} < \infty.\end{aligned}$$

The Borel-Cantelli lemma implies that $\mathbb{P}_\pi(\limsup A_{n,\varepsilon}) = 0$, and consequently that $\mathbb{P}_\pi(\liminf A^c_{n,\varepsilon}) = 1$, which proves (7.5). A similar Borel-Cantelli argument also shows that (7.2) holds. Introduce the sequence of events

$$B_{n,\varepsilon} = \left\{ \sup_{n \leq T \leq n+1} \|W(T) - W(n)\| > n^{(1+\varepsilon)/q} \right\},$$

for given $\varepsilon > 0$ and some $q > p$. Since all moments of $\sup_{n \leq T \leq n+1} \|W(T) - W(n)\|$ are finite, we have by Markov's inequality that the introduced sequence of events satisfies

$$\begin{aligned}\sum_{n=1}^{\infty} \Pr\left(B_{n,\varepsilon}\right) &\leq \sum_{n=1}^{\infty} \Pr\left(\sup_{n \leq T \leq n+1} \|W_T - W_n\|^q > n^{1+\varepsilon}\right)\\ &\leq \sum_{n=1}^{\infty} \Pr\left(\sup_{0 \leq T \leq 1} \|W_T - W_0\|^q > n^{1+\varepsilon}\right)\\ &\leq \mathbb{E}\left[\sup_{0 \leq T \leq 1} \|W(T)\|^q\right] \sum_{n=1}^{\infty} \frac{1}{n^{1+\varepsilon}} < \infty.\end{aligned}$$

Let W denote a d-dimensional Brownian motion and let $\|W_t\|$ denote the corresponding Bessel process, then we have by Lemma 7.4 that for $q > p$

$$\Pr\left(\max_{T\in[0,1]} \|W(T)\| > u\right) = \frac{\pi^{(d-1)/2}}{2^{d/2-1}\Gamma(d/2)} u^{d-2} e^{-u^2/2}(1 + o(1)),$$

as $u \to \infty$. This implies the existence of all moments of the maximum of the Bessel process, since for all q we have that for all $\varepsilon' > 0$ we can find an M sufficiently large such that

$$\begin{aligned}&\mathbb{E}\left[\left(\max_{T\in[0,1]} \|W(T)\|\right)^q\right]\\ &\quad = \mathbb{E}\left[\max_{T\in[0,1]} \|W(T)\|^q\right]\end{aligned}$$

$$
\begin{aligned}
&= \int_0^\infty q u^{q-1} \Pr\left(\max_{T\in[0,1]} \|W(T)\| > u\right) du \\
&\le \int_0^M q u^{q-1} du + \int_M^\infty q u^{q-1} \Pr\left(\max_{T\in[0,1]} \|W(T)\| > u\right) du \\
&\le M^q + \frac{q\pi^{(d-1)/2}}{2^{d/2-1}\Gamma(d/2)} \int_M^\infty u^{q+d-3} e^{-u^2/2} du (1+\varepsilon') < \infty.
\end{aligned}
$$

By a Borel-Cantelli argument we see that

$$
\sup_{n\le T\le n+1} \|W(T) - W(n)\| = O(n^{1/q}) = o(T^{1/p}) \text{ a.s. for } T \to \infty.
$$

Therefore the term (7.2) will be asymptotically negligible. Finally, we see that by Lemma 7.3 the asserted asymptotic variance Σ_f is finite, i.e., all entries

$$
(\Sigma_f)_{ij} = \int_0^\infty \mathrm{Cov}_\pi(f_i(X_0), f_j(X_s))\, ds + \int_0^\infty \mathrm{Cov}_\pi(f_i(X_s), f_j(X_0))\, ds, \quad (7.6)
$$

for $1 \le i, j \le d$ converge absolutely. Indeed, since α-mixing sequences are monotonically decreasing and bounded by $1/4$, an application of Lemma 7.3 gives us

$$
\begin{aligned}
&\int_0^\infty |\mathrm{Cov}_\pi(f_i(X_0), f_j(X_s))| ds \\
&\quad \le 8 \int_0^\infty \alpha_X(s)^{\delta/p} \pi(|Y_{i0}|^p)^{1/p} \pi(|Y_{js}|^p)^{1/p} ds \\
&\quad \le 8\pi(|Y_{i0}|^p)^{1/p} \pi(|Y_{j0}|^p)^{1/p} \left(\frac{1}{4} + \pi(V)^{\delta/p} \int_1^\infty \Psi(s)^{\delta/p} ds\right),
\end{aligned}
$$

which is finite since the integral converges due to the rate of polynomial ergodicity:

$$
\int_1^\infty \Psi(s)^{\delta/p} ds \le \int_1^\infty (1+s)^{-\frac{\delta}{p}(1+\varepsilon)(1+2/\delta)} ds < \infty,
$$

since $\frac{\delta}{p}(1+\varepsilon)(1+2/\delta) > 1$. The second term of (7.6) is treated similarly. In order to show that $\Sigma_f = \Sigma_\xi$, we will show that all entries are equal. Firstly, we decompose the asymptotic covariance matrix as follows

$$
\begin{aligned}
&\lim_{T\to\infty} \mathrm{Var}_\pi\left(\frac{1}{\sqrt{T}} \int_0^T Y_t dt\right) = \lim_{T\to\infty} \mathrm{Var}_\pi\left(\frac{1}{\sqrt{T}}\left(\int_0^n Y_t dt + \int_n^T Y_t dt\right)\right) \\
&\quad = \lim_{T\to\infty} \frac{1}{T} \mathrm{Var}_\pi\left(\int_0^n Y_t dt\right) + \lim_{T\to\infty} \frac{1}{T} \mathrm{Var}_\pi\left(\int_n^T Y_t dt\right) \\
&\qquad + \lim_{T\to\infty} \frac{1}{T} \mathrm{Cov}_\pi\left(\int_0^n Y_t dt, \int_n^T Y_t dt\right) + \lim_{T\to\infty} \frac{1}{T} \mathrm{Cov}_\pi\left(\int_n^T Y_t dt, \int_0^n Y_t dt,\right).
\end{aligned}
\quad (7.7)
$$

Let $\Sigma_{T1}, \Sigma_{T2}, \Sigma_{T3}$ and Σ_{T4} denote the four terms in (7.7). We will show that the entry-wise convergence gives us the desired result. For $1 \leq i, j \leq d$ we obtain the following expressions for the elements of the matrices in (7.7):

$$(\Sigma_{T1})_{ij} = \frac{1}{T}\int_0^n \int_0^n \mathrm{Cov}_\pi(Y_{it}, Y_{js})\, dtds, \tag{7.8}$$

$$(\Sigma_{T2})_{ij} = \frac{1}{T}\int_n^T \int_n^T \mathrm{Cov}_\pi(Y_{it}, Y_{js})\, dtds, \tag{7.9}$$

$$(\Sigma_{T3})_{ij} = \frac{1}{T}\int_0^n \int_n^T \mathrm{Cov}_\pi(Y_{it}, Y_{js})\ dtds, \tag{7.10}$$

$$(\Sigma_{T4})_{ij} = \frac{1}{T}\int_0^n \int_n^T \mathrm{Cov}_\pi(Y_{is}, Y_{jt})\, dtds. \tag{7.11}$$

We see that $(\Sigma_{T1})_{ij}$ tends to the asymptotic variance $(\Sigma_f)_{ij}$ as $T \to \infty$, since

$$\left(\mathrm{Var}_\pi\left(\frac{1}{\sqrt{n}}\sum_{k=1}^n \xi_k\right)\right)_{ij} = \frac{T}{n}\cdot\frac{1}{T}\int_0^n \int_0^n \mathrm{Cov}_\pi(Y_{it}, Y_{js})\, dtds.$$

Finally, we claim that $(\Sigma_{T2})_{ij}, (\Sigma_{T3})_{ij}$ and $(\Sigma_{T4})_{ij}$ tend to zero as $T \to \infty$. An application of Lemma 7.2 and Lemma 7.3 gives us that

$$\begin{aligned}\frac{1}{T}\int_0^n \int_n^T |\mathrm{Cov}_\pi(Y_{it}, Y_{js})|\, dtds &\leq C_{f,V}\frac{1}{T}\int_0^n \int_n^T \Psi(|t-s|)^{1-2/p}\, dtds \\ &= C_{f,V}\frac{1}{T}\int_0^n \int_n^T (1+t-s)^{-\beta\delta/p}\, dtds,\end{aligned}$$

where $C_{f,V} = 8\pi(V)^{\delta/p}\pi(|Y_{i0}|^p)^{1/p}\pi(|Y_{j0}|^p)^{1/p} < \infty$ and the last equality follows since we assumed polynomial ergodicity of degree β and since we always have $t \geq s$ on the considered integration region. Since $\beta\delta/p > 1$, it follows that

$$\begin{aligned}\int_0^n \int_n^T (1+t-s)^{-\beta\delta/p}\, dtds &\leq (T-n)\int_0^n \sup_{t\in[n,T]} (1+t-s)^{-\beta\delta/p}\, ds \\ &\leq (T-n)\int_0^n (1+n-s)^{-\beta\delta/p}\, ds \\ &= \frac{(T-n)}{\beta\delta/p-1}\left(1-\frac{1}{(1+n)^{\frac{\beta\delta}{p}-1}}\right).\end{aligned}$$

Consequently, it follows that

$$\frac{1}{T}\int_0^n \int_n^T |\mathrm{Cov}_\pi(Y_{it}, Y_{js})|\, dtds \leq C_{f,V}\frac{T-n}{T}\frac{p}{\beta\delta-p}\left(1-\frac{1}{(1+n)^{\frac{\beta\delta}{p}-1}}\right) = o(1).$$

By the same argument, we have that

$$\frac{1}{T}\int_0^n \int_n^T |\mathrm{Cov}_\pi(Y_{is}, Y_{jt})|\, dtds = o(1).$$

Finally, we also have that

$$\begin{aligned}\frac{1}{T}\int_n^T\int_n^T |\mathrm{Cov}_\pi(Y_{it},Y_{js})|\,dtds &\le C_{f,V}\frac{1}{T}\int_n^T\int_n^T (1+|t-s|)^{-\beta\delta/p}\,dtds\\ &\le C_{f,V}\frac{(n-T)^2}{T}\sup_{(s,t)\in[n,T]\times[n,T]}(1+|t-s|)^{-\beta\delta/p}\\ &= C_{f,V}\frac{(n-T)^2}{T}=o(1).\end{aligned}$$

Hence we have shown that $(\Sigma_{T2})_{ij}, (\Sigma_{T3})_{ij}$ and $(\Sigma_{T4})_{ij}$ tend to zero as T tends to infinity and thus we have that $\Sigma_f = \Sigma_\xi$. Note that we have now proven our result assuming stationarity, i.e., with initial distribution π. However, by following the argument of [63, Proposition 17.1.6] it follows that the strong invariance principle holds for every initial distribution. Let

$$h(x)=\mathbb{P}_x\left(\left\|\int_0^T[f(X_t)-\pi(f)]dt\,dt-\Sigma_f^{1/2}W(T)\right\|=O(\psi_T)\middle|X_0=x\right).$$

We have currently shown that the strong approximation results holds for initial distribution π, i.e.,

$$\int h(x)\pi(dx)=1.$$

Now we will show that h is harmonic, i.e., $h(x)=P_sh(x)$. Indeed, for every x in E and $s\ge 0$ we have

$$\begin{aligned}&P_sh(x)\\ &=\int_E P_s(x,dy)h(y)\\ &=\mathbb{E}_xh(X_s)\\ &=\mathbb{E}\left[\mathbb{P}_x\left(\left\|\int_s^{s+T}[f(X_t)-\pi(f)]dt-\Sigma_f^{1/2}W(T)\right\|=O(\psi_T)\middle|X_s=y\right)\middle|X_0=x\right]\end{aligned}$$

By the Markov property and the tower property of conditional expectation, we have that

$$\begin{aligned}&P_sh(x)\\ &=\mathbb{E}\Bigg[\mathbb{P}_x\Bigg(\left\|\int_s^{s+T}[f(X_t)-\pi(f)]dt-\Sigma_f^{1/2}W(T)\right\|\\ &\qquad\qquad=O(\psi_T)\Big|X_s=y;X_0=x\Bigg)\Bigg|X_0=x\Bigg]\\ &=\mathbb{P}_x\left(\left\|\int_s^{s+T}[f(X_t)-\pi(f)]dt-\Sigma_f^{1/2}W(T)\right\|=O(\psi_T)\middle|X_0=x\right)\end{aligned}$$

$$= h(x),$$

where the last inequality follows since for all fixed $s \geq 0$ we have, by the same argument of (7.5), that

$$\int_0^s [f(X_t) - \pi(f)]dt - \Sigma_f^{1/2} W(s) \text{ and } \int_T^{T+s} [f(X_t) - \pi(f)]dt - \Sigma_f^{1/2} W(s)$$

are $O(\psi_T)$ almost surely. By [50, Theorem 20.10], we have that for ergodic Markov processes every bounded harmonic function is constant, hence it follows that $h(x) = 1$ for all $x \in E$. It immediately follows that for every initial distribution ν we have that

$$\begin{aligned}\mathbb{P}_\nu & \left(\left\| \int_0^T f(X_t)\, dt - T\pi(f) - \Sigma_f^{1/2} W(T) \right\| = O(\psi_T) \right) \\ &= \int_E \mathbb{P}_x \left(\left\| \int_0^T f(X_t)\, dt - T\pi(f) - \Sigma_f^{1/2} W(T) \right\| = O(\psi_T) \middle| X_0 = x \right) \nu(dx) = 1\end{aligned}$$

Hence the strong invariance principle holds for every initial distribution. □

7.2. Proposition 4.3

In [5] a strong invariance principle for weakly m-dependent processes is given, which are defined as processes that can be approximated by m-dependent processes in the L^p-sense, with a sufficiently decaying approximation error (rate function in terminology of [5]). Their strong invariance principle, stated in Theorem 7.5, is obtained through a classical blocking argument for m-dependent random variables. By dividing an m-dependent sequence into non-overlapping long and short blocks, two sequences of independent random variables are obtained; these can both be approximated by a Brownian motion. Trivially, stationary m-dependent processes satisfying appropriate moment conditions fall into their framework. For more details we refer to [5].

Theorem 7.5 ([5, Theorem 2])**.** *Let $\xi = (\xi_k)_{k=1}^\infty$ be a centered stationary sequence with $\sup_k \mathbb{E}|\xi_k|^p < \infty$, for some $\delta > 0$. Moreover, let ξ be weakly m-dependent in L^p with an exponentially decaying rate function κ, i.e.,*

$$\kappa(m) \ll \exp(-cm), \quad \textit{for some } c > 0.$$

Then the series

$$\sigma_\xi^2 = \sum_{k=0}^\infty \mathbb{E}\xi_0\xi_k$$

is absolutely convergent, and we can redefine ξ on a new probability space on which we can also construct two standard Brownian motions W_1 and W_2 such that

$$\left| \sum_{k=1}^n \xi_k - W_1(s_n^2) - W_2(t_n^2) \right| = O(n^{1/p} \log^2 n) \ \textit{a.s.},$$

where $\{s_n\}$ *and* $\{t_n\}$ *are non-decreasing deterministic sequences with*

$$s_n^2 = \sigma_\xi^2 n + O(n/\log n),$$
$$t_n^2 = O(n/\log n),$$

and $\limsup_n(s_{n+1}^2 - s_n^2) = \limsup_n(t_{n+1}^2 - t_n^2) = \sigma_\xi^2$.

As noted by [5] the perturbed time sequences $\{s_n\}$ and $\{t_n\}$ are deterministic and can be explicitly calculated.

7.2.1. Proof of Proposition 4.3

Proof. Firstly, assume that the initial distribution of X is equal to ν. By Proposition 3.1 we see that we can redefine our process such that it is embedded in a richer process Z. We will identify X as the first coordinate of the process Z. Following Proposition 3.2, we introduce the sequence of stopping times (S_n, R_n) defined as $S_0 = R_0 := 0$ and

$$S_{n+1} := \inf\{T_m > R_n : Z_{T_m} \in A\} \quad \text{and} \quad R_{n+1} := \inf\{T_m : T_m > S_{n+1}\}.$$

Then Z_{R_n} is independent of $\mathcal{F}_{R_{n-1}}$ for all $n \geq 1$ and $(Z_{R_n})_{n\geq 1}$ is an i.i.d sequence with

$$Z_{R_n} \sim \nu(dx)\lambda(du)K((x,u),dx') \quad \text{for all } n \geq 1,$$

where λ denotes the law of a standard Uniform random variable. As a direct consequence, the sequence $\{\xi_n\}_n$ defined as

$$\xi_n := \int_{R_{n-1}}^{R_n} \{f(X_s) - \pi(f)\}\, ds, \quad n \geq 1, \tag{7.12}$$

is stationary under $\mathbb{P}_\nu$. Moreover, by Proposition 3.3 for $n \geq 2$, ξ_n is independent of $\mathcal{F}_{R_{n-2}}$. Let $N(T)$ denote the number of regenerations of the resolvent chain up to time T, namely

$$N(T) = \max\{k : R_k \leq T\}.$$

It immediately follows that

$$\int_0^T \{f(X_s) - \pi(f)\}\, ds = \sum_{k=1}^{N(T)} \xi_k + \int_{R_{N(T)}}^T \{f(X_s) - \pi(f)\}\, ds.$$

Consequently, we have that

$$\left| \int_0^T \{f(X_s) - \pi(f)\}\, ds - \sum_{k=1}^{N(T)} \xi_k \right| \leq \int_{R_{N(T)}}^T |f(X_s) - \pi(f)| ds. \tag{7.13}$$

By an argument analogous to the one given in Theorem 4.1 for the remainder term defined in (7.4), we will show that

$$\int_{R_{N(T)}}^T |f(X_s) - \pi(f)| ds = O(T^{1/p}) \quad \text{a.s. for } T \to \infty. \tag{7.14}$$

In order to show that (7.14) holds, we note that

$$\int_{R_{N(T)}}^{T} |f(X_s) - \pi(f)| ds \leq \int_{R_{N(T)}}^{R_{N(T)+1}} |f(X_s) - \pi(f)| ds. \tag{7.15}$$

By a Borel-Cantelli argument it will follow that

$$\int_{R_{N(T)}}^{R_{N(T)+1}} |f(X_s) - \pi(f)| ds = O(T^{1/p}) \quad \text{a.s. for } T \to \infty. \tag{7.16}$$

Indeed, introduce the event

$$A_n = \left\{ \int_{R_n}^{R_{n+1}} |f(X_s) - \pi(f)| ds > n^{1/p} \right\}.$$

By the stationarity of $\{\xi_n\}_{n\in\mathbb{N}}$ under $\mathbb{P}_\nu$ it follows that the introduced sequence of events satisfies

$$\begin{aligned}
\sum_{n=1}^{\infty} \mathbb{P}_\nu\left(A_{n,\varepsilon}\right) &= \sum_{n=1}^{\infty} \mathbb{P}_\nu\left(\left| \int_{R_n}^{R_{n+1}} |f(X_s) - \pi(f)| ds \right|^p > n \right) \\
&= \sum_{n=1}^{\infty} \mathbb{P}_\nu\left(\left| \int_{0}^{R_1} |f(X_s) - \pi(f)| ds \right|^p > n \right) \\
&\leq \mathbb{E}_\nu \left| \int_{0}^{R_1} |f(X_s) - \pi(f)| ds \right|^p < \infty.
\end{aligned}$$

The Borel-Cantelli lemma states that $\mathbb{P}_\nu(\limsup A_n) = 0$. Consequently, we have that $\mathbb{P}_\nu(\liminf A_n^c) = 1$. Hence it follows that

$$\int_{R_n}^{R_{n+1}} |f(X_s) - \pi(f)| ds = O(n^{1/p}) \quad \text{a.s.} \tag{7.17}$$

Moreover, since $N(T)$ is almost surely increasing and $N(T) = O(T)$, as shown in (7.21), it follows that

$$\int_{R_{N(T)}}^{R_{N(T)+1}} |f(X_s) - \pi(f)| ds = O(N(T)^{1/p}) = O(T^{1/p}) \quad \text{a.s.}$$

Hence proving the claim formulated in (7.16) and as a direct consequence also the bound stated in (7.14). Furthermore, by Proposition 3.3, the sequence $\{\xi_k\}_{k=1}^{\infty}$ is a stationary m-dependent sequence. By the imposed moment conditions and stationarity, we have by the reasoning given in [5, Section 3.1] that $\{\xi_k\}_{k=1}^{\infty}$ is also a weakly m-dependent process with a rate function $\kappa(m)$ equal to zero for $m \geq 1$. Hence by Theorem 7.5, we can redefine $(\xi_k)_k$ on a new probability space on which we can also construct two standard Brownian motions W_1 and W_2 such that

$$\left| \sum_{k=1}^{n} \xi_k - n\mathbb{E}_\nu \xi_1 - W_1(s_n^2) - W_2(t_n^2) \right| = O(n^{1/p} \log^2 n) \text{ a.s.,} \tag{7.18}$$

where $\{s_n\}$ and $\{t_n\}$ are increasing deterministic sequences with $s_n^2 = \sigma_\xi^2 n + O(n/\log n)$ and $t_n^2 = O(n/\log n)$. Note that by Proposition 3.4 we have that

$$\pi(f) = \frac{1}{\varrho}\, \mathbb{E}_\nu \int_0^{R_1} f(X_s) ds.$$

Hence

$$\mathbb{E}_\nu \xi_1 = \mathbb{E}_\nu \int_0^{R_1} \{f(X_s) - \pi(f)\}\, ds = \varrho \cdot \pi(f - \pi(f)) = 0.$$

Furthermore, by definition of big O in (7.18), there exists an almost surely finite random variable C such that for almost all sample paths ω we have that for all $n \geq N_0 \equiv N_0(\omega)$ we have that

$$\frac{1}{n^{1/p}\log^2 n}\left|\sum_{k=1}^{n} \xi_k(\omega) - W_1(s_n^2, \omega) - W_2(t_n^2, \omega)\right| < C(\omega) \tag{7.19}$$

Since we have that $\mathbb{E}_\nu R_1^q < \infty$, by [20, Theorem 2.4] with $q = \beta - 1$, we can construct a Brownian motion $\tilde{W}$ such that

$$\left|N(T) - \frac{T}{\varrho} - \frac{\mathrm{Var}_\nu(R_1)}{\varrho^{3/2}}\tilde{W}_T\right| = o(T^{1/q}). \tag{7.20}$$

By the law of iterated logarithm for Brownian motion we obtain

$$N(T) = \frac{T}{\varrho} + O(\sqrt{T\log\log T}) \quad \text{a.s.} \tag{7.21}$$

Since $N(T)$ is almost surely increasing and tends to infinity, we have that for almost every sample path ω there exists a $T_0 \equiv T_0(\omega)$ such that $N(T)(\omega) \geq N_0$ for all $T \geq T_0$. Hence we obtain from (7.19) that

$$\limsup_{T\to\infty} \frac{|\sum_{k=1}^{N(T)} \xi_k - W_1(s_{N(T)}^2) - W_2(t_{N(T)}^2)|}{N(T)^{1/p}\log^2(N(T))} < C \quad \text{a.s.}, \tag{7.22}$$

where $s_{N(T)}^2$ and $t_{N(T)}^2$ are almost surely increasing sequences, which given $N(T)$ are deterministic with

$$\begin{aligned} s_{N(T)}^2 &= \sigma_\xi^2 N(T) + O(N(T)/\log N(T)) \\ t_{N(T)}^2 &= O(N(T)/\log N(T)). \end{aligned}$$

We see that (7.22) can be reformulated as

$$\left|\sum_{k=1}^{N(T)} \xi_k - W_1(s_{N(T)}^2) - W_2(t_{N(T)}^2)\right| = O(N(T)^{1/p}\log^2 N(T))) \quad \text{a.s.} \tag{7.23}$$

$$= O(T^{1/p}\log^2 T)) \quad \text{a.s.} \tag{7.24}$$

Here the second equality follows by (7.21). Furthermore, the asymptotic behaviour of $N(T)$ motivates the introduction of σ_T^2, τ_T^2 defined as

$$\begin{aligned} \sigma_T^2 &= s_n^2/\varrho, \quad \text{for } T \in [n, n+1), \\ \tau_T^2 &= t_n^2/\varrho, \quad \text{for } T \in [n, n+1). \end{aligned}$$

By Theorem 6.1 (see also Theorem 1.2.1 of [22]) we see that

$$\left|W_1(s_{N(T)}^2) - W_1(\sigma_T^2)\right| \text{ and } \left|W_2(t_{N(T)}^2) - W_1(\tau_T^2)\right| \text{ are both } O(T^{1/4}\log T) \text{ a.s.} \tag{7.25}$$

with

$$\sigma_T^2 = \frac{\sigma_\xi^2}{\varrho} T + O(T/\log T) \text{ and } \tau_T^2 = O(T/\log T).$$

Combining results (7.14), (7.23), and (7.25) concludes the proof. By the same arguments given in the proof of Theorem 4.1, the strong invariance principle holds for every initial distribution. ☐

Furthermore, we have by [5, Proposition 1] that

$$\operatorname{Corr}(W_1(s_n), W_2(t_m)) \to 0 \quad \text{as } m, n \to \infty.$$

Hence (4.13) immediately follows.

7.3. Theorem 4.4

For this proof, we will rely on the following properties of the resolvent chain. Granted that the process X is aperiodic and positive Harris recurrent, then also the resolvent $\bar{X}$ will inherit these properties, as seen in [63, Propostion 5.4.5] and [87, Theorem 3.1] respectively. Moreover, by [32, Theorem 5.3], exponential convergence to stationarity is equivalent for X and $\bar{X}$. The split chain of the resolvent in turn obtains aperiodicity and positive Harris recurrence from $\bar{X}$, as seen in for example [68]. Following a co-de-initialising argument of [78], we see that the split chain inherits the rate of convergence of the resolvent chain. To conclude, we see that the split chain inherits aperiodicity, positive Harris recurrence, and the rate of ergodicity from the process X.

Note that by Proposition 3.1 $(Z_{T_n}^1, Z_{T_n}^2)_n$, the jump chain of the first two coordinates of Z, has the same distribution as the split chain of the resolvent. From (3.4) and (3.5) we see that $(Z_{T_n}^1, Z_{T_n}^2)_n$ is a Markov chain taking values in $E' := E \times [0, 1]$ that moves according to the kernel

$$U'((x,u),(dy,dv)) = \nu(dy)\lambda(dv)\mathbb{1}_{\{u \le \alpha \mathbb{1}_C(x)\}} + W(x,dy)\lambda(dv)\mathbb{1}_{\{u > \alpha \mathbb{1}_C(x)\}}, \tag{7.26}$$

where λ denotes Lebesgue measure on the unit interval. Observe that the kernel of the split chain of the resolvent also satisfies a one-step minorisation condition $U' \ge s \otimes \nu \otimes \lambda$, i.e.,

$$U'((x,u),(dy,dv)) \ge s(x,u)\nu(dy)\lambda(dv), \tag{7.27}$$

where

$$s(x,u) = \mathbb{1}_{\{u \leq \alpha \mathbb{1}_C(x)\}}.$$

Moreover, the split chain of the resolvent is aperiodic, positive Harris recurrent and inherits the rate of convergence to stationarity from X.

Lemma 7.6 ([46, Lemma 1])**.** *Let $(X_t)_{t\geq 0}$ be a positive Harris recurrent Markov process with invariant distribution π. Let U denote the transition kernel of the resolvent chain of X and assume that the following minorisation condition holds:*

$$U(x,dy) \geq \alpha \mathbb{1}_C(x)\nu(dy). \tag{7.28}$$

Then for any π-integrable function $g : E^{[0,\infty)} \to \mathbb{R}$ we have the following inequality holds

$$\mathbb{E}_\pi|g| \geq c\ \mathbb{E}_\nu|g|, \tag{7.29}$$

where $c = \alpha\pi(C)$.

Proof. Since the resolvent chain has the same stationary distribution as the process X, i.e., $\pi = \pi U$, the claim follows with the identical argument of [46, Lemma 1]. □

7.3.1. Proof of Theorem 4.4

Proof. Firstly, by the construction of the randomised stopping times $(S_n)_n$ and $(R_n)_n$ we see that $R_n = S_n + \sigma_{n+1}$, where σ_{n+1} has a standard exponential distribution. Hence, by the triangle inequality in $L^q(\pi)$ we only need to show that $\mathbb{E}_\pi[{S_1}^q] < \infty$, with

$$S_1 = \inf\{T_n : Z_{T_n} \in C \times [0,\alpha] \times E\}$$

Let $\bar{Z} = (\bar{Z}_n)_n$ denote the jump chain of the process Z, i.e., $\bar{Z}_n = Z_{T_n}$, where the $(T_n)_n$ denote the jump times. Let $\bar{X} = (\bar{X}_n)_{n\geq 0}$ again denote the resolvent chain. Let N_t denote the number of jumps up to time t. Let $\bar{\tau}_A$ denote the hitting time of the recurrent atom for jump chain $\bar{Z}$, i.e.,

$$\bar{\tau}_A := \inf\{n \geq 0 : \bar{Z}_n \in A\} = \inf\{n \geq 0 : \bar{Z}_n \in C \times [0,\alpha] \times E\}.$$

For technical convenience, we introduce $q := \beta - 1$, note that by the assumed ergodicity assumptions we have that $q > p(p+\varepsilon)/\varepsilon$. From the relation between the expectation of positive random variables and tail probabilities we can express the expectation of interest as follows

$$\mathbb{E}_\pi S_1^q = \int_0^\infty q t^{q-1} \mathbb{P}_\pi\left(S_1 > t\right) dt$$
$$= \int_0^\infty q t^{q-1} \sum_{m=0}^\infty \mathbb{P}_\pi\left(\bar{\tau}_A > m; N_t = m\right) dt$$

$$= \int_0^\infty q t^{q-1} \sum_{m=0}^\infty \left(\underbrace{\mathbb{P}_\pi\left(\bar{\tau}_A > m; N_t = m; \bar{Z}_0 \in A\right)}_{=0} + \mathbb{P}_\pi\left(\bar{\tau}_A > m; N_t = m; \bar{Z}_0 \notin A\right) \right) dt$$
$$= \int_0^\infty q t^{q-1} \int_{E'} \int_{E'} \sum_{m=0}^\infty \frac{t^m}{m!} e^{-t} \sum_{k=m+1}^\infty \left(U' - \nu \otimes \lambda \otimes s\right)^k (x, dz) \mathbb{1}_A(z) \pi(dx) dt$$
$$= \int_{E'} \int_{E'} \sum_{m=0}^\infty \int_0^\infty \frac{t^{m+q-1}}{m!} q e^{-t} \sum_{k=m+1}^\infty \left(U' - \nu \otimes \lambda \otimes s\right)^k (x, dz) \mathbb{1}_A(z) dt \pi(dx)$$
$$= \int_{E'} \int_{E'} \sum_{k=1}^\infty \left(U' - \nu \otimes \lambda \otimes s\right)^k (x, dz) \mathbb{1}_A(z) q \sum_{m=0}^{k-1} \frac{\Gamma(m+q)}{m!} \pi(dx)$$
$$= \int_{E'} \int_{E'} \sum_{k=1}^\infty \frac{\Gamma(k+q)}{\Gamma(k)} \left(U' - \nu \otimes \lambda \otimes s\right)^k (x, dz) s(z) \pi(dx).$$

Here we obtained the last equality by using

$$\sum_{m=0}^{k-1} \frac{\Gamma(m+q)}{m!} = \frac{\Gamma(k+q)}{q\Gamma(k)},$$

which can easily be proven by mathematical induction and the fact that for every $k > 0$ we have that $\Gamma(k+2) = k\Gamma(k+1) + \Gamma(k+1)$. Note that $\Gamma(k+q)/\Gamma(k-1)$ can be dominated by some polynomial $\psi(k)$ with a leading term of order k^{q+1}. By [70, Proposition 1.6] we have that

$$\int_{E'} \int_{E'} \sum_{k=0}^\infty \psi(k) \left(U' - \nu \otimes \lambda \otimes s\right)^k (x, dz) s(z) \pi(dx) < \infty.$$

It follows that $\mathbb{E}_\pi S_1^q < \infty$.

For the second statement of Theorem 4.4 we follow the argument of [4, Theorem 2] with some minor adaptations. We give the proof for completion.

$$\begin{aligned}
\left[\mathbb{E}_\pi \xi_1^p\right]^{1/p} &\leq \left[\mathbb{E}_\pi \left| \int_0^{R_1} |f(X_s)| ds \right|^p \right]^{1/p} \\
&= \left[\mathbb{E}_\pi \left| \int_0^\infty |f(X_s)| \mathbb{1}_{\{R_1 \geq s\}} ds \right|^p \right]^{1/p} \\
&\leq \int_0^\infty \left[\mathbb{E}_\pi \left(|f(X_s)|^p \mathbb{1}_{\{R_1 \geq s\}} ds\right)\right]^{1/p} \\
&\leq \int_0^\infty \left[\mathbb{E}_\pi |f(X_s)|^{p+\varepsilon}\right]^{1/(p+\varepsilon)} \left[\mathbb{E}_\pi \mathbb{1}_{\{R_1 \geq s\}}\right]^{\varepsilon/p(p+\varepsilon)} ds \\
&\leq \pi\left(|f|^{p+\varepsilon}\right)^{1/(p+\varepsilon)} \int_0^\infty \left[\mathbb{P}_\pi(R_1 \geq s)\right]^{\varepsilon/p(p+\varepsilon)} ds \\
&\leq \pi\left(|f|^{p+\varepsilon}\right)^{1/(p+\varepsilon)} \left(1 + \pi(R_1^q)^{\varepsilon/p(p+\varepsilon)} \int_1^\infty s^{-\varepsilon q/p(p+\varepsilon)} ds\right) < \infty.
\end{aligned}$$

Here the inequalities follow by Minkowski's integral inequality, Hölder's inequality, stationarity, and Markov's inequality. Note that the integral on the last line is finite due to the imposed condition on the rate of polynomial ergodicity since $q = \beta - 1 > p(p+\varepsilon)/\varepsilon$. An application of Lemma 7.6 concludes the proof. $\square$

Remark 7.7. Note that if we assume polynomial ergodicity of rate $\beta > 1$, without any further requirements, then we can only guarantee the existence of moments up to order p' where $p' < \frac{1}{2}(\sqrt{\varepsilon(\varepsilon + 4(\beta-1))} - \varepsilon)$ if $p > \frac{1}{2}(\sqrt{\varepsilon(\varepsilon + 4(\beta-1))} - \varepsilon)$ and $p' = p$ otherwise. $\square$

Remark 7.8. For the exponentially ergodic case we would make use of [69, Lemma 2.8] which states that for an exponentially ergodic Markov chain there exists an $r > 1$ such that

$$\int_{E'} \int_{E'} \sum_{k=0}^{\infty} r^k \left(U' - \nu \otimes \lambda \otimes s\right)^k (x, dy) \mathbb{1}_C(y) \pi(dx) < \infty. \qquad \square$$

7.4. Theorems 4.6 and 4.7

Lemma 7.9 ([61, Lemma 2.4])**.** *Let B be a standard Brownian motion and N be a Poisson process with intensity λ, independent of B. Then there exists a standard Brownian motion W that is also independent of N such that*

$$\left| B(n) - \frac{1}{\sqrt{\lambda}} W(N(n)) \right| = O(\log(n)) \quad a.s.$$

Proof. The claim immediately follows from [61, Lemma 2.4] and a Borel-Cantelli argument. $\square$

7.4.1. Proof of Theorem 4.6

Proof. We will first assume that our initial distribution is equal to the stationary distribution. Let x_0 denote the smallest local optimum of the density π, i.e.,

$$x_0 = \min\{x : \pi'(x) = 0\}.$$

Since the tails of π are diminishing, we must have that x_0 is a local maximum. Moreover, for some $M > 0$, define the set A as follows

$$A = [x_0 - M, x_0] \times \{+1\}.$$

Note that on $(-\infty, x_0)$ the density on π is increasing, and therefore the potential $U = -\log \pi$ is decreasing and thus the derivative of U is negative. Consequently, for all $(x, v) \in (-\infty, x_0) \times \{+1\}$ we have that the switching intensity $\lambda(x, v) = (U'(x))^+ = 0$, since the process is moving toward a higher density region. If the process moves from $(-\infty, x_0 - M) \times \{+1\}$ to A, the process will thus not switch and move deterministically from A to $x_0 \times \{+1\}$ in time M. If the process hits A from $[x_0 - M, x_0] \times \{-1\}$, i.e., when the position component is in $[x_0 - M, x_0]$ and the velocity switches from -1 to $+1$, then the point $x_0 \times \{+1\}$ will be reached in time at most M. Note that these are the only possibilities for reaching the set A. We see that when the process hits A, the process must move deterministically for time at most M until the point $x_0 \times \{+1\}$ is reached and the probability of a velocity switch becomes positive. This motivates the introduction of the stopping times R_n defined as

$$R_0 = \inf\{t \geq 0 : (X_t, V_t) = (x_0, 1)\}$$

and

$$R_n = \inf\{t \geq R_{n-1} : (X_t, V_t) = (x_0, 1)\}.$$

By the Markov property, the sequence $\{\xi_n\}$ defined as

$$\xi_n := \int_{R_{n-1}}^{R_n} \{f(X_s) - \pi(f)\}\, ds, \quad n \geq 1,$$

is i.i.d under $\mathbb{P}_\nu$, with ν a Dirac measure at the point $x_0 \times \{+1\}$. Note that this argument holds for any local optimum by the smoothness assumptions on π. Note that we also have that $R_n \leq M + \tau_A$ with τ_A again denoting the hitting time of set A. Since we have that

$$\{\tau_A > t\} \subset \bigcup_{m=1}^{\infty} \{\bar{\tau}_A > m; N_t = m\},$$

where $\bar{\tau}_A$ again denotes the hitting time of the resolvent chain, we can follow the argument of Theorem 4.4 to obtain that

$$\mathbb{E}_\nu[(R_1)^{\beta-1}] < \infty.$$

Moreover, for all measurable $f : E \to \mathbb{R}$ with $\pi(|f|^{p+\varepsilon}) < \infty$ where $p \geq 1$ and $\beta > (p + 2\varepsilon)/\varepsilon$, we have that

$$\mathbb{E}_\pi \left| \int_0^{R_0} [f(X_s) - \pi(f)] ds \right|^p < \infty \text{ and } \mathbb{E}_\nu \left| \int_{R_0}^{R_1} [f(X_s) - \pi(f)] ds \right|^p < \infty.$$

Thus we see that $\xi_0 := \int_0^{R_0} [f(X_s) - \pi(f)] ds$ is asymptotically negligible. Define $(\tau_k)_{k \in \mathbb{N}}$ as $\tau_k = R_k - R_{k-1}$ and let ϱ and σ_τ^2 denote the mean and variance of this random variable. The sequence of random vectors (ξ_k, τ_k) are independent and identically distributed. If we choose $\alpha = \operatorname{Cov}_\nu(\xi_1, \tau_1)/\operatorname{Var}_\nu(\tau_1)$, then it immediately follows that $\xi_k - \alpha(\tau_k - \varrho)$ and τ_k are uncorrelated.

Applying the multivariate Komlós-Major-Tusnády approximation given in [35, Theorem 1] and [20, Theorem 2.1], there exists two independent Brownian motions B_1 and B_2 such that

$$\left| \sum_{k=1}^{n} \xi_k - \alpha\left(\sum_{k=1}^{n} \tau_k - \varrho\right) - \tilde{\sigma} B_1 \right| = o\left(\psi_n\right) \quad \text{a.s.} \tag{7.30}$$

$$|R_n - n\varrho - \sigma_\tau B_2(n)| = o\left(\psi_n\right) \quad \text{a.s.}, \tag{7.31}$$

with

$$\psi_n = n^{\max\left(\frac{1}{\beta-1}, \frac{1}{p}\right)}. \tag{7.32}$$

Note that in (7.30) we have that $\mathbb{E}_\nu \xi_1 = 0$ by Theorem 3.4 and that $\tilde{\sigma}^2 = \mathrm{Var}_\nu(\xi_1 - \alpha(\tau_1 - \varrho))$. From the assumptions on the rate of ergodicity, we see that the approximation error simplifies to $o(n^{1/p})$. By [52, Theorem 1(ii)], a Poisson Process N with intensity $\lambda = \varrho^2/\sigma_\tau^2$ can be constructed from the Brownian motion B_2 such that

$$\left|N(n) - \frac{\varrho}{\gamma}n - \frac{\sigma_\tau}{\gamma}B_2(n)\right| = O(\log n) \quad \text{a.s.}, \tag{7.33}$$

where $\gamma = \sigma_\tau^2/\varrho$ and N is constructed increment-wise from B_2 in a determinstic way and is therefore also independent of B_1. From (7.31) and (7.33) it follows that

$$|R_n - \gamma N(n)| = o(n^{1/p}) \quad \text{a.s.} \tag{7.34}$$

We claim that it therefore follows that

$$\left|\int_0^{R_n} [f(X_s) - \pi(f)]ds - \int_0^{\gamma N(n)} [f(X_s) - \pi(f)]ds\right| = O(n^{1/p}) \quad \text{a.s.} \tag{7.35}$$

Indeed, we have that

$$\left|\int_0^{R_n} [f(X_s) - \pi(f)]ds - \int_0^{\gamma N(n)} [f(X_s) - \pi(f)]ds\right| = \left|\int_{b_n}^{c_n} f(X_s) - \pi(f)ds\right|, \tag{7.36}$$

where $b_n := \min\{R_n, \gamma N(n)\}$ and $c_n := \max\{R_n, \gamma N(n)\}$. Therefore we can introduce the positive sequence α_n as follows

$$\alpha_n := c_n - b_n = |R_n - \gamma N(n)|.$$

From (7.34) it follows that $\alpha_n = o(n^{1/p})$ a.s., hence for almost every ω it holds that for all $\varepsilon_1 > 0$ there exists an $N_1 := N_1(\omega)$ such that for all $n \geq N_1$ we have that $\alpha_n < \varepsilon_1 n^{1/p}$ and hence $c_n = b_n + \alpha_n \leq b_n + \varepsilon_1 n^{1/p}$. Note that the stopping times $(R_k)_{k\geq 0}$ are regeneration epochs of the process, and hence the corresponding cycles $\mathcal{C}_k := (X_s : R_k \leq s < R_{k+1})$ are independent and identically distributed. Let $\eta(T) := \max\{k : R_k \leq T\}$ denote the number of regenerative cycles up to time T and let $Y_k = \int_{R_k}^{R_{k+1}} |f(X_s) - \pi(f)|ds$. Then we see that for $n > N_1(\omega)$ we have that

$$\left|\frac{1}{n^{\frac{1}{p}}} \int_{b_n}^{c_n} f(X_s) - \pi(f)ds\right|$$

$$
\begin{aligned}
&= \frac{1}{n^{\frac{1}{p}}}\left|\int_0^{c_n-b_n} f(X_{b_n+u}) - \pi(f)du\right| \\
&\le \frac{1}{n^{\frac{1}{p}}}\int_0^{\alpha_n} |f(X_{b_n+u}) - \pi(f)|du \\
&\le \frac{1}{n^{\frac{1}{p}}}\int_0^{\varepsilon_1 n^{1/p}} |f(X_{b_n+s}) - \pi(f)|ds \\
&\le \frac{1}{n^{\frac{1}{p}}}\sum_{j=\eta(b_n)}^{\eta(b_n+\varepsilon_1 n^{1/p})} Y_j + \frac{1}{n^{\frac{1}{p}}}\int_{R_{\eta(b_n+\varepsilon_1 n^{1/p})}}^{b_n+\varepsilon_1 n^{1/p}} |f(X_s) - \pi(f)|ds \qquad (7.37)
\end{aligned}
$$

From (7.21) we see that $\eta(T)$ tends to infinity as $T \to \infty$ and $\lim_{T\to\infty}\eta(T)/T = 1/\varrho$ almost surely. Also for every positive sequence m_T that tends to infinity as $T \to \infty$ we have that $\lim_{T\to\infty}\eta(m_T)/m_T = 1/\varrho$ almost surely. By an application of the law of iterated logarithm to (7.31) and (7.33) we obtain $R_n = n/\varrho + O(\sqrt{n\log\log n})$ *a.s.* and $N_n = n/\lambda + O(\sqrt{n\log\log n})$ *a.s.* respectively. Hence we have that $b_n = O(n)$ *a.s.*, and consequently $\eta(b_n) = O(n)$ almost surely. Note that $\eta(b_n+\varepsilon_1 n^{1/p})$, the number of regenerations until time $b_n+\varepsilon_1 n^{1/p}$ is equal to the number of generation until time b_n and the number of regenerations in the time interval $(b_n, b_n+\varepsilon_1 n^{1/p})$, i.e., $\eta(b_n+\varepsilon_1 n^{1/p}) = \eta(b_n)+\eta(b_n+\varepsilon_1 n^{1/p})-\eta(b_n)$. Since $\eta(T)$ is a renewal process it is clear that we should have

$$\eta(b_n+\varepsilon_1 n^{1/p}) - \eta(b_n) = O(\eta(\varepsilon_1 n^{1/p})) \quad a.s. \qquad (7.38)$$

Indeed, since we have that $\mathbb{E}_\nu R_1^q < \infty$, by [20, Theorem 2.4] we can construct a Brownian motion $\tilde{B}_2$ such that

$$\left|\eta(T) - \frac{T}{\mu_\eta} - \sigma_\eta \tilde{B}_2(T)\right| = o(T^{1/q}) \quad a.s., \qquad (7.39)$$

for some constants μ_η and σ_η. Hence for almost all sample paths ω there exists a $T_1(\omega)$ such that for all $T \ge T_1(\omega)$ we have that

$$\frac{1}{T^{1/q}}\left|\eta(T) - \frac{T}{\mu_\eta} - \sigma_\eta \tilde{B}_2(T)\right| < \varepsilon. \qquad (7.40)$$

Since b_n is non-decreasing and tends to infinity almost surely, it follows that for all sample paths ω there exists a $N_2(\omega)$ such that $\eta(b_n)(\omega) \ge T_1(\omega)$ for all $n \ge N_2(\omega)$ and hence

$$\frac{1}{b_n^{1/q}}\left|\eta(b_n) - \frac{b_n}{\mu_\eta} - \sigma_\eta \tilde{B}_2(b_n)\right| < \varepsilon. \qquad (7.41)$$

Since $b_n = O(n)$ almost surely, it follows that

$$\left|\eta(b_n) - \frac{b_n}{\mu_\eta} - \sigma_\eta \tilde{B}_2(b_n)\right| = o(b_n^{1/q}) = o(n^{1/q}) \quad a.s. \qquad (7.42)$$

Let $a_n := \varepsilon_1 n^{1/p}$, then by the triangle inequality, we obtain

$$\eta(b_n + a_n) - \eta(b_n) \leq \left|\eta(b_n + a_n) - (b_n + a_n)/\mu_\eta - \sigma_\eta \tilde{B}_2(\eta(b_n) + a_n)\right| \tag{7.43}$$

$$+ a_n/\mu_\eta + \left|-\eta(b_n) + b_n/\mu_\eta + \sigma_\eta \tilde{B}_2(b_n)\right| \tag{7.44}$$

$$+ \sigma_\eta \left|\tilde{B}_2(b_n + a_n) - \tilde{B}_2(b_n)\right| \tag{7.45}$$

By (7.42) the rhs of (7.43) and the second term in (7.44) are both $o(n^{1/q})$ and thus $o(n^{1/p})$. Furthermore, by [21, Theorem 2] we have that

$$\limsup_{n\to\infty} \sup_{0\leq s\leq a_n} \frac{\left|\tilde{B}_2(n+s) - \tilde{B}_2(n)\right|}{\left[a_n(\log(n/a_n) + \log\log n)\right]^{1/2}} = 1 \quad a.s. \tag{7.46}$$

Since we have $a_n = \varepsilon_1 n^{1/p}$ it follows that

$$\sup_{0\leq s\leq a_n} \left|\tilde{B}_2(n+s) - \tilde{B}_2(n)\right| = O\left(n^{1/2p}\log(n)\right) = o\left(n^{1/p}\right) \quad a.s. \tag{7.47}$$

Moreover, since $\eta(b_n) = O(n)$ $a.s.$ and almost surely non-decreasing we also have that

$$\sup_{0\leq s\leq a_n} \left|\tilde{B}_2(\eta(b_n) + s) - \tilde{B}_2(\eta(b_n))\right| = o\left(\eta(b_n)^{1/p}\right) = o\left(n^{1/p}\right) \quad a.s. \tag{7.48}$$

Hence, we have shown that

$$\eta(b_n + a_n) - \eta(b_n) \leq a_n/\mu_\eta + o\left(n^{1/p}\right)$$

almost surely. Therefore there exists a $K > 0$ such that for almost all sample paths there exits an $N_3(\omega)$ sufficiently large such that $\eta(b_n+a_n)-\eta(b_n) < Kn^{1/p}$ almost surely. Hence we have shown that the claim formulated in (7.38) indeed holds.

For technical convenience let $\tilde{a}_n$ be defined as $Kn^{1/p}$. Since $(Y_k)_{k\geq 0}$ form an i.i.d sequence with $\mathbb{E}_\nu|Y_1|^p < \infty$ we have by the Komlós-Major-Tusnády approximation that there exists a Brownian motion B_3 such that

$$\left|\sum_{k=0}^{n} Y_k - n\mu_Y - \sigma_Y B_3(n)\right| = o(n^{1/p}) \quad \text{a.s.}, \tag{7.49}$$

where μ_Y and σ_Y denote the mean and standard deviation of Y_1 respectively. It immediately follows that we also have

$$\left|\sum_{k=0}^{\eta(b_n)} Y_k - \eta(b_n)\mu_Y - \sigma_Y B_3(\eta(b_n))\right| = o(\eta(b_n^{1/p})) = o(n^{1/p}) \quad \text{a.s.} \tag{7.50}$$

By the triangle inequality, we obtain

$$\left|\sum_{k=\eta(b_n)}^{\eta(b_n)+\tilde{a}_n} Y_k\right| \leq \left|\sum_{k=0}^{\eta(b_n)+\tilde{a}_n} Y_k - (\eta(b_n) + \tilde{a}_n)\mu_Y - \sigma_Y B_3(\eta(b_n) + \tilde{a}_n))\right| \tag{7.51}$$

$$+ \tilde{a}_n \mu_Y + \left| - \sum_{k=0}^{\eta(b_n)} Y_k + \eta(b_n)\mu_Y + \sigma_Y B_3(\eta(b_n)) \right| \tag{7.52}$$

$$+ \sigma_y |B_3(\eta(b_n) + \tilde{a}_n) - B_3(\eta(b_n))| \tag{7.53}$$

$$\leq \tilde{a}_n \mu_Y + o(n^{1/p}) \quad \text{a.s.} \tag{7.54}$$

The last inequality follows, since by (7.50) both the term in (7.51) and the second term in (7.53) are $o(n^{1/p})$ almost surely. Furthermore, by (7.48) the last inequality also follows. Hence it follows that

$$\mathbb{P}_\nu \left(\limsup_{n\to\infty} \frac{1}{n^{1/p}} \left| \sum_{k=\eta(b_n)}^{\eta(b_n+a_n)} Y_k \right| \leq K\mu_Y \right) = 1. \tag{7.55}$$

Hence the first term in the upper bound (7.37) is $O(1)$ almost surely. For the second term, we see that from (7.17), it follows that

$$Y_n = \int_{R_n}^{R_{n+1}} |f(X_s) - \pi(f)| ds = O(n^{1/p}) \quad \text{a.s.} \tag{7.56}$$

Therefore

$$\int_{R_{\eta(b_n+a_n)}}^{b_n+a_n} |f(X_{R_n+s}) - \pi(f)| ds \leq \int_{R_{\eta(b_n+a_n)}}^{R_{\eta(b_n+a_n)+1}} |f(X_{R_n+s}) - \pi(f)| ds \tag{7.57}$$

$$= Y_{\eta(b_n+\varepsilon_1 n^{1/p})} \tag{7.58}$$

$$= O\left((\eta(b_n + \varepsilon_1 n^{1/p}))^{1/p} \right) \quad \text{a.s.} \tag{7.59}$$

$$= O\left((n + n^{1/p})^{1/p} \right) = O\left(n^{1/p} \right) \quad \text{a.s.} \tag{7.60}$$

Hence our claim (7.37) follows, and consequently we have also shown (7.35). Combining (7.30), (7.34), and (7.35) it follows that

$$\left| \int_0^{\gamma N(n)} [f(X_s) - \pi(f)] ds - \alpha\gamma N(n) + \alpha\varrho n - \tilde{\sigma} B_1(n) \right| = o\left(n^{\frac{1}{p}} \right) \text{ a.s.} \tag{7.61}$$

Let $(\Gamma_s)_{s\geq 0}$ be defined as $\Gamma_0 := 0$ and $\Gamma_s := N^{-1}(s)$, the right-continuous inverse of the Poisson process. Recall that N is a Poisson process with intensity $\lambda = \varrho^2/\sigma_\tau^2$. Taking $n = \Gamma_{n'}$ in (7.61) and subsequently making the substitution $n = \gamma n'$, it follows that

$$\begin{aligned} \left| \int_0^n [f(X_s) - \pi(f)] ds - \alpha n + \alpha\varrho \Gamma_{n/\gamma} - \tilde{\sigma} B_1(\Gamma_{n/\gamma}) \right| &= o\left({\Gamma_n}^{1/p} \right) \text{ a.s.} \\ &= o\left(n^{1/p} \right) \text{ a.s.,} \end{aligned} \tag{7.62}$$

where we used the fact that Γ is a non-decreasing process that tends to infinity. Moreover, since Γ_n has a Gamma distribution it follows from the Komlós-Major-Tusnády approximation [52, Theorem 1] that there exists a Brownian motion B_4 such that

$$\left|\Gamma_n - \frac{n}{\lambda} - \frac{1}{\lambda}B_4(n)\right| = O(\log n) \quad \text{a.s.} \tag{7.63}$$

Note that the Poisson process N and therefore its corresponding event time process Γ are independent of B_1. Therefore by an application of Lemma 7.9 with $n = \Gamma_n$ it follows that there exists a standard Brownian motion B_5 independent of N and Γ such that

$$\left|B_1(\Gamma_n) - \frac{1}{\sqrt{\lambda}}B_5(n)\right| = O(\log n) \quad \text{a.s.} \tag{7.64}$$

Applying the obtained approximations given in (7.63) and (7.64) to (7.62) it follows that

$$\left|\int_0^n f(X_s) - \pi(f)ds - \left(\frac{\tilde{\sigma}}{\sqrt{\lambda\gamma}}B_5(n) - \frac{\alpha\varrho}{\lambda\sqrt{\gamma}}B_4(n)\right)\right| = o\left(n^{1/p}\right) \quad \text{a.s.} \tag{7.65}$$

Note that since B_4 and B_5 are independent we have that

$$W_n = \frac{1}{\sigma_f}\left(\frac{\tilde{\sigma}}{\sqrt{\lambda\gamma}}B_5(n) - \frac{\alpha\varrho}{\lambda\sqrt{\gamma}}B_4(n)\right) \tag{7.66}$$

is a standard Brownian motion since

$$\frac{\tilde{\sigma}^2}{\gamma\lambda} + \frac{\alpha^2\varrho^2}{\gamma\lambda^2} = \frac{\mathbb{E}_\nu\xi_1^2}{\varrho} = \sigma_f^2. \tag{7.67}$$

Furthermore, by definition of big O, there exists an almost surely finite random variable C such that for almost all sample paths ω we have that for all $n \geq N_0 \equiv N_0(\omega)$ we have that

$$\frac{1}{n^{1/p}}\left|\int_0^n f(X_s(\omega))ds - T\pi(f) - \sigma_f^2 W_n(\omega)\right| < C(\omega). \tag{7.68}$$

It immediately follows that (7.68) also holds for T sufficiently large and hence carries over for $T \to \infty$. By the same argument given in the proof of Theorem 4.1, the strong invariance principle holds for every initial distribution. □

7.4.2. Proof of Theorem 4.7

Proof. From [9, Proposition 2.8] we see that the Zig-Zag process with a stationary distribution of product form $\pi(x) = \prod_{i=1}^d \pi_i(x_i)$ can be decomposed into d independent Zig-Zag processes, each with stationary distribution π_i. Since we have that

$$\left\|\int_0^T f(X_t)\,dt - T\pi(f) - \Sigma_f^{1/2}W(T)\right\|$$

$$\leq \sqrt{d} \max_i \left| \int_0^T f_i(X_t^i)\,dt - T\pi_i(f_i) - \sigma_{f_i} W^i(T) \right|, \tag{7.69}$$

the theorem follows if we can show that a strong invariance principle holds for every component on the same probability space. Firstly, assume that the initial distribution of Z is π.

In order to obtain a Brownian approximation for every coordinate we will use a regenerative argument along the lines of Theorem 4.6. For every component $i = 1, \dots, d$ we define the following: x_0^i the smallest local maximum of the density π_i, i.e., $x_0^i = \min\{x : \pi_i'(x) = 0\}$ and corresponding set set $A_i = [x_0^i - M, x_0^i] \times \{+1\}$, and the sequences of stopping times $\{R_n^i\}_{n\in\mathbb{N}}$ as follows

$$R_0^i = \inf\{t \geq 0 : (X_t^i, V_t^i) = (x_0^i, 1)\},$$

and

$$R_n^i = \inf\{t \geq R_{n-1} : (X_t^i, V_t^i) = (x_0^i, 1)\}.$$

Furthermore, we also introduce for every coordinate i the sequence $\{\xi_n^i\}$ defined as

$$\xi_n^i := \int_{R_{n-1}^i}^{R_n^i} \{f(X_s) - \pi(f)\}\,ds, \quad n \geq 1.$$

Note that for all components $\{\xi_n^i\}_n$ is i.i.d under $\mathbb{P}_{\nu_i}$, with ν_i a Dirac measure at the point $x_0^i \times \{+1\}$. We can follow the argument of Theorem 4.4 to obtain that

$$\mathbb{E}_{\nu_i}\left[(R_1^i)^{\beta-1}\right] < \infty \quad \text{for } i = 1, \dots, d.$$

Moreover, for all measurable $f : E \to \mathbb{R}$ with $\pi(|f|^{p+\varepsilon}) < \infty$ where $p \geq 1$ and $\beta > 2 + p/\varepsilon$, we have that

$$\mathbb{E}_{\nu_i}\left| \int_{R_0^i}^{R_1^i} f_i(X_s^i) - \pi(f)ds \right|^p < \infty \quad \text{for } i = 1, \dots, d.$$

Note that for every coordinate i we have that

$$\begin{aligned}\left| \int_0^T f_i(X_t^i)\,dt - T\pi_i(f_i) - \sigma_{f_i} W^i(T) \right| &\leq \left| \int_0^{R_1^i} f_i(X_t^i) - \pi_i(f_i)dt \right| \\ &\quad + \left| \int_{R_1^i}^T f_i(X_t^i) - \pi_i(f_i)dt - \sigma_{f_i} W^i(T) \right|.\end{aligned} \tag{7.70}$$

By assuming that the process starts at its stationary distribution, it follows by the argument in the proof of Theorem 4.4 that $\left|\int_0^{R_1^i} f_i(X_t^i) - \pi_i(f_i)dt\right|$ is almost surely finite and hence asymptotically negligible.

Define $(\tau_k^i)_{k\in\mathbb{N}}$ as $\tau_k^i = R_k^i - R_{k-1}^i$ and let ϱ_i and $\sigma_{\varrho_i}^2$ denote the mean and variance respectively. The sequence of random vectors (ξ_k^i, τ_k^i) are independent and identically distributed. If we choose $\alpha_i = \operatorname{Cov}_\nu(\xi_1^i, \tau_1^i)/\operatorname{Var}_\nu(\tau_1^i)$, then it immediately follows that $\xi_k^i - \alpha_i(\tau_k^i - \varrho_i)$ and τ_k^i are uncorrelated. By applying the multivariate Komlós-Major-Tusnády approximation given in [35, Theorem 1] and [20, Theorem 2.1] to the sequence of random vectors

$$z_k = (z_k^1, \dots, z_k^d)^T = ((\xi_k^1 - \alpha_1(\tau_k^1 - \varrho_1), \tau_k^1), \dots, (\xi_k^d - \alpha_d(\tau_k^d - \varrho_d), \tau_k^d))^T,$$

it follows that there exists a $2d$-dimensional Brownian motion such that

$$\left| \sum_{k=1}^n z_k - \mathbb{E}_\nu z_1 - \tilde{\Sigma}_z B_n \right| = o\left(n^{1/p}\right) \text{ a.s.,} \tag{7.71}$$

where $\tilde{\Sigma}_z = \operatorname{diag}(\operatorname{Var}_\nu(z_1), \dots, \operatorname{Var}_\nu(z_k))$. All components of z_k are independent and therefore also the corresponding components of the Brownian motion are independent. Note that we have that for every component z_k^i of z_k we have that there exists two independent Brownian motions B_1 and B_2 such that

$$\left| \sum_{k=1}^n \xi_k^i - \alpha_i \left(\sum_{k=1}^n \tau_k^i - \varrho_i \right) - \tilde{\sigma}_i B_{i1} \right| = o\left(n^{1/p}\right) \text{ a.s.} \tag{7.72}$$

$$\left| R_n^i - n\varrho_i - \sigma_{\tau_i} B_{i2}(n) \right| = o\left(n^{1/p}\right) \text{ a.s.} \tag{7.73}$$

Note that in (7.72) we have that $\mathbb{E}_\nu \xi_1^i = 0$ by Theorem 3.4 and that $\tilde{\sigma}_i = \operatorname{Var}_\nu(\xi_1^i - \alpha_i(\tau_1^i - \varrho_i))$ a.s. By following the argument of the proof of Theorem 4.6 for every component, we see that

$$\left| \int_{R_1^i}^n f_i(X_t^i) - \pi_i(f_i) ds - \sigma_{f_i} W_n^i \right| = o\left(n^{1/p}\right) \text{ a.s.} \quad \text{for } i = 1, \dots, d. \tag{7.74}$$

By combining (7.69), (7.70) and (7.74) the claim follows. By the argument given in the proof of Theorem 4.1, the strong invariance principle holds for every initial distribution. □

7.5. Proof of Theorem 6.2

Proof. Firstly, by Proposition 4.3 there exist two standard Brownian motions W_1 and W_2 such that

$$\left| \int_0^T f(X_s) ds - W_1(\sigma_T^2) - W_2(\tau_T^2) \right| = O(\psi_T) \text{ a.s.,}$$

where $\{\sigma_T^2\}$ and $\{\tau_T^2\}$ are non-decreasing sequences with

$$\sigma_T^2 = \frac{\sigma_\xi^2}{\varrho} T + O(T/\log T) \text{ and } \tau_T^2 = O(T/\log T).$$

as $T \to \infty$, where σ_ξ^2 and ϱ are defined in Theorem 4.3. An application of our strong invariance principle gives the following

$$\begin{aligned}
&\limsup_{T\to\infty} \max_{0\le t\le T-a_T} \max_{0\le u\le a_T} \beta_T \left| \int_0^{t+u} f(X_u)du - \int_0^t f(X_u)du \right| \\
&\le \limsup_{T\to\infty} \max_{0\le t\le T-a_T} \max_{0\le u\le a_T} \beta_T \big| W_1(\sigma_{t+u}^2) - W_1(\sigma_t^2) \big| \\
&\quad + \limsup_{T\to\infty} \max_{0\le t\le T-a_T} \max_{0\le u\le a_T} \beta_T \big| W_2(\tau_{t+u}^2) - W_2(\tau_t^2) \big| \\
&\quad + \beta_T O(\psi_T) \\
&=: A_1 + A_2 + A_3.
\end{aligned}$$

Since $\beta_T\psi_T = o(1)$, it immediately follows that $\limsup_T A_3 = 0$ almost surely. In order to use the arguments of [5, Theorem 4] for the terms A_1 and A_2, we require the following properties of the sequence σ_T^2; for any $\varepsilon > 0$ there exists some T_0 such that for all $T \ge T_0$

$$\sigma_T^2 \le \left(\frac{\sigma_\xi^2}{\varrho} + \varepsilon \right) T \quad \text{and} \quad \sup_{u\ge 0} \{ \sigma_{u+a_T}^2 - \sigma_u^2 \} \le \left(\frac{\sigma_\xi^2}{\varrho} + \varepsilon \right) a_T. \tag{7.75}$$

Since $\sigma_T^2 = \frac{\sigma_\xi^2}{\mu} T + O(T/\log(T))$, the first required property described in (7.75) follows directly. From the proof of Theorem 4.5 we have that

$$\sigma_T^2 = s_n^2/\varrho, \quad \text{for } T \in [n, n+1). \tag{7.76}$$

Note that (7.76) is equivalent to $\sigma_T^2 = s_{\lfloor T\rfloor}^2/\varrho$ and therefore $\limsup_{u\to\infty}(\sigma_{\lfloor u\rfloor+1}^2 - s_{\lfloor u\rfloor}^2) \le \limsup_{k\to\infty}(s_{k+1}^2 - s_k^2)/\varrho = \sigma_\xi^2/\varrho$, where the last equality follows since by Theorem 7.5 we have that $\limsup_k(s_{k+1}^2 - s_k^2) = \sigma_\xi^2$. Since a_T tends to infinity, we have for T and U_0 sufficiently large that

$$\begin{aligned}
\sup_{u>U_0} \{\sigma_{u+a_T}^2 - \sigma_u^2\} &= \sup_{u>U_0} \{\sigma_{\lfloor u+a_T\rfloor}^2 - \sigma_u^2\} \\
&\le \sup_{u>U_0} \{\sigma_{\lfloor u+a_T\rfloor}^2 - \sigma_{\lfloor u\rfloor}^2\} \\
&\le \limsup_{u\to\infty} \sum_{j=1}^{\lfloor a_T\rfloor} (\sigma_{\lfloor u\rfloor+j}^2 - \sigma_{\lfloor u\rfloor+j-1}^2) \\
&\le 1/\varrho \sum_{j=1}^{\lfloor a_T\rfloor} \limsup_{k\to\infty} (s_{\lfloor u\rfloor+j}^2 - s_{\lfloor u\rfloor+j-1}^2) \\
&\le (\sigma_\xi^2/\varrho + \varepsilon)\lfloor a_T\rfloor \le (\sigma_\xi^2/\varrho + \varepsilon)a_T,
\end{aligned} \tag{7.77}$$

where the first equality follows from (7.76), the first inequality due to the fact that $(\sigma_u)_{u\ge 0}$ is a non-decreasing sequence. Note that for all $U_0 > 0$ we have that

$$\sup_u \{\sigma_{u+a_T}^2 - \sigma_u^2\} = \max\left\{ \sup_{u\le U_0} \{\sigma_{u+a_T}^2 - \sigma_u^2\}, \sup_{u>U_0} \{\sigma_{u+a_T}^2 - \sigma_k^2\} \right\} \tag{7.78}$$

Since $(\sigma_n^2)_{n\geq 0}$ is a non-decreasing sequence and a_T tends to infinity we have that for sufficiently large T that

$$\sup_{u\leq U_0}\{\sigma^2_{u+a_T}-\sigma^2_u\}\leq \sigma^2_{U_0+a_T}\leq (\sigma_\xi^2/\varrho+\varepsilon)a_T. \tag{7.79}$$

Combining (7.77) and (7.79) gives (7.78). Consequently, we have also shown that the required properties given in (7.75) hold. Hence for $T\geq T_0$ we obtain

$$\begin{aligned}
&\max_{0\leq t\leq T-a_T}\max_{0\leq u\leq a_T}\beta_T\big|W_1(\sigma^2_{t+u})-W_1(\sigma^2_t)\big|\\
&\leq \sup_{0\leq t\leq \sigma^2_{T-a_T}}\ \sup_{0\leq u\leq (\sigma_\xi^2/\varrho+\varepsilon)a_T}\beta_T|W_1(t+u)-W_1(t)|\\
&\leq \sup_{0\leq t\leq (\sigma_\xi^2/\varrho+\varepsilon)(T-a_T)}\ \sup_{0\leq u\leq (\sigma_\xi^2/\varrho+\varepsilon)a_T}\beta_T|W_1(t+u)-W_1(t)|\\
&= \sup_{0\leq t\leq \tilde T_\varepsilon-\tilde a_{T,\varepsilon}}\ \sup_{0\leq u\leq \tilde a_{T,\varepsilon}}\beta_T|W_1(t+u)-W_1(t)|,
\end{aligned}$$

where $\tilde T_\varepsilon$ and $\tilde a_{T,\varepsilon}$ are defined as $(\sigma_\xi^2/\varrho+\varepsilon)T$ and $(\sigma_\xi^2/\varrho+\varepsilon)a_T$ respectively. Introduce

$$\tilde\beta_{T,\varepsilon}:=\left(2\tilde a_{T,\varepsilon}\left[\log\frac{\tilde T_\varepsilon}{\tilde a_{T,\varepsilon}}+\log\log\tilde T_\varepsilon\right]\right)^{-1/2},$$

then by Theorem 6.1 we have that

$$\limsup_{T\to\infty}\ \sup_{0\leq t\leq \tilde T_\varepsilon-a_{T,\varepsilon}}\ \sup_{0\leq u\leq a_{T,\varepsilon}}\tilde\beta_{T,\varepsilon}|W(t+u)-W_t|\leq \sigma_\xi^2/\varrho\quad \text{a.s.}$$

Similarly, it can be shown that $\limsup A_2=0$ almost surely, which completes the proof. ☐

Acknowledgments

The authors thank the associate editor and the referees for their constructive comments and suggestions which helped to improve the manuscript. This work is part of the research programme 'Zigzagging through computational barriers' with project number 016.Vidi.189.043, which is financed by the Dutch Research Council (NWO).

References

[1] Adler, R., Feldman, R. and Taqqu, M. (1998). *A practical guide to heavy tails: statistical techniques and applications*. Springer Science & Business Media. MR1652283

[2] Andrieu, C., Dobson, P. and Wang, A. Q. (2021). Subgeometric hypocoercivity for piecewise-deterministic Markov process Monte Carlo methods. *Electronic Journal of Probability* **26** 1–26. MR4269208

[3] Athreya, K. B. and Ney, P. (1978). A new approach to the limit theory of recurrent Markov chains. *Transactions of the American Mathematical Society* **245** 493–501. MR0511425

[4] Bednorz, W. and Łatuszyński, K. (2007). A few remarks on "Fixed-width output analysis for Markov chain Monte Carlo" by Jones et al. *Journal of the American Statistical Association* **102** 1485–1486. MR2412582

[5] Berkes, I., Hörmann, S., Schauer, J. et al. (2011). Split invariance principles for stationary processes. *The Annals of Probability* **39** 2441–2473. MR2932673

[6] Berkes, I., Liu, W. and Wu, W. B. (2014). Komlós–Major–Tusnády approximation under dependence. *The Annals of Probability* **42** 794–817. MR3178474

[7] Bhattacharya, R. (1978). Criteria for recurrence and existence of invariant measures for multidimensional diffusions. *The Annals of Probability* 541–553. MR0494525

[8] Bierkens, J. and Duncan, A. (2017). Limit theorems for the zig-zag process. *Advances in Applied Probability* **49** 791–825. MR3694318

[9] Bierkens, J., Fearnhead, P., Roberts, G. et al. (2019). The zig-zag process and super-efficient sampling for Bayesian analysis of big data. *The Annals of Statistics* **47** 1288–1320. MR3911113

[10] Bierkens, J. and Roberts, G. (2017). A piecewise deterministic scaling limit of lifted Metropolis–Hastings in the Curie-Weiss model. *The Annals of Applied Probability* **27** 846–882. MR3655855

[11] Bierkens, J., Roberts, G. O. and Zitt, P.-A. (2019). Ergodicity of the zigzag process. *Ann. Appl. Probab.* **29** 2266–2301. `https://doi.org/10.1214/18-AAP1453`. MR3983339

[12] Bouchard-Côté, A., Vollmer, S. J. and Doucet, A. (2018). The bouncy particle sampler: A nonreversible rejection-free Markov chain Monte Carlo method. *Journal of the American Statistical Association* **113** 855–867. MR3832232

[13] Bradley, R. C. (2005). Basic properties of strong mixing conditions. A survey and some open questions. *Probability Surveys* **2** 107–144. MR2178042

[14] Brockwell, A. E. and Kadane, J. B. (2005). Identification of regeneration times in MCMC simulation, with application to adaptive schemes. *Journal of Computational and Graphical Statistics* **14** 436–458. MR2161623

[15] Cattiaux, P., Chafai, D. and Guillin, A. (2011). Central limit theorems for additive functionals of ergodic Markov diffusions processes. *arXiv preprint arXiv:1104.2198*. MR3069369

[16] Chakraborty, S., Bhattacharya, S. K. and Khare, K. (2019). Estimating accuracy of the MCMC variance estimator: a central limit theorem for batch means estimators. *arXiv preprint arXiv:1911.00915*. MR4363701

[17] Chien, C., Goldsman, D. and Melamed, B. (1997). Large-sample results for batch means. *Management Science* **43** 1288–1295.

[18] Chien, C.-H. (1988). Small-sample theory for steady state confidence

intervals. In *Proceedings of the 20th Conference on Winter Simulation* 408–413.

[19] Csáki, E. and Csörgő, M. (1995). On additive functionals of Markov chains. *Journal of Theoretical Probability* **8** 905–919. MR1353559

[20] Csörgö, M. and Horváth, L. (1993). *Weighted approximations in probability and statistics*. J. Wiley & Sons. MR1215046

[21] Csörgö, M. and Révész, P. (1979). How big are the increments of a Wiener process? *The Annals of Probability* 731–737. MR0537218

[22] Csörgö, M. and Révész, P. (2014). *Strong approximations in probability and statistics*. Academic Press. MR0666546

[23] Csörgö, S. and Hall, P. (1984). The Komlós-Major-Tusnády approximations and their applications. *Australian Journal of Statistics* **26** 189–218. MR0766619

[24] Cuny, C., Dedecker, J. and Merlevède, F. (2018). On the Komlós, Major and Tusnády strong approximation for some classes of random iterates. *Stochastic Processes and their Applications* **128** 1347–1385. MR3769665

[25] Damerdji, H. (1991). Strong consistency and other properties of the spectral variance estimator. *Management Science* **37** 1424–1440.

[26] Damerdji, H. (1994). Strong consistency of the variance estimator in steady-state simulation output analysis. *Mathematics of Operations Research* **19** 494–512. MR1290511

[27] Damerdji, H. (1995). Mean-square consistency of the variance estimator in steady-state simulation output analysis. *Operations Research* **43** 282–291. MR1327416

[28] Davydov, Y. A. (1968). Convergence of distributions generated by stationary stochastic processes. *Theory of Probability & Its Applications* **13** 691–696. MR0243586

[29] Deheuvels, P. (2000). Uniform limit laws for kernel density estimators on possibly unbounded intervals. In *Recent advances in reliability theory* 477–492. Springer. MR1783500

[30] Deligiannidis, G., Bouchard-Côté, A., Doucet, A. et al. (2019). Exponential ergodicity of the bouncy particle sampler. *The Annals of Statistics* **47** 1268–1287. MR3911112

[31] Douc, R., Moulines, E., Priouret, P. and Soulier, P. (2018). *Markov chains*. Springer. MR3889011

[32] Down, D., Meyn, S. P. and Tweedie, R. L. (1995). Exponential and uniform ergodicity of Markov processes. *The Annals of Probability* 1671–1691. MR1379163

[33] Duncan, A. B., Lelievre, T. and Pavliotis, G. (2016). Variance reduction using nonreversible Langevin samplers. *Journal of Statistical Physics* **163** 457–491. MR3483241

[34] Durmus, A., Guillin, A., Monmarché, P. et al. (2020). Geometric ergodicity of the bouncy particle sampler. *Annals of Applied Probability* **30** 2069–2098. MR4149523

[35] Einmahl (1989). Extensions of results of Komlós, Major, and Tusnády

to the multivariate case. *Journal of Multivariate Analysis* **28** 20–68. MR0996984

[36] El-Nouty, C. (1999). On the large increments of fractional Brownian motion. *Statistics & Probability Letters* **41** 169–178. MR1665268

[37] Fearnhead, P., Bierkens, J., Pollock, M., Roberts, G. O. et al. (2018). Piecewise deterministic Markov processes for continuous-time Monte Carlo. *Statistical Science* **33** 386–412. MR3843382

[38] Flegal, J. M. and Jones, G. L. (2010). Batch means and spectral variance estimators in Markov chain Monte Carlo. *The Annals of Statistics* **38** 1034–1070. MR2604704

[39] Fort, G. and Roberts, G. O. (2005). Subgeometric ergodicity of strong Markov processes. *The Annals of Applied Probability* **15** 1565–1589. MR2134115

[40] Gilks, W. R., Roberts, G. O. and Sahu, S. K. (1998). Adaptive Markov chain Monte Carlo through regeneration. *Journal of the American Statistical Association* **93** 1045–1054. MR1649199

[41] Glynn, P. W. and Whitt, W. (1991). Estimating the asymptotic variance with batch means. *Operations Research Letters* **10** 431–435. MR1141337

[42] Glynn, P. W. and Whitt, W. (1992). The asymptotic validity of sequential stopping rules for stochastic simulations. *The Annals of Applied Probability* **2** 180–198. MR1143399

[43] Goldsman, D., Meketon, M. and Schruben, L. (1990). Properties of standardized time series weighted area variance estimators. *Management Science* **36** 602–612. MR1053642

[44] Gong, L. and Flegal, J. M. (2016). A practical sequential stopping rule for high-dimensional Markov chain Monte Carlo. *Journal of Computational and Graphical Statistics* **25** 684–700. MR3533633

[45] Heunis, A. J. (2003). Strong invariance principle for singular diffusions. *Stochastic Processes and Their Applications* **104** 57–80. MR1956472

[46] Hobert, J. P., Jones, G. L., Presnell, B. and Rosenthal, J. S. (2002). On the applicability of regenerative simulation in Markov chain Monte Carlo. *Biometrika* **89** 731–743. MR1946508

[47] Höpfner, R. and Löcherbach, E. (2003). *Limit theorems for null recurrent Markov processes*. American Mathematical Soc. MR1949295

[48] Hwang, C.-R., Hwang-Ma, S.-Y. and Sheu, S.-J. (1993). Accelerating gaussian diffusions. *The Annals of Applied Probability* 897–913. MR1233633

[49] Jones, G. L., Haran, M., Caffo, B. S. and Neath, R. (2006). Fixed-width output analysis for Markov chain Monte Carlo. *Journal of the American Statistical Association* **101** 1537–1547. MR2279478

[50] Kallenberg, O. (1997). *Foundations of modern probability* **2**. Springer. MR1464694

[51] Komlós, J., Major, P. and Tusnády, G. (1976). An approximation of partial sums of independent RV's, and the sample DF. II. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **34** 33–58. MR0402883

[52] Komlos, M. (1975). Tusnady (1975) An approximation of partial sums of RV's, and the sample DF, I, Z. *Wahrscheinlichkeitstheorie und Verw. Gebiete* **32** 111–131. MR0375412
[53] Kuelbs, J. and Philipp, W. (1980). Almost sure invariance principles for partial sums of mixing B-valued random variables. *The Annals of Probability* 1003–1036. MR0602377
[54] Lazi, P. and Sandri, N. (2021). On sub-geometric ergodicity of diffusion processes. *Bernoulli* **27** 348–380. MR4177373
[55] Lelievre, T., Nier, F. and Pavliotis, G. A. (2013). Optimal non-reversible linear drift for the convergence to equilibrium of a diffusion. *Journal of Statistical Physics* **152** 237–274. MR3082649
[56] Li, W. V. (1992). Limit theorems for the square integral of Brownian motion and its increments. *Stochastic Processes and Their Applications* **41** 223–239. MR1164176
[57] Liu, Y., Vats, D. and Flegal, J. M. (2021). Batch size selection for variance estimators in MCMC. *Methodology and Computing in Applied Probability* 1–29. MR4379481
[58] Löcherbach, E. and Loukianova, D. (2008). On Nummelin splitting for continuous time Harris recurrent Markov processes and application to kernel estimation for multi-dimensional diffusions. *Stochastic Processes and their Applications* **118** 1301–1321. MR2427041
[59] Löcherbach, E. and Loukianova, D. (2009). The law of iterated logarithm for additive functionals and martingale additive functionals of Harris recurrent Markov processes. *Stochastic Processes and Their Applications* **119** 2312–2335. MR2531093
[60] Major, P. (1979). An improvement of Strassen's invariance principle. *The Annals of Probability* **7** 55–61. MR0515812
[61] Merlevède, F., Rio, E. et al. (2015). Strong approximation for additive functionals of geometrically ergodic Markov chains. *Electronic Journal of Probability* **20**. MR3317156
[62] Meyn, S. P. and Tweedie, R. L. (1993). Stability of Markovian processes II: Continuous-time processes and sampled chains. *Advances in Applied Probability* 487–517. MR1234294
[63] Meyn, S. P. and Tweedie, R. L. (2012). *Markov chains and stochastic stability*. Springer Science & Business Media. MR2509253
[64] Mihalache, S. (2012). Strong approximations and sequential change-point analysis for diffusion processes. *Statistics & Probability Letters* **82** 464–472. MR2887460
[65] Monmarché, P. (2014). Piecewise deterministic simulated annealing. *arXiv preprint arXiv:1410.1656*. MR3487077
[66] Mykland, P., Tierney, L. and Yu, B. (1995). Regeneration in Markov chain samplers. *Journal of the American Statistical Association* **90** 233–241. MR1325131
[67] Nummelin, E. (1978). A splitting technique for Harris recurrent Markov chains. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **43** 309–318. MR0501353

[68] Nummelin, E. (2004). *General irreducible Markov chains and non-negative operators*. Cambridge University Press. MR0776608

[69] Nummelin, E. and Tuominen, P. (1982). Geometric ergodicity of Harris recurrent Marcov chains with applications to renewal theory. *Stochastic Processes and Their Applications* **12** 187–202. MR0651903

[70] Nummelin, E. and Tuominen, P. (1983). The rate of convergence in Orey's theorem for Harris recurrent Markov chains with applications to renewal theory. *Stochastic Processes and Their Applications* **15** 295–311. MR0711187

[71] Ortega, J. (1984). On the size of the increments of nonstationary Gaussian processes. *Stochastic Processes and Their Applications* **18** 47–56. MR0757346

[72] Parzen, E. (1979). Nonparametric statistical data modeling. *Journal of the American Statistical Association* **74** 105–121. MR0529528

[73] Piterbarg, V. I. and Rodionov, I. V. (2020). High excursions of Bessel and related random processes. *Stochastic Processes and their Applications* **130** 4859–4872. MR4108474

[74] Révész, P. (1982). On the increments of Wiener and related processes. *The Annals of Probability* 613–622. MR0659532

[75] Revuz, D. (2008). *Markov chains*. Elsevier. MR0415773

[76] Rey-Bellet, L. and Spiliopoulos, K. (2015). Irreversible Langevin samplers and variance reduction: a large deviations approach. *Nonlinearity* **28** 2081. MR3366637

[77] Rio, E. (1993). Covariance inequalities for strongly mixing processes. In *Annales de l'IHP Probabilités et Statistiques* **29** 587–597. MR1251142

[78] Roberts, G. O. and Rosenthal, J. S. (2001). Markov chains and de-initializing processes. *Scandinavian Journal of Statistics* **28** 489–504. MR1858413

[79] Rogers, L. C. G. and Williams, D. (2000). *Diffusions, Markov processes and martingales: Volume 2, Itô calculus* **2**. Cambridge University Press. MR1780932

[80] Sherman, M. and Goldsman, D. (2002). Large-sample normality of the batch-means variance estimator. *Operations Research Letters* **30** 319–326. MR1932879

[81] Shorack, G. R. and Wellner, J. A. (2009). *Empirical processes with applications to statistics*. SIAM. MR3396731

[82] Sigman, K. (1990). One-dependent regenerative processes and queues in continuous time. *Mathematics of Operations Research* **15** 175–189. MR1038240

[83] Song, W. T. and Schmeiser, B. W. (1995). Optimal mean-squared-error batch sizes. *Management Science* **41** 110–123.

[84] Stramer, O. and Tweedie, R. (1997). Existence and stability of weak solutions to stochastic differential equations with non-smooth coefficients. *Statistica Sinica* 577–593. MR1467449

[85] Stramer, O. and Tweedie, R. (1999). Langevin-type models I: Diffusions with given stationary distributions and their discretizations. *Methodology*

and Computing in Applied Probability **1** 283–306. MR1730651

[86] Stroock, D. W. and Varadhan, S. S. (1997). *Multidimensional diffusion processes* **233**. Springer Science & Business Media. MR0532498

[87] Tweedie, S. (1993). Generalized resolvents and Harris recurrence. *Doeblin and Modern Probability* **149**.

[88] Vats, D., Flegal, J. M. and Jones, G. L. (2019). Multivariate output analysis for Markov chain Monte Carlo. *Biometrika* **106** 321–337. MR3949306

[89] Vats, D., Flegal, J. M., Jones, G. L. et al. (2018). Strong consistency of multivariate spectral variance estimators in Markov chain Monte Carlo. *Bernoulli* **24** 1860–1909. MR3757517

[90] Williams, D. (2006). In *Stochastic integrals: proceedings of the LMS Durham Symposium*, July 7–17, 1980 **851**. Springer. MR0620983