Uncertainty Quantification for High-Dimensional Markov Chain Monte Carlo

Pengel, A.L.

DOI
10.4233/uuid:905e2bb6-034e-40a3-ab16-e7e677c5a976

Publication date
2024

Document Version
Final published version

Citation (APA)
Pengel, A. L. (2024). *Uncertainty Quantification for High-Dimensional Markov Chain Monte Carlo*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:905e2bb6-034e-40a3-ab16-e7e677c5a976

Uncertainty Quantification for High-Dimensional Markov Chain Monte Carlo

Pengel, A.L.

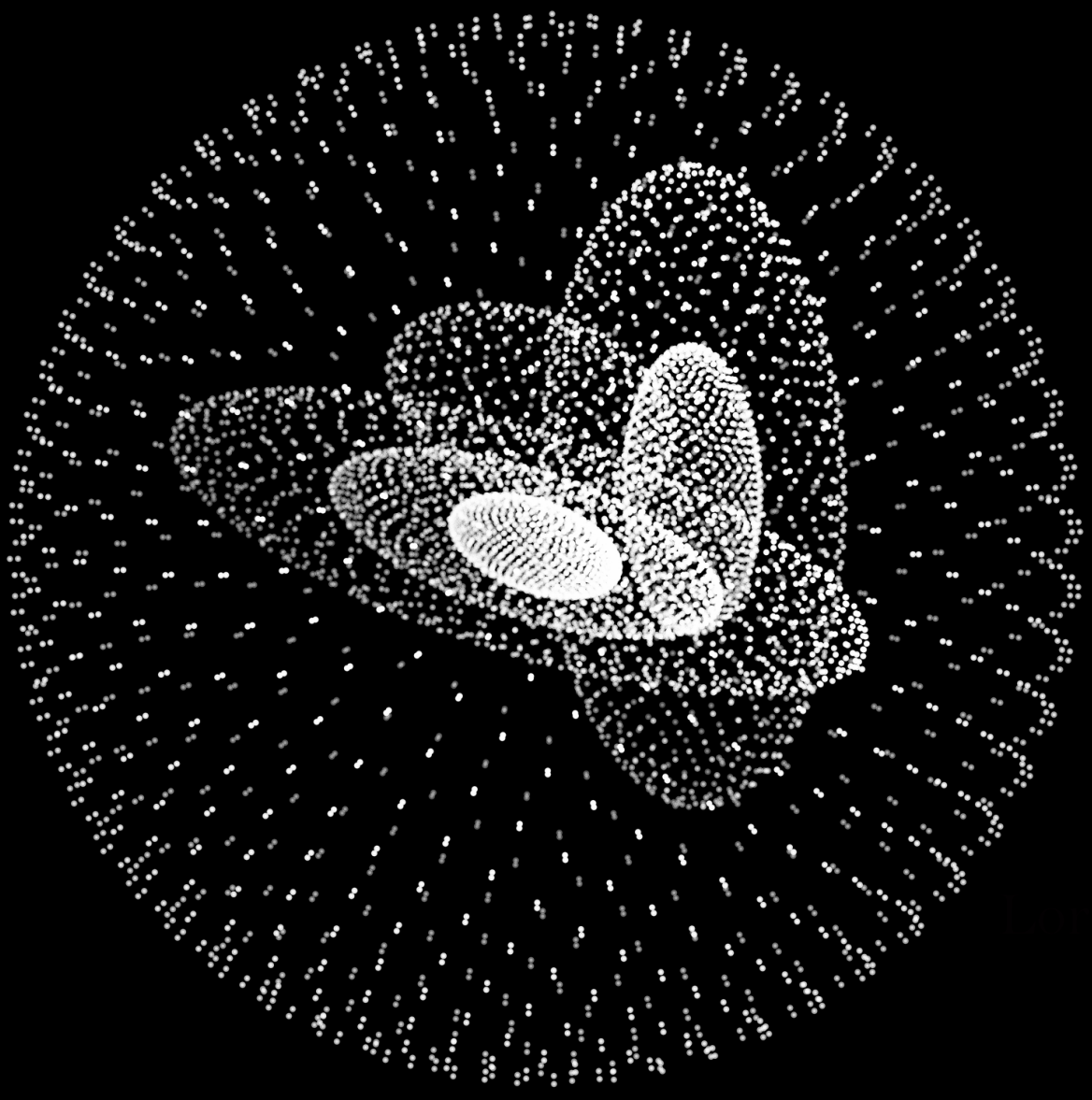
DOI
10.4233/uuid:905e2bb6-034e-40a3-ab16-e7e677c5a976

Publication date
2024

Document Version
Final published version

Citation (APA)
Pengel, A. L. (2024). *Uncertainty Quantification for High-Dimensional Markov Chain Monte Carlo*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:905e2bb6-034e-40a3-ab16-e7e677c5a976

Uncertainty Quantification for High-Dimensional Markov Chain Monte Carlo



Ardjen Laurence Pengel

Uncertainty Quantification for High-Dimensional Markov Chain Monte Carlo

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology
by the authority of the Rector Magnificus,
Prof.dr.ir. T.H.J.J. van der Hagen, chair of the Board for
Doctorates to be defended publicly on
Monday, the 9$^{\text{th}}$ of December 2024 at 3:00 p.m.

by

Ardjen Laurence PENGEL,
Master of Science in Mathematics,
Master of Science in Econometrics,
University of Amsterdam, Netherlands,

born in Ghent, Belgium.

This dissertation has been approved by the promotors.

Composition of the doctoral committee:

Rector Magnificus	*chairperson*
Dr. ir. G.N.J.C. Bierkens	Delft University of Technology, *promotor.*
Prof. dr. ir. G. Jongbloed	Delft University of Technology, *promotor.*

Independent members:

Dr. R. Altmeyer	Imperial College London.
Dr. K. Łatuszyński	University of Warwick.
Prof. dr. ir. F. H. v. d. Meulen	Vrije Universiteit Amsterdam.
Prof. dr. F. H. J. Redig	Delft University of Technology.
Prof. dr. A.W. van der Vaart	Delft University of Technology, *reserve member.*

Other member:

Dr. J. Yang	University of Copenhagen, *external collaborator.*

The research in this dissertation was funded by NWO as a part of the research programme 'Zigzagging through computational barriers' with project number 016.Vidi.189.043

***Cover page**: the figure on the cover is an illustration of a sequence of converging confidence ellipsoids*

"I know of scarcely anything so apt to impress the imagination as the wonderful form of cosmic order expressed by the 'Law of Frequency of Error.' The Law would have been personified by the Greeks and deified, if they had known of it." - Francis Galton

Contents

List of Symbols

i.i.d.	independent and identically distributed
$\xrightarrow{\text{w}}$	convergence in distribution; weak convergence
$\xrightarrow{\text{p}}$	convergence in probability
$\xrightarrow{\text{a.s.}}$	convergence almost surely
$\sim$	has distribution
$\mathcal{B}(E)$	Borel σ-algebra on E
$\mathcal{L}(X)$	distribution of X
$\mathcal{N}_d(m, \Sigma)$	d-dimensional Normal with mean $m \in \mathbb{R}^d$ and covariance matrix $\Sigma \in \mathbb{R}^{d \times d}$
$\mathcal{N}(m, \sigma^2)$	univariate Normal distribution with mean $m \in \mathbb{R}$ and variance $\sigma^2 \in \mathbb{R}^+$
$\text{Uniform}(A)$	Uniform random variable on the set A
$\text{Exp}(\lambda)$	Exponential random variable with rate λ
$\mathbb{E}_\pi(\cdot)$	Expectation under measure π
$\text{Var}_\pi(\cdot)$	Variance-covariance matrix under measure π
$\text{Cov}_\pi(\cdot, \cdot)$	Cross-covariance matrix under measure π
$a \wedge b$	$\min(a, b)$ for $a, b \in \mathbb{R}$
$a \vee b$	$\max(a, b)$ for $a, b \in \mathbb{R}$
$a \lesssim b$	inequality up to a universal constant
$\log^*(x)$	$\log(x) \vee 1$ for $x > 0$
$\text{tr}(A)$	trace of matrix A
$\det(A)$	determinant of matrix A
$\lvert A \rvert_*$	Spectral norm of matrix A
$\lvert A \rvert$	Frobenius norm of matrix A
$\text{Vol}(\cdot)$	Volume element

Chapter 1

Introduction

The main idea behind the Monte Carlo method is that by simulating a process many times, the average outcome of the simulations will correspond with the expected behaviour of our process of interest. This methodology is especially useful in settings where we can generate simulations but are unable to perform direct calculations. These methods were developed by physicists Nicholas Metropolis and Stanislaw Ulam in the 1940s during their work on the Manhattan Project for the development of the first nuclear weapons. Through simulation of neutron diffusion, they gained insights into nuclear chain reactions that could not be obtained through exact calculations.

For many applications, it is not even possible to obtain exact simulations of the object of interest. To address this problem, the first Markov Chain Monte Carlo (MCMC) algorithms were introduced by Metropolis et al. [104]. The main idea is to construct a relatively simple process such that, in its equilibrium state, the process can be used to approximate the distribution of interest. This approximating process will be constructed such that it is Markovian, i.e., the future evolution of the process can only depend on its past states through its present value.

Any Markov Chain Monte Carlo algorithm can be conceptualised as a set of rules dictating the movement of a particle through space. The implied motion of the particle ensures that, over time, the particle explores the space in accordance with the probability distribution of interest. In the long run, the particle thus spends time in any region of the space that is proportional to the probability mass assigned to that region by our distribution of interest. The simulated path of our particle can subsequently be used to estimate quantities related to the distribution of interest. MCMC methods

are generally acknowledged to be the most versatile algorithms for simulating a probability distribution. MCMC algorithms are extensively applied in a wide array of fields, ranging from statistics and machine learning to physics.

We consider the problem of sampling from the *target* distribution π defined on $E \subseteq \mathbb{R}^N$. Typically, the objective is to generate samples from π and compute some collection of features of this distribution that can usually be expressed as expectations with respect to π; in other words, we are interested in

$$\pi(f) := \int_E f(x)\pi(dx), \tag{1.1}$$

for some appropriately integrable function $f : E \to E'$ with $E' \subseteq \mathbb{R}^d$. We will refer to E and E' as the state space and feature space, respectively.

We note that even for problems of moderate dimension, alternative deterministic numerical integration techniques for computing $\pi(f)$ are not feasible; see for example, the results of Hinrichs et al. [80] and their given references. Hinrichs et al. [80] show that the number of required computations in order to approximate integrals up to some given precision level will grow super-exponentially with the dimension of the problem.

One of the largest fields that relies on MCMC is Bayesian statistics. In Bayesian statistics, practitioners are interested in the so-called posterior distribution, which quantifies the plausibility of the underlying statistical model after observing the data. Suppose that our data X takes values in a Polish space $(\mathfrak{X}, \mathcal{B}(\mathfrak{X}))$. Let our statistical model be parameterised with a Polish parameter space $(\Theta, \mathcal{B}(\Theta))$ such that it is given by a collection of probability measures on $(\mathfrak{X}, \mathcal{B}(\mathfrak{X}))$:

$$\mathcal{P} = \{P_\theta : \theta \in \Theta\}. \tag{1.2}$$

We endow the parameters of our model with prior distribution Π. By treating the unknown parameter as stochastic, the Bayesian paradigm inherently views statistical inference as the quantification of uncertainty. The prior is often interpreted as the quantification of our initial beliefs regarding the parameter values and, in high-dimensional settings, is practically used to induce regularisation. In this setting, the model (1.2) is interpreted as a collection of conditional laws of the data given the parameter. The posterior distribution represents our beliefs regarding the parameters after taking the observed data into consideration and is described by the conditional distri-

bution of the parameter given the data. The posterior is the primary object of interest in Bayesian inference. If the model is dominated by some σ-finite measure, with $\{p_\theta : \theta \in \Theta\}$ denoting the corresponding densities then the posterior is given by Bayes's formula

$$\Pi(B|X) = \frac{\int_B p_\theta(X) d\Pi(\theta)}{\int_\Theta p_\theta(X) d\Pi(\theta)},$$

where B denotes a measurable subset of Θ. The posterior can thus be obtained by reweighing the prior according to the plausibility of the parameter values implied by the observed sample. MCMC algorithms are in most practical problems necessary to infer important aspects of this posterior distribution, such as its mean and spread.

The fundamental idea behind MCMC is to construct a Markov chain such that its equilibrium distribution is given by the distribution of interest. For simplicity, we will consider discrete-time processes in this introduction. We say that $(X_k)_{k\in\mathbb{N}}$ with initial distribution ν_0 is a time-homogeneous Markov chain on some filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_k)_{k\in\mathbb{N}}, \mathbb{P}_{\nu_0})$ taking values in $(E, \mathcal{E})$, where E denotes the state-space and $\mathcal{E}$ the corresponding Borel σ-algebra, if for all $B \in \mathcal{E}$ we have that $\mathbb{P}_{\nu_0}(X_0 \in B) = \nu_0(B)$ and for all $m, n \in \mathbb{N}_0$ we have

$$\mathbb{P}_{\nu_0}\left(X_{n+m} \in B|\mathcal{F}_n\right) = P^m(X_n, B) \ \mathbb{P}_{\nu_0}\text{–a.s.},$$

where $P^m(x, \cdot)$ denotes the m-step transition kernel of the chain defined as

$$P^m(x, B) = \mathbb{P}(X_m \in B|X_0 = x), \quad m \in \mathbb{N}, x \in E, B \in \mathcal{E}.$$

We say that π is the equilibrium distribution of the Markov chain if $\pi P = \pi$, which is shorthand notation for

$$\int_E P(x, B)\pi(dx) = \pi(B) \quad \forall B \in \mathcal{E}.$$

Note that if the starting distribution is equal to the equilibrium distribution, then so will all subsequent marginal distributions of the chain. Therefore, the equilibrium distribution is often also referred to as the stationary or invariant distribution. We will choose the transition kernel of the chain so that the equilibrium distribution coincides with our target distribution of interest.

We refer to Chapter 2 for some concrete examples.

Under some regularity conditions, we can show that for any initial state $x \in E$, the chain will, in the long run, explore the state space according to its equilibrium distribution. More formally stated, it can be shown that the law of the chain converges in total variation to the stationary measure, i.e.,

$$\lim_{k\to\infty} \sup_{B\in\mathcal{E}} \left| P^k(x, B) - \pi(B) \right| = 0. \tag{1.3}$$

Moreover, we have that the ergodic law of large numbers holds:

$$\hat{\pi}_T(f) := \frac{1}{T}\sum_{k=0}^{T-1} f(X_k) \xrightarrow{a.s.} \int_E f(x)\pi(dx) \text{ as } T \to \infty, \tag{1.4}$$

provided that $\pi(\|f\|) < \infty$. This can be interpreted as the time average of the process converging to the space average with respect to the equilibrium distribution. Consequently, we see that $\hat{\pi}_T(f)$ is a consistent estimator of our features of interest $\pi(f)$. However, the ergodic law of large numbers does not describe the convergence rate of $\hat{\pi}_T(f)$, nor does it provide any insight regarding the precision of the estimator in finite samples. In order to assess the accuracy of $\hat{\pi}_T(f)$, we require a central limit theorem (CLT) to hold;

$$\sqrt{T}\left(\frac{1}{T}\sum_{k=0}^{T-1} f(X_k) - \int_E f(x)\pi(dx)\right) \xrightarrow{w} \mathcal{N}_d(0, \Sigma_f) \text{ as } T \to \infty, \tag{1.5}$$

where the asymptotic covariance matrix is given by

$$\Sigma_f = \mathrm{Var}_\pi(f(X_0)) + \sum_{k=1}^{\infty} \mathrm{Cov}_\pi(f(X_0), f(X_k)) + \sum_{k=1}^{\infty} \mathrm{Cov}_\pi(f(X_0), f(X_k))^T. \tag{1.6}$$

Jones [85] gives an expository survey on conditions under which the Markov chain CLT holds. Note that due to the dependence structure of the chain, the asymptotic variance Σ_f contains auto-covariance matrices, whereas if we were able to obtain i.i.d. samples from the target distribution, the asymptotic covariance matrix Σ_f would simplify to $\mathrm{Var}_\pi(f(X_0))$.

1.1. Mixing and Termination times

There are two practical questions that need to be addressed for every application of MCMC. Firstly, when is it reasonable to assume that the sampling algorithm is exploring the state space according to its stationary distribution? And when is it justified to terminate the simulation? Additionally, in high-dimensional settings, the scale of the problem introduces further challenges to these key inquiries. In Qin and Hobert [126] and Rajaratnam and Sparks [128] it is shown that many results regarding the convergence of MCMC samplers, that do not explicitly take the impact of dimensionality into consideration are inapplicable to high-dimensional scenarios. This highlighted the importance of the so-called convergence complexity of MCMC, which entails understanding how these convergence properties scale as the dimension of the problem grows.

In order to determine when the chain is sufficiently close to stationarity, we analyse the mixing time of the chain, which is defined as

$$t^*(\epsilon, \nu_0) := \inf\{t > 0 : D(\nu_0 P^t, \pi) < \epsilon\}, \tag{1.7}$$

where ν_0 denotes the initial distribution of the chain, P^t denotes the Markov transition kernel of the chain, ϵ is the specified tolerance level, and D is some metric on the space of probability measures, usually total variation or Wasserstein distance. The dependence of mixing times on the dimension of the underlying state space has been studied by Bou-Rabee and Eberle [25]; Hairer et al. [77]; Dalalyan [43]; Durmus and Moulines [57; 58]; Qin and Hobert [125; 127]; Yang and Rosenthal [156] among others. In the setting of Bayesian computation Altmeyer [2]; Belloni and Chernozhukov [11]; Tang and Yang [147]; Nickl and Wang [114] obtained dimension-dependent mixing time results, utilising the concentration properties of posterior distributions arising from various statistical models.

While a significant amount of research has been done to provide dimension-dependent performance guarantees for MCMC algorithms, the focus has been primarily on the first question. The issue of when to terminate the simulation has not been examined as thoroughly. This dissertation seeks to fill that gap by giving a rigorous exploration of the convergence complexity of MCMC output analysis and providing theoretical guarantees for termination criteria in high-dimensional settings. As noted by Gong and Flegal [75], many output

analysis tools used for addressing the termination question, such as visual inspection of trace plots and classical convergence diagnostics, are only appropriate for low dimensional problems. Termination criteria are usually defined as stopping times that follow a confidence set for the features of interest over the simulation trajectory. They allow termination of the simulation when a specified precision requirement is met. Analysing these termination criteria in high-dimensional settings remains an unaddressed problem.

While Markov Chain Monte Carlo methods are indeed less susceptible to the curse of dimensionality compared to deterministic integration methods, they are *not* unaffected by the challenges posed by high-dimensional settings. Since MCMC is approximate inference, both the mixing time and the termination time can potentially lead to computation times that are exponentially increasing in dimension. In Bandeira et al. [7], it has been demonstrated that with a worst-case initialisation, local MCMC samplers will require $\mathcal{O}(e^d)$ number of iterations to reach the bulk of the mass of the posterior distribution resulting from a non-linear regression model. Furthermore, it is also well-known that multi-modal distributions present a significant challenge for MCMC algorithms, as the time for transitioning between modes typically depends exponentially on their relative difference in potential energy, as shown in, for example Bovier et al. [27] and Monmarché [110]. Similarly, it is not straightforward to determine whether for an arbitrary problem an estimate of $\pi(f)$ that is within some desired precision level can even be obtained within polynomial time.

1.2. Gaussian approximations for MCMC algorithms

Fixed running-time implementations of MCMC offer no uncertainty guarantees, and prematurely terminating our simulation algorithm could even result in inconsistency of the estimator $\hat{\pi}_T(f)$. Even if the algorithm is already in its stationary phase, if it mixes slowly, i.e., it takes a long time to explore the state-space and move between different regions of high π-probability, the simulation output might not adequately represent the target distribution. Therefore, it is of great importance to use termination criteria that can give us precision guarantees. This problem is closely related to the uncertainty quantification of our algorithm. In Glynn and Whitt [73], asymptotic validity of several sequential termination rules is established under the assumption of a functional central limit theorem (FCLT) for the simulation process. Consider the rescaled partial-sum process of our Markov chain X given by $S^{(T)} := (S_t^{(T)})_{t\in[0,1]}$ where

$$S_t^{(T)} := \frac{1}{\sqrt{T}} \sum_{k=1}^{\lfloor Tt \rfloor} (f(X_k) - \pi(f)),\ t \in [0,1], \tag{1.8}$$

and let $Z := (Z_t)_{t\in[0,1]}$ be defined as $Z_t := \Sigma_f^{1/2} W_t$, where $\Sigma_f^{1/2}$ denotes the square root of the time-average covariance matrix of X and $W = (W_t)_{t\in[0,1]}$ is a standard d–dimensional Brownian motion, and let $D[0,1]$ denote the Skorokhod space, which consists of all $\mathbb{R}^d$-valued càdlàg functions with domain $[0,1]$, and let $\mathscr{D}$ denote the Borel σ–algebra generated by the Skorokhod topology. The functional central limit theorem states that

$$\left(\frac{1}{\sqrt{T}} \sum_{k=1}^{\lfloor Tt \rfloor} (f(X_k) - \pi(f)) \right)_{t\in[0,1]} \xrightarrow{w} (Z_t)_{t\in[0,1]},$$

by which we mean that for every continuity set $A \in \mathscr{D}$ of Z, we have that

$$\left| \mathbb{P}\left(S^{(T)} \in A\right) - \mathbb{P}\left(Z \in A\right) \right| = o(1) \text{ as } T \to \infty. \tag{1.9}$$

In order to generalise the results of Glynn and Whitt [73] and obtain asymptotic validity of termination criteria in high-dimensional settings, we require quantitative Gaussian approximations of our Markov chain X. These Gaussian approximation results quantify the rate at which the trajectories of the partial sum process can be approximated by the appropriately scaled

trajectories of a Gaussian process and are thus a refinement of the functional central limit theorem, as shown in for example [123; Theorem 1.E]. We say that a weak Gaussian approximation holds for $X = (X_t)_{t\in\mathbb{N}}$ if the process can be defined on a probability space, together with a Brownian motion W, such that

$$\lim_{T\to\infty} \mathbb{P}\left(\frac{1}{\Psi_T}\left|\sum_{t=0}^{T} f(X_t) - T\pi(f) - \Sigma_f^{1/2} W_T\right| \leqslant K\bar{\psi}_N\psi_d\right) = 1, \tag{1.10}$$

where $|\cdot|$ denotes the Euclidean norm, K denotes a dimension-independent almost surely finite constant, and $\bar{\psi}_N, \psi_d$, and Ψ_T denote the dependence of the approximation error on the dimension of the state space, the dimension of the feature space, and the sampling time respectively. Similarly, we say that a strong Gaussian approximation holds for $X = (X_t)_{t\in\mathbb{N}}$ if

$$\mathbb{P}\left(\limsup_{T\to\infty} \frac{1}{\Psi_T}\left|\sum_{t=0}^{T} f(X_t) - T\pi(f) - \Sigma_f^{1/2} W_T\right| \leqslant K\bar{\psi}_N\psi_d\right) = 1. \tag{1.11}$$

We use the customary notation for Gaussian approximations, i.e.,

$$\left|\sum_{t=0}^{T} f(X_t) - T\pi(f) - \Sigma_f^{1/2} W_T\right| = \begin{cases} \mathcal{O}_P\left(\bar{\psi}_N\psi_d\Psi_T\right) \\ \\ \mathcal{O}_{a.s.}\left(\bar{\psi}_N\psi_d\Psi_T\right) \end{cases}, \tag{1.12}$$

where $\mathcal{O}_P$ and $\mathcal{O}_{a.s.}$ denote the weak and strong approximation respectively. These Gaussian approximation results are closely related to the convergence rate of the FCLT, as shown in Csörgö and Horváth [38; Theorem 1.16 and Theorem 1.17]. These approximation results are powerful tools used to obtain numerous results in both probability and statistics as seen in, e.g., Csörgö and Hall [41] and Shorack and Wellner [139]. For applications in MCMC, these results tell us how the trajectory of the sampling process will fluctuate around the value it is designed to approximate. In the one-dimensional case, Gaussian approximation results for MCMC were obtained by Flegal and Jones [66]; Csáki and Csörgő [37]; Jones et al. [87] and Merlevède et al. [103]. These results were extended to multivariate setting by Banerjee and Vats [8] and Li and Qin [95]. For a more extensive overview of Gaussian approximation results, we refer to Section 3.1 of Chapter 3. We note that Gaussian approximations are in the literature often also referred to as strong invariance principles and Gaussian embeddings.

Since MCMC methods are only able to sample approximately from the target distribution, quantifying the uncertainty of the simulation output is of central importance. Stating only the point estimate $\hat{\pi}_T(f)$ for an application where we are interested in $\pi(f)$ can be misleading since it does not convey the reliability of the estimate. In order to give a standard error, we need to estimate the asymptotic covariance matrix Σ_f. Gaussian approximation results also play a central role in the analysis of estimators of the asymptotic variance and MCMC output analysis, see for example Damerdji [44; 45]; Flegal and Jones [66]; Jones and Hobert [86]; Jones et al. [87]; Vats et al. [151; 152]. Furthermore, the asymptotic variance is also necessary for computing many convergence diagnostics and for the implementation of commonly used termination criteria. Convergence diagnostics are statistical tests that can be used to assess stationarity and proper exploration of our MCMC sampler.

1.3. Contributions and organisation of the thesis

In Chapter 2, we give examples of classic MCMC algorithms and introduce sampling algorithms based on Piecewise Deterministic Markov Processes (PDMPs). Contrary to most commonly used MCMC algorithms, these are non-reversible and continuous-time processes. We also introduce drift and minorisation conditions, which are commonly used to obtain convergence guarantees for MCMC algorithms. Furthermore, we give a more extensive introduction to MCMC output analysis, which encompasses estimation of the asymptotic variance and obtaining validity of termination criteria. Additionally, we provide a brief introduction to convergence diagnostics, which are also a significant part of MCMC output analysis. Furthermore, we also discuss some suggestions for possible convergence diagnostics suitable for PDMP based sampling algorithms.

In Chapter 3, we obtain novel Gaussian approximation results for a broad class of ergodic (continuous-time) Markov processes. The arguments used to obtain strong invariance principles for discrete-time processes do not directly carry over to the continuous-time case. For the Zig-Zag sampler, which is a specific PDMP, we show that the optimal Gaussian approximation rate can be obtained. This is the first MCMC sampler, within the considered class of problems, for which this optimal rate has been obtained. We demonstrate how these results can be used to analyse the batch means method for simulation output of Piecewise Deterministic Monte Carlo (PDMC) samplers.

Consequently, many results regarding uncertainty quantification now carry over to PDMC samplers. Since previous work on estimation of the MCMC standard error is based on strong invariance principles with limited accuracy, our results weaken the currently available regularity conditions guaranteeing strong convergence of the batch means estimator in any MCMC setting. We also derive a fluctuation result for additive functionals of ergodic diffusions using our Gaussian approximation results. The results of Chapter 3 resulted in the publication

> A. Pengel and J. Bierkens. Strong invariance principles for ergodic Markov processes. *Electronic Journal of Statistics*, 18(1):191–246, 2024.

In Chapter 4, we consider the problem of uncertainty quantification for MCMC in the high-dimensional setting. Firstly, in the high-dimensional setting, it is not even clear if a Markov central limit theorem holds. The critical challenge in the high-dimensional setting is to determine how long the simulation time must be to ensure the validity of the Gaussian approximations as the dimension of the problem grows. This dissertation contributes to the field by introducing novel dimension-dependent bounds for the Gaussian approximation of a wide class of MCMC samplers. Our obtained approximation results provide explicit bounds for $\bar{\psi}_N$ and ψ_d, the dependence of the dimension of the state space and feature space on the approximation rate. For applications in Bayesian statistics, our results provide a direct link between the statistical model complexity and the computational complexity of the MCMC algorithm. Furthermore, this work obtains the first result attaining the optimal approximation rate for MCMC samplers in a multivariate setting. We also consider a wider class of ergodic processes than previously examined in MCMC output analysis and give novel results that quantify the effect of slower ergodicity rates on the Gaussian approximation error. Our results provide us with explicit simulation requirements such that a Markov CLT holds. Moreover, our results can be used to adapt the tuning parameters of the variance estimation methods considered in Vats et al. [152; 151] to high-dimensional settings and give simulation requirements that guarantee the validity of these variance estimation methods. Therefore, we are able to give conditions for valid uncertainty quantification for high-dimensional MCMC algorithms.

Finally, our results enable us to study the convergence complexity of a broad range of termination criteria and show that the termination rules introduced in Glynn and Whitt [73] and Vats et al. [152] can be applied to high-dimensional settings. These results explicitly describe how the termination time scales with the desired precision and the complexity of the problem. We give conditions that guarantee that the termination time of an MCMC algorithm scales polynomially in dimension while ensuring a desired level of precision. The results presented in Chapter 4, resulted in the following papers:

> A. Pengel, J. Yang, and Z. Zhou. Gaussian approximation and output analysis for high-dimensional MCMC.
> *arXiv preprint, arXiv:2407.05492*, 2024.

> A. Pengel, J. Yang, and Z. Zhou. High-dimensional Gaussian approximation for continuous-time processes. *In preparation.*

Chapter 2

MCMC and Output Analysis

The purpose of Chapter 2 is to give an overview and discussion on some existing results related to MCMC and output analysis. In Section 2.2, we discuss several discrete-time and continuous-time MCMC algorithms. We also give a brief introduction to drift and minorisation conditions. In Section 2.3, we give an overview of some existing results on MCMC output analysis, including results on termination criteria and estimation of the asymptotic variance. Furthermore, we also briefly discuss some proposed ideas for convergence diagnostics for PDMPs.

2.1. Introduction

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

$$\pi(x) = \frac{e^{-U(x)}}{\int_E e^{-U(x)}\ dx}, \tag{2.1}$$

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.

2.2. Markov Chain Monte Carlo Algorithms

The Metropolis adjusted Langevin algorithm (MALA) and the Hamiltonian Monte Carlo (HMC) sampler, introduced by Roberts and Tweedie [134] and

Duane et al. [55] respectively, are widely used MCMC samplers in high-dimensional settings. Also the Gibbs sampler, Gelfand and Smith [69], and slice sampler, Neal [113] and Murray et al. [111], have enjoyed widespread use.

We give a brief introduction to MALA. The main idea originates from the so-called overdamped Langevin diffusion, which is the process $X = (X_t)_{t\geq 0}$ that solves the stochastic differential equation

$$dX_t = -\nabla U(X_t)dt + \sqrt{2}dW_t, \tag{2.2}$$

where $(W_t)_{t\geq 0}$ is a standard d-dimensional Brownian motion. Under mild conditions on the potential U it can be shown that π is the stationary distribution of the Langevin diffusion. However, since, in general, the transition kernel of this process is intractable, we cannot simulate this process exactly. The Euler–Maruyama discretisation, which approximates the transition kernel over a time step h with a Gaussian distribution, gives us the following Markov chain $Y = (Y_k)_{k\in\mathbb{N}}$

$$Y_{k+1} = Y_k - h\nabla U(Y_k) + \sqrt{2h}Z_k, \tag{2.3}$$

where (Z_k) denotes a sequence of i.i.d standard Gaussians. However, due to the incurred approximation error, for fixed discretisation step $h > 0$, the process Y will not have π as its stationary distribution. One possible solution to compensate for this is by adding a so-called accept-reject step. Inspired by the classical Metropolis-Hastings algorithm, proposed by Metropolis et al. [104] and Hastings [78], the idea is to use the chain Y as proposed steps and to subsequently accept these steps according to some probability $\alpha(x, y)$ which depends on the current and proposed state, such that the resulting stationary distribution is π. We give the simulation scheme of MALA in Algorithm 1 below.

Algorithm 1. Metropolis adjusted Langevin algorithm

1. Initialise X_0.
   For $k = 1, 2, \ldots, T$ do
2. Simulate a proposal Y_k according to

$$Y_k \sim \mathcal{N}_d(X_{k-1} - h\nabla U(X_{k-1}), 2hI_N).$$

3. Simulate $U \sim \text{Unif}([0, 1])$
4. If $U \leq \alpha(X_{k-1}, Y_k)$:
   Set $X_k = Y_k$.
5. Else:
   Set $X_k = X_{k-1}$.

The Metropolis-Hastings acceptance probability is given by

$$\alpha(x, y) = 1 \wedge \frac{\pi(y)q(x; y - h\nabla U(y), 2hI_d)}{\pi(x)q(y; x - h\nabla U(y), 2hI_d)}, \tag{2.4}$$

where $q_d(\cdot; m, \Sigma)$ denotes the $\mathcal{N}_d(m, \Sigma)$ density function. Note that the ratio of the evaluations of π appearing in (2.4) will tend to give a higher acceptance probability when the proposal is in a higher π-density region and will decrease the acceptance probability for proposals in lower π-density regions. Similarly, the ratio of the proposal densities corrects for the asymmetry of the proposals. It is not difficult to show that the transition kernel of MALA satisfies the detailed balance condition, namely,

$$\pi(dx)P(x, dy) = \pi(dy)P(y, dx). \tag{2.5}$$

Note that the detailed balance condition can be interpreted as follows: in stationarity, the probability of the chain being in dx and subsequently moving to dy is equal to the probability of the chain being in dy and then moving to dx. It can be shown that this is equivalent to the chain being invariant with respect to time reversal, i.e., under the stationary measure, we have that $(X_s)_{0 \leq s \leq t} \sim (X_{t-s})_{0 \leq s \leq t}$. Hence, we say that the chain X is π-reversible. It is easy to show that a π-reversible process leaves π invariant. However, this reversibility can lead to diffusive behaviour and, thus slow mixing. Say we are in a region where π stays roughly equal, for example, in the tail of a fat-tailed distribution. Then, for any proposal, $\pi(y)$ will not be substantially

different from the current state $\pi(x)$. Hence, (2.5) implies that $P(x, dy)$ and $P(y, dx)$ will be closely matched, which can lead to diffusive behaviour. The sampler will thus explore the state-space in small steps, similar to a random walk. Consequently, it will take longer to reach distant regions and thus the sampling algorithm will exhibit poor convergence rates.

2.2.1. Piecewise Deterministic Monte Carlo

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 has been shown that non-reversibility can significantly improve the performance of sampling methods, in terms of both convergence rate to equilibrium and asymptotic variance, see for example, the results of Hwang et al. [83] and Lelievre et al. [93] regarding convergence to stationarity and Duncan et al. [56] and Rey-Bellet and Spiliopoulos [131] 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, the Bouncy Particle Sampler, and the Boomerang sampler, introduced in Bierkens and Roberts [20], Bouchard-Côté et al. [26], and Bierkens et al. [23] 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 Bierkens et al. [21].

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 (ZZP), the set of possible velocities is given by $\mathcal{V} = \{-1, +1\}^d$. We distinguish N 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

$$\lambda_i(x, v) = (v_i \partial_{x_i} U(x))^+ , \tag{2.6}$$

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

$$\lambda_Z(x, v) = \sum_{i=1}^{d} \lambda_i(x, v) = \sum_{i=1}^{d} (v_i \partial_{x_i} U(x))^+ . \tag{2.7}$$

We see that given an initial position and velocity, in components where we are moving against the gradient of our potential U, then the switching intensity will be low. Since the direction of this component contributes to moving in a direction with higher probability mass relative to our current position. Analogously, components that are moving in the direction of the gradient of the potential will have a higher switching intensity. The simulation scheme for Zig-Zag is given in Algorithm 2 below.

Algorithm 2. Zig-Zag Sampler

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

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

   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 Bierkens et al. [21] 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

$$\lambda_i(x, v) = \left(-v_i \frac{\partial_{x_i} \pi_i(x_i)}{\pi_i(x_i)}\right)^+ = \left(v_i \partial_{x_i} U_i(x)\right)^+, \tag{2.8}$$

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 $i = 1, \ldots, d$.

Bouncy Particle Sampler

For the Bouncy Particle Sampler (BPS), the state-space of the velocity is given by $\mathcal{V} = \mathbb{R}^d$. The rate at which the BPS changes its velocity is determined by the directional derivative of the potential U along the direction of the velocity evaluated at the current position

$$\lambda_B(x, v) := \left(\langle \nabla U(x), v \rangle\right)^+. \tag{2.9}$$

If $\langle \nabla U(x), v \rangle < 0$, then moving according to our velocity from our current position will result in a decrease of potential energy. This corresponds to moving to a region with higher probability mass, and therefore we desire the switching intensity to be zero. Similarly, the larger $\langle \nabla U(x), v \rangle$, the more increase in potential energy is gained by moving along direction v relative to our current position. This specification of the intensity motivates the following changes for the velocity at events:

$$R(x)v := v - 2\frac{\langle \nabla U(x), v \rangle}{\|\nabla U(x)\|^2} \nabla U(x). \tag{2.10}$$

Note that the projection of the velocity v onto $\nabla U(x)$, the gradient of the potential, is given by

$$\frac{\langle \nabla U(x), v \rangle}{\|\nabla U(x)\|^2} \nabla U(x).$$

Therefore an application of the reflection $R(x)$ to velocity v will result in flipping the components of v that are in the direction of $\nabla U(x)$, whereas, the components orthogonal to $\nabla U(x)$ remain unchanged. Flipping the velocity components in the direction of $\nabla U(x)$ at events, is desirable since given our position, $-\nabla U(x)$ gives the direction of the steepest descent in potential energy.

Unfortunately, the canonical rates (2.9) can result in a reducible process for elliptical symmetric targets and therefore, we add a refreshment rate $\gamma : \mathfrak{X} \to \mathbb{R}^+$ to our switching intensity

$$\bar{\lambda}(x, v) := \lambda_B(x, v) + \gamma(x).$$

At events, we will refresh with probability determined by the relative intensity of the refreshment and the canonical rate. At refreshments, we will draw the velocity according to its stationary distribution ν. The Bouncy Particle Sampler is given in Algorithm 3 below.

Algorithm 3. Bouncy Particle Sampler

1. Initialise $(X_0, V_0) \leftarrow (x, v)$ and $T_0 \to 0$
2. For $k = 1, 2, \cdots$ simulate τ_k according to

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

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 $T_k = T_{k-1} + \tau_k$
5. Update velocity at the event time
   with probability $\lambda(X_{\tau_k}, V_{\tau_k})/\bar{\lambda}(X_{\tau_k}, V_{\tau_k})$ reflect the velocity $V_{T_k} = R(X_{T_k})V_{T_{k-1}}$
   Otherwise, refresh the velocity according to $V_{T_k} \sim \nu$

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 Fearnhead et al. [65]. It can be shown that under very mild regularity conditions, both sampling processes admit a stationary distribution given by

$$\mu(dx, dv) = \pi(dx)\upsilon(dv), \tag{2.11}$$

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 $\pi(f)$.

Remark 2.2.1. Both the ZZS and BPS can be simulated without discretisation error through Poisson thinning; see Lewis and Shedler [94]. However, this requires finding a dominating function for the switching intensities, which in some applications might be difficult. We refer to the results of Bertazzi

et al. [15] and Bertazzi et al. [16] for discretisation schemes of PDMPs, which at the cost of exactness, only require point-wise evaluations of the gradient for simulating event times. △

2.2.2. Drift and Minorisation conditions

Drift and minorisation conditions are widely used for obtaining quantitative bounds for the mixing time of Markov chains. The drift condition describes how fast the Markov chain moves towards subsets of the state space, while the minorisation condition controls how fast the Markov chain forgets its past. We say that the Markov chain satisfies a geometric drift condition if Drift Condition 1 holds.

Drift Condition 1. *Let there exist a function $V : E \to \mathbb{R}^+$, some set C, constants $\lambda \in (0,1)$ and $0 < b, \upsilon_C < \infty$ such that $\upsilon_C = \sup_{x \in C} V(x)$ and*

$$PV(x) = \int_E V(y)P(x,dy) \leq \lambda V(x) + b\mathbb{1}_C(x),$$

for some set $C \in \mathcal{E}$.

More specifically, the function V in Drift Condition 1 describes how fast the chain in expectation will move towards the set C given that the chain is currently in state x and how long the chain is expected to stay in this set C. An appropriate drift function should have low values in high-probability regions of the state space. Note that Drift Condition 1 implies that while the chain is not in C, the value of the drift function will decrease geometrically. In many applications, we can only guarantee that the drift function decays at a polynomial rate while the process is not in C. This corresponds to the following polynomial drift condition.

Drift Condition 2. *Let there exist a function $V : E \to \mathbb{R}^+$, some set C, constants $0 < c, b, \upsilon_C < \infty$, such that $\upsilon_C = \sup_{x \in C} V(x)$ and $\eta \in (0,1)$ such that*

$$PV(x) \leq V(x) - cV(x)^\eta + b\mathbb{1}_C(x),$$

for some set $C \in \mathcal{E}$ with $\pi(C) > 0$.

We say that an associated local m_0-step minorisation condition holds for the Markov chain if the following holds.

Minorisation Condition 1. *Let ν be some probability measure defined on C such that*

$$P^{m_0}(x,\cdot) \geq \alpha \mathbb{1}_C(x)\nu(\cdot),$$

the minorisation volume $\alpha \in (0,1]$, $m_0 \in \mathbb{N}$ and small set C with $\pi(C) > 0$.

It is known that a multi-step minorisation condition holds for all widely used MCMC algorithms, see for example Meyn and Tweedie [107; Proposition 5.4.5]. Often it can even be shown that the Markov chain satisfies a one-step minorisation condition. The minorisation condition controls how fast the process forgets its past. In Section 3.2 and Section 4.5.1 we give a detailed discussion. The main idea is that every time we hit set C, we have a probability α the m_0-skeleton point of the chain will be independent of the current state.

For continuous-time processes, we can also introduce the drift and minorisation framework. 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 $(E, \mathscr{E})$ with transition semigroup given by $(P_t)_{t\geq 0}$ with finite invariant measure π. For continuous-time processes, the petite set will now play the role of the small set. We say that a set C is petite if there exists some non-trivial measure ν such that for all $x \in C$ we have that

$$\int P_t(x,\cdot)a(dt) \geq \alpha\nu(\cdot), \tag{2.12}$$

with $\alpha \in (0,1]$, $\pi(C) > 0$, and sampling scheme a, which is a probability distribution on $\mathbb{R}^+$. Note that if $a(dt) = \delta_{t_0}(dt)$, then we have a small set condition for the continuous-time process. In order to define drift conditions for continuous-time processes, we introduce the generator. We say that $\mathcal{L}$ is the generator of process X with domain $\mathcal{D}$ if for every function $f \in \mathcal{D}$ we have that

$$\mathcal{L}f(x) = \lim_{t\to 0} \frac{\mathbb{E}_x f(X_t) - f(x)}{t}, \tag{2.13}$$

where the limit is uniform with respect to the supremum norm and $\mathcal{D}$ is the set of continuous functions that are vanishing at infinity such that the limit (2.13) exists. The generator of a process thus describes the expected infinitesimal evolution. Note that we can extend this concept and introduce to

so-called extended generator, which is defined as the solution of the martingale problem of process X. The extended generator has a larger domain and coincides with the generator on $\mathcal{D}$. We refer to Deligiannidis et al. [50] and Bierkens et al. [22] for a characterisation of the (extended) generator of BPS and ZZS, respectively. We say that X satisfies an exponential drift condition if Drift condition 3 holds.

Drift Condition 3. *Let there exist a function $V : E \to \mathbb{R}^+$, $\lambda \in (0, 1)$, $b < \infty$, and petite set C such that*

$$\mathcal{L}V \leq -\lambda V + b\mathbb{1}_C(x)$$

Similarly, we can define a polynomial drift condition for X;

Drift Condition 4. *Let there exist a function $V : E \to \mathbb{R}^+$, $\alpha \in (0, 1)$, $\zeta \in (\alpha, 1)$, $c_\eta > 0$, $b < \infty$, and petite set C such that*

$$\mathcal{L}V^\eta \leq -c_\eta V^{\eta-\zeta} + b\mathbb{1}_C(x).$$

Drift and minorisation conditions provide a general approach for determining the speed of convergence to stationarity. We say that a Markov process X is *ergodic* with convergence rate Ψ if

$$\|P_t(x, \cdot) - \pi\|_{TV} \leq V(x)\Psi(t), \quad \text{for all } x \in E \text{ and } t \geq 0, \tag{2.14}$$

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 $\beta, \gamma > 0$. The definition is applicable to both the discrete-time and continuous-time settings. For a more thorough discussion of these definitions, we refer to Meyn and Tweedie [105].

In Deligiannidis et al. [50] and Bierkens et al. [22], exponential ergodicity for the Bouncy Particle Sampler and the Zig-Zag process are established respectively through the drift and minorisation approach of Meyn and Tweedie [107]. Note that explicit ergodicity rates can directly be used to obtain mixing time bounds. For other general bounds for the mixing time based on the drift and minorisation framework, we refer to Rosenthal [136].

2.3. Output Analysis

2.3.1. Termination Criteria

Because MCMC methods can only approximate the target distribution, it is crucial to assess the uncertainty in the simulation results. Therefore, the point estimate $\hat{\pi}_T(f)$ should always be accompanied by its standard error. However, if we keep our simulation going until the standard error is sufficiently small, we are effectively using a termination rule. The same is the case if we terminate based on another statistic. Practitioners often terminate the simulation when the effective sample size is sufficiently large. The effective sample size is defined in (2.20) and (2.23). The effective sample size (ESS) is a measure of the efficiency of an MCMC algorithm, incorporating the difference in magnitude between the MCMC standard error and the independent Monte Carlo error. The ESS quantifies how many independent samples the MCMC simulation output is equivalent to. Note that the computation of the ESS also requires an estimate of the asymptotic covariance matrix Σ_f. Moreover, it is well known that stopping according to the effective sample size is equivalent to termination rules based on confidence sets, see for example Vats et al. [152].

The Fixed Volume Stopping Rule (FVSR) is a sequential stopping procedure where the simulation is terminated when the volume of a confidence set for $\pi(f)$ falls below some predetermined tolerance level. By the CLT, we can immediately construct a confidence ellipsoid $C(T)$ for our parameter of interest, namely,

$$C(T) = \{\theta \in \mathbb{R}^p : T(\hat{\pi}_T(f) - \theta)^\top \hat{\Sigma}_T^{-1}(\hat{\pi}_T(f) - \theta) < q_\alpha\}, \tag{2.15}$$

where $\hat{\Sigma}_T$ denotes some consistent estimator of the asymptotic covariance matrix, which is evaluated using simulation output until time T, and where q_α denotes, depending on the choice of estimator $\hat{\Sigma}_T$, the appropriate quantile of a Chi-squared or Hotelling's T-squared distribution such that the coverage of the confidence ellipsoid is $(1 - \alpha)$.

Given some user-specified tolerance level ε, the FVSR defines the time of termination $T_1(\varepsilon)$ for our simulation experiment as

$$T_1(\varepsilon) = \inf\{t \geq 0 : \mathrm{Vol}(C(t))^{1/d} + \Lambda(t) \leq \varepsilon \}. \tag{2.16}$$

Here $\Lambda(t)$ is some sequence of order $o(t^{-1/2})$. The role of $\Lambda(t)$ is to prevent early termination due to a possible inaccurate estimate of the covariance matrix due to small sample size. Moreover, the minimum simulation threshold should be chosen such that the MCMC sampler is in its equilibrium phase. A common choice is $\Lambda(t) = \mathbb{1}_{\{t>T^*\}} + 1/T$, for some minimal desired simulation sample size T^*.

Unfortunately, since our sample size at termination is random, it cannot be immediately guaranteed that the resulting confidence set indeed has the correct coverage. Glynn and Whitt [73] give an example where the FVSR is employed with an estimator of the covariance matrix that only converges in probability to the true asymptotic covariance matrix, and the resulting confidence interval does not have the correct coverage. If $\hat{\Sigma}_T \xrightarrow{p} \Sigma_f$ then we say that $\hat{\Sigma}_T$ is (weakly) consistent. If $\hat{\Sigma}_T$ converges almost surely to Σ_f, we say that the estimator is strongly consistent. We say that a functional weak law of large numbers (FWLLN) holds for $\hat{\Sigma}_T$ if

$$(\hat{\Sigma}_{Tt})_{t\in[0,1]} \xrightarrow{w} (\Sigma_f)_{t\in[0,1]},$$

here the weak convergence is in the Skorokhod topology. Note that in general the weak law of large numbers does not imply an FWLLN, as shown in Glynn and Whitt [71]. Conditions that result in asymptotic validity of the FVSR are described in the following theorem.

Theorem 2.3.1. *(Glynn and Whitt [73; Theorem 1; Theorem 2]) Let $\hat{\pi}_T(f)$ satisfy an FCLT and suppose that a FWLLN holds for $\hat{\Sigma}_T$ or that $\hat{\Sigma}_T$ is a strongly consistent estimator of Σ_f. Then as ε tends to zero we have that*

1. *The termination time $T_1(\varepsilon)$ is asymptotically equivalent to*

$$\varepsilon^2 T_1(\varepsilon) \xrightarrow{a.s.} c_{\alpha,p}^{2/d} |\Sigma_f|^{1/d}, \tag{2.17}$$

   where $|\cdot|$ denotes the determinant of a matrix and $c_{\alpha,p}$ denotes the product of $q_\alpha^{p/2}$ and the volume of a standard p-dimensional hyperball.

2. *The following central limit theorem holds at the termination time*

$$\varepsilon^{-1}\left(\frac{1}{T_1(\varepsilon)}\sum_{k=1}^{T_1(\varepsilon)} f(X_k) - \pi(f)\right) \xrightarrow{w} \mathcal{N}_d\left(0, c_{\alpha,p}^{-4/d}|\Sigma_f|^{-2/d}\ \Sigma_f\right) \tag{2.18}$$

3. Asymptotic validity of the resulting confidence set

$$\Pr\left(C(T_1(\varepsilon)) \ni \pi(f)\right) \to 1-\alpha. \tag{2.19}$$

We see that it is of central importance that we have an estimator $\hat{\Sigma}_T$ of the asymptotic covariance matrix Σ_f that is either strongly consistent or satisfies a functional weak law of large numbers. While a CLT is enough to guarantee the validity of a confidence interval, for termination criteria we follow the confidence interval over a simulation trajectory. Hence, we need to describe how the trajectory of both the sample average and the empirical confidence interval fluctuate around the values they are designed to estimate.

In Vats et al. [152] and Vats et al. [151], the strong consistency of the multivariate batch means and spectral variance estimator is proven under the assumption of a strong Gaussian approximation. Hence Theorem 2.3.1 justifies the use of the FVSR implemented with these estimators of the asymptotic variance.

Part 1 of Theorem 2.3.1, can be used to show that termination according to $T_1(\varepsilon)$ is asymptotically equivalent to termination according to the following definition of the effective sample size

$$\mathrm{ESS}_1 = \frac{T}{\left|\hat{\Sigma}_f\right|^{1/d}} \tag{2.20}$$

Since $\hat{\pi}(f)$ and $\hat{\Sigma}_f$ are in practice often correlated, choosing a justifiable tolerance ε is not straightforward. Therefore Vats et al. [152] propose the use of the Relative Fixed Volume Stopping Rule (RFVSR), where our stopping criteria takes the spread of our target distribution into account. This gives the following termination rule

$$T_2(\varepsilon) = \inf\{t \geq 0 : \mathrm{Vol}(C(t))^{1/d} + \Lambda(t) \leq \varepsilon \, |\hat{\Gamma}_f|^{1/2d}\}, \tag{2.21}$$

where

$$\hat{\Gamma}_f = \frac{1}{T}\sum_{k=1}^{T} \left(f(X_k) - \hat{\pi}_T(f)\right)\left(f(X_k) - \hat{\pi}_T(f)\right)^\top. \tag{2.22}$$

Under the conditions of Theorem 2.3.1, Vats et al. [152] show asymptotic validity of this stopping rule and show that it is asymptotically equivalent to

the following version of the effective sample size

$$\mathrm{ESS}_2 = T \left(\frac{|\hat{\Gamma}_f|}{|\hat{\Sigma}_f|} \right)^{1/d}. \tag{2.23}$$

Note that while the results of Vats et al. [151; 152] offer strongly consistent estimators of Σ_f, they are based on a strong Gaussian approximation with an implicit rate. Consequently, the results do not guarantee a convergence rate for any of the proposed estimators. Moreover, it is noted in Glynn and Whitt [73] that the minimum simulation threshold T^* should be chosen to ensure that the estimator of the asymptotic variance is reliable. Furthermore, Glynn and Whitt [73] assumes an FCLT for the simulation process. Therefore, the simulation threshold should also enforce the error introduced by the Gaussian approximation to be at least of smaller magnitude than the desired precision level. This would require a quantitative analysis of the introduced termination rules. In Section 4.4.2 of Chapter 4, we provide a solution to this challenge.

2.3.2. Convergence Diagnostics for Piecewise Deterministic Monte Carlo

Convergence diagnostics are statistical tests that can be used to assess the convergence of our Markovian sampler to stationarity and to determine if the algorithm is exploring the state-space properly. If our sampler is not moving according to its stationary distribution, the simulated output will misrepresent the target distribution. This can be the case if, for example, the sampler is initialised in a low-probability region of the target. Slow mixing can also occur when dealing with multi-modal target distributions. When the target distribution has multiple local maxima, the sampler can get 'trapped' in a high probability region around a local mode. Consequently, the sampler could not have explored the full support of the target distribution during a specified length of simulation.

A methodological approach would be to assess the convergence of our algorithm through statistical tests where the null hypothesis asserts that our process is moving according to its long-run dynamics and has explored the state space properly. Following this approach, we can never prove that our process has reached equilibrium. Despite this shortcoming of convergence

diagnostics, practitioners often assume that stationarity of their MCMC sampler is reasonable if they find insufficient evidence to reject this hypothesis.

We consider testing procedures for the null hypothesis, which postulates that starting from some time $\bar{t}$, we are in stationarity, i.e.,

$$H_0 : \ Z_t \sim \mu, \ \text{for all } t \geq \tilde{t},$$

where $\bar{t}$ denotes the burn-in time. For an extensive overview of commonly used convergence diagnostics in MCMC literature, we refer to Mengersen et al. [102], Roy [137], and Cowles and Carlin [36]. Most of these methods can also be applied to PDMC algorithms. The usefulness of this approach for examining stationarity will depend upon the power of the employed statistical tests. Note that this is a very difficult testing problem. This is a direct consequence of the obvious but critical factor that the target distribution π is unknown. Moreover, the positive correlation resulting from these samplers will result in larger standard errors, which leads to a decrease in power for most testing procedures. This will especially be the case if the sampler is mixing slowly.

Nevertheless, convergence diagnostics are still widely used. In the PDMP setting, we have the auxiliary velocity component at our disposal. Consequently, the null of stationarity has more components, which allows the use of more diagnostic tools. In stationarity, the distribution of the velocity component is known. Moreover, the position and velocity components are independent. We exploit these characteristics of existing PDMC algorithms to construct convergence diagnostics. As cornerstones for these testing procedures, we introduce the empirical distribution for the joint and marginal position and velocity components:

$$\mu_T(A \times B) = \frac{1}{T} \int_0^T \mathbb{1}_{\{X_t \in A; V_t \in B\}} dt, \tag{2.24}$$

$$\pi_T(A) = \mu_T(A \times \mathcal{V}) = \frac{1}{T} \int_0^T \mathbb{1}_{\{X_t \in A\}} dt, \tag{2.25}$$

$$\nu_T(B) = \mu_T(\mathfrak{X} \times B) = \frac{1}{T} \int_0^T \mathbb{1}_{\{V_t \in B\}}, dt \tag{2.26}$$

where $A \in \mathcal{B}(\mathfrak{X})$ and $A \in \mathcal{B}(\mathcal{V})$. By the ergodic law of large numbers, it immediately follows that these are strongly consistent estimators. Moreover, if we have that a CLT holds, then for every $C_1, \cdots, C_p \in \mathcal{B}(E)$, we have that

$$\sqrt{T}\left(\mu_T(C_1), \cdots, \mu_T(C_p)\right)^\top - \left(\mu(C_1), \cdots, \mu(C_p)\right)^\top\right) \xrightarrow{w} \mathcal{N}_p(0, \Sigma_C), \quad (2.27)$$

where C denotes the sets $C_1, \cdots, C_p$ and the asymptotic covariance matrix is given given by $(\Sigma_C)_{ij}$ for $1 \leq i, j \leq p$ by its entries

$$\Sigma_{Cii} = 2\int_0^\infty \Pr\left(Z_0 \in C_i; Z_t \in C_i\right)\ dt - 2\mu(C_i)^2$$

and for $i \neq j$

$$\Sigma_{Cij} = \int_0^\infty \Pr\left(Z_0 \in C_i; Z_t \in C_j\right)\ dt + \int_0^\infty \Pr\left(Z_0 \in C_j; Z_t \in C_i\right)\ dt - 2\mu(C_i)\mu(C_j),$$

where

$$\Pr\left(Z_0 \in D; Z_t \in C\right) = \int_D P_t(z, C)\mu(dz),$$

for measurable sets C and D.

Testing for equilibrium of velocity

We can immediately reject the hypothesis of stationarity of our process if the velocity component does not possess the correct distribution. The equilibrium distribution of the velocity component is fully specified. Therefore, the construction of an appropriate statistic is straightforward. For the Zig-Zag process, the stationary distribution of the velocity component is a uniform distribution over the velocity space $\{-1, 1\}^d$. A naive approach would consist of checking that the process attains every possible velocity vector for an equal amount of time. However, it would be more efficient to consider the components individually. Observe that under the null hypothesis of stationarity, the probability of every velocity component being -1 or $+1$ must be equal. Otherwise, it would contradict the uniform distribution of the velocity over the set $\{-1, 1\}^d$. Moreover, all the velocity components must be independent. This motivates a diagnostic tool that detects if the amount of time each velocity component was equal to -1 or $+1$ is statistically dissimilar. Let V_t^i denote the i-th velocity component at time t and introduce the following events

$$B_i := \{(v_1, \cdots, v_d) \in \mathcal{V} : v_i = +1\} \quad (2.28)$$

for $i = 1, \cdots, d$. In order to assess the convergence of the velocity component, we introduce the following statistic

$$C_T^1 = \left((v_T(B_1), \cdots, v_T(B_d)) - \iota/2\right)^\top \hat{\Sigma}_{B,T}^{-1} \left((v_T(B_1), \cdots, v_T(B_d)) - \iota/2\right), \tag{2.29}$$

where

$$\nu_T(B_j) = \frac{1}{T} \int_0^T \mathbb{1}_{\{V_t^j = +1\}} dt, \text{ for } j = 1, \cdots, d,$$

$\hat{\Sigma}_{B,T}$ denotes a consistent estimator of the asymptotic covariance matrix of $(v_T(B_1), \cdots, v_T(B_d))$, and ι denotes a d-dimensional vector of ones.

Theorem 2.3.2. *Let $Z_t = (X_t, V_t)$ be a stationary d-dimensional Zig-Zag process such that the CLT holds and that $\hat{\Sigma}_{B,T}$ is a consistent estimator of the asymptotic variance, given in (2.27) with $C = B_1, \cdots, B_d$, then*

$$C_T^1 \xrightarrow{w} \chi^2(d)$$

Proof. The proof follows immediately from the central limit theorem for a vector of empirical probabilities as given in (2.27) with $C_i = \mathfrak{X} \times B_i$ for $i = 1, \cdots, d$, and the continuous mapping theorem. □

Similarly, the velocity of the Bouncy Particle Sampler has a multivariate normal distribution under the null hypothesis. Multiple approaches for testing the Gaussianity of stationary data are available. In some specifications of the BPS, the velocity component has a uniform distribution over an ellipsoid. In this situation, we can construct a moment-based test statistic, which consists of a vector of sample moments that characterise a uniform distribution on an ellipsoid.

Testing for independence of position and velocity

The null hypothesis of stationarity asserts that the position and velocity component of both the Zig-Zag sampler and the Bouncy Particle sampler are independent. This result suggests that a statistic that can measure the dependence between the position and velocity can give us a reasonable indication if the process is not exploring the state space according to its stationary distribution. If diagnostic test C_T^1 is unable to reject the stationary distribution of the velocity component, a natural next step would be to consider

the distribution of the velocity given the position of the process. This can be done by selecting a partition $\{A_j\}_{j=1}^{\ell}$ of the position-space and considering the distribution of the velocity for each subset of the position. This is, in fact, a Pearson Chi-square test applied to stationary Markovian data. Therefore, we introduce the test statistic

$$C_T^2 = (\mu_T(A \times B) - \pi_T(A)/2)^\top \hat{\Sigma}_{C,T}^{-1}(\mu_T(A \times B) - \pi_T(A)/2),$$

where $\mu_T(A \times B)$ denotes a vector of all pairs $\mu_T(A_i \times B_j)$ for $i = 1, \cdots, \ell$ and $j = 1, \cdots, 2d$. Here $\{A_j\}_{j=1}^{\ell}$ denotes a partition of the position-space and $\{B_i\}_{i=1}^{2d}$ are defined in (2.28), with adding the complementary events where the velocity is equal to -1 to our vector of empirical probabilities.

Theorem 2.3.3. *Let $Z_t = (X_t, V_t)$ be a stationary d-dimensional Zig-Zag process such that the CLT holds and let $\hat{\Sigma}_{C,T}$ be a consistent estimator estimator of the asymptotic variance of $\mu_T(A \times B)$ given in (2.27) then*

$$C_T^2 \xrightarrow{w} \chi^2(2d(\ell - 1))$$

Proof. By the central limit theorem for a vector of empirical probabilities as given in (2.27) we have that

$$(\mu_T(C) - \mu(C))^\top \Sigma_2^{-1}(\mu_T(C) - \mu(C)) \xrightarrow{w} \chi^2(2d\ell)$$

with $C_i = A_\ell \times B_i$ for $i = 1, \cdots, 2d$ and $i = 1, \cdots, \ell$. By a standard projection argument as stated in Van der Vaart [149; Lemma 16.6] and the continuous mapping theorem, the claim follows. □

Note that under the null hypothesis, the distribution of the velocity is fully specified and does not have to be estimated. Therefore less degrees of freedom are lost in the asymptotic distribution of the Pearson statistic. Unfortunately, for the implementation of both considered test statistics, we need to estimate an asymptotic covariance matrix, which will result in a loss of statistical power.

Remark 2.3.4. For the Pearson statistic C_T^2, we have not shown that it is allowed to choose the partition of the position variable according to simulation output. Presumably, an argument along the lines of Van der Vaart [149; Theorem 17.9] could validate this procedure. △

Chapter 3

Gaussian Approximations for Continuous-time Processes

While Gaussian approximation results have many applications, results for continuous-time settings have been limited. In this chapter, we obtain strong invariance principles for a broad class of ergodic Markov processes. As an application, we show how many useful results regarding the estimation of the asymptotic variance of Markov chain simulation output carry over to PDMP-based methods.

3.1. Introduction

Let $S = (S_t)_{t\geq 0}$ denote a stochastic process and let $\psi = (\psi_t)_{t\geq 0}$ be a positive sequence, we remind the reader that we write

$$S_T = o_{a.s.}(\psi_T) \qquad \text{and} \quad S_T = \mathcal{O}_{a.s.}(\psi_T)$$

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. 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 $S_n = \sum_{k=1}^{n} X_k$. Without loss of generality, suppose that the sequence has zero mean and unit variance. The Komlós-Major-Tusnády approximation Komlós et al. [89]; Komlos [90] 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

$$S_n = W(n) + o_{a.s.}(n^{1/p}) \tag{3.1}$$

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

$$S_n = W(n) + \mathcal{O}_{a.s.}(\log n) \tag{3.2}$$

Furthermore, if only existence of the the second moment is assumed, Major [101] showed that there exists a sequence $t_n \sim n$ such that

$$S_n = W(t_n) + o_{a.s.}(n^{1/2}) \tag{3.3}$$

The error terms appearing in the strong invariance principles (3.1), (3.2), and (3.3) are optimal. Given their broad range of applications, 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 Berkes et al. [13]. In Markovian settings, strong approximation results were obtained by Cuny et al. [42], Csáki and Csörgő [37], Vats et al. [151], and Merlevède et al. [103], among others. The strong invariance principle of Merlevède et al. [103] attains the Komlós-Major-Tusnády bound given in (3.2). The results of Csáki and Csörgő [37] and Merlevède et al. [103] are established through an application of Nummelin splitting, introduced in the seminal papers of Athreya and Ney [6] and Nummelin [115]. 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 Damerdji [44; 45], Flegal and Jones [66], and Vats et al. [151] show strong consistency of the batch means and spectral variance estimators for MCMC simulation output using the appropriate strong invariance principles.

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, Löcherbach and Loukianova [98] 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 Berkes et al. [12] to obtain a strong invariance principle attaining a convergence rate of order $\mathcal{O}_{a.s.}(T^{1/4}\log T)$. This result is formulated in Theorem 3.3.5 and covers a wide range of Markov processes including ergodic diffusions. Although the nearly optimal bound $\mathcal{O}_{a.s.}(T^{1/p}\log(T)^2)$ of Berkes et al. [12] 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 3.3.5.

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 Merlevède et al. [103] such that the optimal strong approximation error of $\mathcal{O}_{a.s.}(T^{1/p})$ 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 Merlevède et al. [103] can be extended under less restrictive conditions such that the optimal approximation error (3.2) 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 3.4.2 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 3.3.6 and 3.3.7 obtain the optimal approximation rate of $\mathcal{O}_{a.s.}(T^{1/p})$ 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 3.4.3. Furthermore, we demonstrate the ap-

plicability of our results to diffusion processes and show that the magnitude of increments can be described with our obtained approximation results.

This chapter is organised as follows. In Section 3.2, we review Nummelin splitting in continuous time as introduced in Löcherbach and Loukianova [98] and discuss other relevant results. In Section 3.3, the main results of this chapter are given. In Section 3.4, we discuss the estimation of the asymptotic variance for PDMC simulation output. Section 3.5 illustrates the applicability of our results to diffusion processes. In Section 3.6, the proofs of the main results are given.

3.2. 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

$$\pi(A) > 0 \implies \mathbb{P}_x\left(\int_0^\infty \mathbb{1}_{\{X_s\in A\}} ds = \infty\right) = 1, \quad x \in E. \tag{3.4}$$

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 thne 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 Höpfner and Löcherbach [82; Thereom 1.4]. The transition kernel of the process $\bar{X} = (\bar{X}_n)_{n\in\mathbb{N}}$ is given by

$$U(x, A) = \int_0^\infty P_t(x, A)e^{-t}dt, \tag{3.5}$$

and satisfies the one-step minorisation condition, see for example Höpfner and Löcherbach [82] or Revuz [130],

$$U(x, A) \geq h \otimes \nu(x, A), \tag{3.6}$$

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

$$K((x,u),dy) = \begin{cases} \nu(dy) & \text{for } (x,u) \in C \times [0,\alpha] \\ W(x,dy) & \text{for } (x,u) \in C \times (\alpha,1] \\ U(x,dy) & \text{for } x \notin C \end{cases}, \tag{3.7}$$

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

$$W(x,dy) = \frac{U(x,dy) - \alpha\nu(dy)}{1-\alpha}. \tag{3.8}$$

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 Löcherbach and Loukianova [98]. 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, Löcherbach and Loukianova [98] 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

$$\left(\frac{p_t(x, x')}{u(x, x')}\mathbb{1}_{\{0<u(x,x')<\infty\}} + \mathbb{1}_{\{u(x,x')\in\{0,\infty\}\}}\right) e^{-t}dt \quad \text{on} \quad \mathbb{R}_+. \tag{3.9}$$

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

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

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

$$Z_{T_n+s}^1 \sim \frac{p_s(x, y)p_{t-s}(y, x')}{p_t(x, x')}\mathbb{1}_{\{p_t(x,x')>0\}}\Lambda(dy). \tag{3.10}$$

Let $Z_{T_n+s}^1 := x_0$ for some fixed $x_0 \in E$ on $\{p_t(x, x') = 0\}$. Moreover, given $Z_{T_n+s}^1 = y$ on $s + u < t$ we have that

$$Z_{T_n+s+u}^1 \sim \frac{p_u(y, y')p_{t-s-u}(y', x')}{p_{t-s}(y, x')}\mathbb{1}_{\{p_{t-s}(y,x')>0\}}\Lambda(dy'). \tag{3.11}$$

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

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

$$Z_{T_{n+1}}^3 \sim K((x', u'), dx''). \tag{3.12}$$

Note that in the construction of Z the inter-sampling times $(\sigma_n)_{n\geq 1}$ are drawn according to (3.9), their conditional distribution given the starting and endpoint of the sampled chain. Equation (3.10) and (3.11), 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 Löcherbach and Loukianova [98; 99] that we work with are given in the following propositions. Firstly, the first coordinate of Z has the desired distribution.

Proposition 3.2.1 (Löcherbach and Loukianova [98; Proposition 2.8])**.** *The constructed process Z from the simulation scheme given in (3.9)-(3.12) 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 3.2.2 (Löcherbach and Loukianova [99; Proposition 4.2])**.** *Let (S_n, R_n) be a sequence of stopping times defined as $S_0 = R_0 := 0$ and*

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

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

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

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 3.2.3 (Löcherbach and Loukianova [99; 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 3.2.3 was also noted by Sigman [140]. 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 Löcherbach and Loukianova [98] 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 3.2.4 (Sigman [140; Theorem 2])**.** *Let X be a positive recurrent one-dependent regenerative process, then we can characterise its stationary measure as follows*

$$\pi(A) = \frac{1}{\varrho}\ \mathbb{E}_\nu \int_0^{R_1} \mathbb{1}_{\{X_s \in A\}}\ ds, \tag{3.13}$$

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

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T f(X_s)ds = \frac{1}{\varrho}\ \mathbb{E}_\nu \int_0^{R_1} f(X_s)ds \quad , \tag{3.14}$$

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

Note that the normalisation constant $1/\varrho$ given in Proposition 3.2.4 is finite and non-zero due to the positive Harris recurrence of the process.

Remark 3.2.5. The framework of Löcherbach and Loukianova [98; 99] does not require ergodicity. Moreover, it is important to note that contrary to the classically regenerative case, Proposition 3.2.4 does not imply convergence in total variation to the stationary measure. For a counterexample see Sigman [140; Remark 3.2]. △

For our applications, we will require ergodicity and hence we must additionally impose this as stated in (2.14). These ergodicity requirements are usually established through Foster–Lyapunov drift conditions; see Down et al. [54] and Fort and Roberts [67] for exponential and polynomial ergodicity respectively. These results have been applied to several classes of diffusion processes, see for example Cattiaux et al. [31; Theorem 8.3 and 8.4] and Stramer and Tweedie [145; Theorem 3.1 and 4.1].

For PDMPs, Bierkens et al. [22] 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 Deligiannidis et al. [50] and Durmus et al. [59] conditions for exponential ergodicity of the Bouncy Particle Sampler are given. Utilising hypocoercivity techniques, Andrieu et al. [4] 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.

3.3. Main Theorems

The most straightforward approach for obtaining a strong approximation result for Markov processes would be through ergodicity requirements. In Kuelbs and Philipp [91] it is shown that a multivariate strong invariance principle holds for sums of random vectors satisfying a strong mixing condition; see also Theorem 3.6.1. This mixing condition is satisfied when one has an appropriate rate of ergodicity of the process. All proofs are provided in Section 3.6.

Theorem 3.3.1. *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*

$$\left\| \int_0^T f(X_t)\, dt - T\pi(f) - \Sigma_f^{1/2} W(T) \right\| = \mathcal{O}_{a.s.}(\psi_T) \tag{3.15}$$

with

$$\psi_T = T^{1/2-\min(\delta/(2\delta+4),\ \lambda)} \text{ for some } \lambda \in (0, 1/2), \tag{3.16}$$

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

$$\Sigma_f = \int_0^\infty \mathrm{Cov}_\pi \left(f(X_0), f(X_s)\right)\ ds + \int_0^\infty \mathrm{Cov}_\pi(f(X_s), f(X_0))\ ds, \tag{3.17}$$

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

Remark 3.3.2. The asymptotic covariance matrix Σ_f given in Theorem 3.3.1 cannot be simplified. Only for the univariate case ($p = 1$) and for reversible processes do we obtain that

$$\Sigma_f = 2 \int_0^\infty \mathrm{Cov}_\pi(f(X_0), f(X_s))\ ds. \tag{3.18}$$

As a result of the reversibility, the cross-covariance matrices in (3.17) will be symmetric and thus the asymptotic covariance can be expressed as (3.18). △

The rate ψ_T appearing in Theorem 3.3.1 will depend on the dependence and moment structure of the considered process. We see that if higher-order moments exist, i.e., δ becomes larger, then ψ_T becomes smaller. Moreover, Kuelbs and Philipp [91] state that λ will be influenced by the decay of dependence within the process. We expect the approximation bound ψ_T to be smaller for processes with faster decaying auto-dependence. 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 (3.15) 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 Löcherbach and Loukianova [98] 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 Berkes et al. [12] to obtain a strong invariance principle; see also Theorem 3.6.5.

Proposition 3.3.3. *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 3.2.2 and 3.2.3. Moreover, assume that*

$$\mathbb{E}_\nu[R_1^q] < \infty \quad \text{for some} \quad q > 2, \tag{3.19}$$

$$\mathbb{E}_\nu \left| \int_0^{R_1} f(X_s) ds \right|^p < \infty \quad \text{for some} \quad p > 2. \tag{3.20}$$

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

$$\left| \int_0^T f(X_s) ds - T\pi(f) - W_1(\sigma_T^2) - W_2(\tau_T^2) \right| = \mathcal{O}_{a.s.}(\psi_T) \tag{3.21}$$

where $\{\sigma_T^2\}$ *and* $\{\tau_T^2\}$ *are non-decreasing sequences with* $\sigma_T^2 = \frac{\sigma_\xi^2}{\varrho} T + \mathcal{O}_{a.s.}\left(\frac{T}{\log T}\right)$, $\tau_T^2 = \mathcal{O}_{a.s.}\left(\frac{T}{\log T}\right)$ *as* $T \to \infty$, *and* $\psi_T, \pi(f), \varrho$, *and* σ_ξ *are defined in equations (3.23) to (3.26) below.*

In Proposition 3.3.3 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 3.3.3 stated in (3.19) and (3.20) 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 Mykland et al. [112] and the strong approximation result of Csáki and Csörgő [37]. The results of Hobert et al. [81] 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 Jones et al. [87] and Bednorz and Łatuszyński [10]; hence simplifying the required conditions of Csáki and Csörgő [37]. 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 3.3.3, we show that the cycle moment conditions (3.19) and (3.20) required for Proposition 3.3.3 can also be guaranteed with more easily verifiable ergodicity and moment conditions.

Theorem 3.3.4. *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 3.3.3 and Theorem 3.3.4 we obtain the desired strong invariance principle.

Theorem 3.3.5. *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*

$$\left| \int_0^T f(X_s)ds - T\pi(f) - W_1(\sigma_T^2) - W_2(\tau_T^2) \right| = \mathcal{O}_{a.s.}(\psi_T) \tag{3.22}$$

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

$$\psi_T = \max\left\{T^{1/4}\log T, T^{1/p}\log^2(T)\right\}, \tag{3.23}$$

$$\pi(f) = \frac{1}{\varrho}\,\mathbb{E}_\nu \int_0^{R_1} f(X_s)\; ds, \tag{3.24}$$

$$\varrho = \mathbb{E}_\nu[R_1], \;\; \textit{and} \tag{3.25}$$

$$\sigma_\xi = \sqrt{\operatorname*{Var}_\nu(\xi_1) + 2\operatorname*{Cov}_\nu(\xi_1, \xi_2)}\;. \tag{3.26}$$

Proof. The assertion follows immediately from Proposition 3.3.3 and Theorem 3.3.4. □

The appearance of the second Brownian motion in Theorem 3.3.5 is inherited from the strong invariance principle of Berkes et al. [12]. Although we obtain different time perturbations of the Brownian motions, all desired properties carry over. The second Brownian motion appearing in (3.22) 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

$$\operatorname{Corr}\left(W_1(\sigma_t^2), W_2(\tau_s^2)\right) \to 0, \quad \text{as } t, s \to \infty. \tag{3.27}$$

Note that the nearly optimal convergence rate $\mathcal{O}_{a.s.}(T^{1/p}\log^2 T))$ obtained by Berkes et al. [12] does not carry over. Instead, we obtain an approximation error that cannot be improved beyond $\mathcal{O}_{a.s.}(T^{1/4}\log T)$. Obtaining a superior approximation error remains an open problem for the class of

processes considered in Theorem 3.3.5. A possible approach for attaining a better convergence rate would be to extend to results of Berkes et al. [12] to a multivariate setting and then follow the approach of Merlevède et al. [103].

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 Merlevède et al. [103] and obtain the optimal bound of $\mathcal{O}_{a.s.}(T^{1/p})$ for the strong approximation of the one-dimensional Zig-Zag process.

Theorem 3.3.6. *Let $Z = (X_t, V_t)_{t\geq 0}$ be an aperiodic, positive Harris recurrent one-dimensional Zig-zag process with an invariant distribution $\pi \otimes \upsilon$, 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*

$$\left| \int_0^T f(X_s)ds - T\pi(f) - \sigma_f^2 W(T) \right| = \mathcal{O}_{a.s.}(T^{1/p}) \tag{3.28}$$

where σ_f^2 can be characterised as (3.30).

In Merlevède et al. [103] 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 3.3.6 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 3.3.7. *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*

$$\left\| \int_0^T f(X_t)\, dt - T\pi(f) - \Sigma_f^{1/2} W(T) \right\| = \mathcal{O}_{a.s.}(T^{1/p}) \tag{3.29}$$

and covariance matrix $\Sigma_f = \mathrm{diag}\{\sigma_{f_1}^2, \cdots, \sigma_{f_d}^2\}$ with

$$\sigma_{f_i}^2 = \int_0^\infty \mathrm{Cov}_\pi(f_i(X_0^i), f_i(X_s^i))\, ds + \int_0^\infty \mathrm{Cov}_\pi(f_i(X_s^i), f_i(X_0^i))\, ds. \tag{3.30}$$

Note that although the proof of Theorem 3.3.7 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 3.3.6, 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 3.3.8. From Theorem 3.3.4 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 3.6.7 and (3.76), we see that the approximation error of Theorem 3.3.6 can in general taken to be of order $\mathcal{O}_{a.s.}(T^\alpha)$ 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 3.3.5 and 3.3.7.

Furthermore, from Remark 3.6.8, we see that under the assumption of exponential ergodicity, the conclusions of Theorem 3.3.4 and all aforementioned strong invariance principles hold with their stated approximation error. △

Remark 3.3.9. Note that in Theorem 3.3.6, 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 3.3.6 and 3.3.7 will hold. △

3.4. 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 Bierkens and Duncan [19] several conditions are given to obtain a CLT for the univariate Zig-Zag process. In Durmus et al. [59], Deligiannidis et al. [50], and Bierkens et al. [22] 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 3.3.1, 3.3.5, 3.3.6, and 3.3.7 immediately imply the following central limit theorems for polynomially ergodic Markov processes.

Corollary 3.4.1. *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:*

$$\frac{1}{\sqrt{T}} \int_0^T (f(X_s, V_s) - \mu(f))\, ds \xrightarrow{w} \mathcal{N}_p(0, \Sigma_f). \tag{3.31}$$

Additionally, also a functional central limit theorem holds:

$$\left(\frac{1}{\sqrt{n}} \int_0^{nt} (f(X_s, V_s) - \mu(f))\, ds \right)_{t\geq 0} \xrightarrow{w} \Sigma_f^{1/2} W \text{ as } n \to \infty, \tag{3.32}$$

where

$$\Sigma_f = \int_0^\infty \operatorname*{Cov}_\mu(f(X_0, V_0), f(X_s, V_s))\, ds + \int_0^\infty \operatorname*{Cov}_\mu(f(X_s, V_s), f(X_0, V_0))\, ds, \tag{3.33}$$

$W = (W_t)_{t\geq 0}$ 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 Csörgö and Horváth [38; Theorem 1.17], the FCLT immediately follows from the strong invariance principle formulated in Theorem 3.3.1. Similarly, by Damerdji [45; Proposition 2.1] the CLT follows. □

By the same argument the CLT follows for the processes considered in Theorems 3.3.5, 3.3.6, and 3.3.7. 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 3.4.1 also a (functional) central limit theorem follows for the position component of the process. We are interested in estimating the asymptotic variance (3.33); 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;

$$\bar{Z}_i(\ell_T) := \frac{1}{\ell_T} \int_{(i-1)\ell_T}^{i\ell_T} f(X_s)ds, \quad i = 1, \ldots, k_T. \tag{3.34}$$

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 $\mathrm{Var}(\bar{Z}_i(\ell_T))$. 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

$$\hat{\sigma}_T^2 = \frac{\ell_T}{k_T - 1} \sum_{i=1}^{k_T} \left(\bar{Z}_i(\ell_T) - \frac{1}{k_T} \sum_{i=1}^{k_T} \bar{Z}_i(\ell_T) \right)^2. \tag{3.35}$$

Following the framework of Damerdji [45], 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 Glynn and Whitt [72]. 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 3.4.2. *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*

$$\frac{T^{2/p}}{\ell_T} \log(T) \to 0, \quad \text{as } T \to \infty, \tag{3.36}$$

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 3.3.1, Jones et al. [87; Proposition 3], and Damerdji [45; Theorem 3.3]. □

Remark 3.4.3. Note that Theorem 3.4.2 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 3.3.6 and 3.3.7 obtain the optimal approximation rate of $\mathcal{O}_{a.s.}(T^{1/p})$ whereas the results of Jones et al. [87] are based upon the strong invariance principle of Csáki and Csörgő [37] which attains the rate $\mathcal{O}_{a.s.}(T^{\gamma} \log T)$, with $\gamma = \max(1/p, 1/4)$. More specifically, for f with $\pi(|f|^p) < \infty$ Jones et al. [87] requires $T^{\gamma} \log^3(T)/\ell_T \to 0$ as $T \to \infty$. In particular for the case where $p > 4$, Theorem 3.4.2 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 3.4.4. Note that a similar conclusion holds for the overlapping batch means and spectral variance estimators considered in Flegal and Jones [66]. △

We see from the required assumption (3.36) 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 3.4.4. *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 (3.112) below. Then we have that*

$$\mathbb{E}_\mu\left|\hat{\sigma}_T^2 - \sigma_f^2\right|^2 = 2\sigma_f^4 \frac{\ell_T}{T} + \mathcal{O}_{a.s.}\left(\frac{T^{1/p}}{\sqrt{T}} \log^{\frac{1}{2}} T\right) + \mathcal{O}_{a.s.}\left(\ell_T^{-1} T^{2/p} \log T\right), \quad (3.37)$$

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

$$\sqrt{k_T}(\hat{\sigma}_T^2 - \sigma_f^2) \xrightarrow{w} \mathcal{N}(0, 2\sigma_f^4) \text{ as } T \to \infty. \quad (3.38)$$

Proof. By the imposed conditions of the process, the strong invariance principle formulated in Theorem 3.3.1 holds. The first claim then follows by Damerdji [46; Theorem 1 and Lemma 3] and the second by Damerdji [46; Proposition 2]. □

The first and second term in (3.37) 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

$$\hat{\Sigma}_T = \frac{\ell_T}{k_T - 1} \sum_{i=1}^{k_T} \left(\bar{Z}_i(\ell_T) - \frac{1}{k_T}\sum_{i=1}^{k_T} \bar{Z}_i(\ell_T)\right)\left(\bar{Z}_i(\ell_T) - \frac{1}{k_T}\sum_{i=1}^{k_T} \bar{Z}_i(\ell_T)\right)^T, \quad (3.39)$$

where $\bar{Z}_i(\ell_T)$ is defined in (3.34). Given the strong invariance principle of Theorem 3.3.1, the results of Vats et al. [152] for the multivariate batch means estimator immediately carry over.

Theorem 3.4.5. *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*

$$\frac{\psi_T^2}{\ell_T} \log(T) \to 0, \quad \text{as } T \to \infty, \tag{3.40}$$

with ψ_T defined in (3.16), 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 3.3.1 and Vats et al. [152; Theorem 2]. □

Furthermore, if the target distribution is of product form and we consider the Zig-Zag Sampler, then Theorem 3.3.7 gives a strong invariance principle with an explicit approximation error. Therefore, we can replace condition (3.40) of Theorem 3.4.5 with (3.36) for every component of the Zig-Zag process. This results in a condition that can more easily be verified.

3.4.1. Discussion

Batch size selection for PDMC

In Glynn and Whitt [72] 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 3.4.4 suggests that $\alpha^* = (2+p)/2p$ would be optimal in the mean squared error sense. The well-known results of Chien et al. [34], Goldsman et al. [74], and Song and Schmeiser [142] obtain a bound for the bias of order $\mathcal{O}_{a.s.}(\ell_T^{-1})$, 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 $\mathcal{O}_{a.s.}(\ell_T^{-1})$ for batch means under milder conditions remains an unaddressed problem. Theorem 3.4.2 and 3.4.4 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 3.4.4 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 autocovariance 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 Flegal and Jones [66]. 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 Adler et al. [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 Chien [35], 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.

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 Sherman and Goldsman [138] 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 3.4.4 places more restrictive conditions on the batch size and excludes the choice $\ell_T^\diamond \asymp T^{1/3}$. In Chakraborty et al. [32] 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 $k_T = o_{a.s.}(\ell_T^2)$. Hence their result is also unable to guarantee asymptotic normality for batch size $\ell_T^\diamond$. We see that Theorem 3.4.4 gives more practical conditions for guaranteeing asymptotic normality of the batch-means estimator, in particular, the results are applicable to non-reversible processes.

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 3.3.1, many results for other estimators of the asymptotic variance also carry over. In Flegal and Jones [66] more convenient alternatives are given for some of the requirements of the framework given in Damerdji [44]. The results of Flegal and Jones [66] 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 Flegal and Jones [66] are only imposed such that the strong invariance principle of Csáki and Csörgő [37] 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 Vats et al. [151] and Liu et al. [97] 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.

Regenerative Simulation

From the proof of Theorem 3.3.6, we see that the univariate Zig-Zag sampler possesses a recurrent atom. Moreover, any local optimum with an appropri-

ate 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 Brockwell and Kadane [29] or Hobert et al. [81]. 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 Brockwell and Kadane [29], 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 Flegal and Jones [66] and Gilks et al. [70], 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 Monmarché [110]. 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.

3.5. 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 3.3.5 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 3.5.1 (Csörgö and Révész [39; 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*

$$\limsup_{T\to\infty} \sup_{0\leq t\leq T-a_T} \sup_{0\leq u\leq a_T} \beta_T |W_{t+u} - W_t| = 1 \quad , \tag{3.41}$$

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 Csörgö and Révész [40; 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 Li [96], El-Nouty [62] and Ortega [120] 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 Révész [129] and Deheuvels [49]. 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 Csörgö and Révész [40; 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 3.3.5. 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 Berkes et al. [12] it is shown that the magnitude of the increments of partial sums of weakly m-dependent sequences are indeed given by Theorem 3.5.1, 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 Berkes et al. [12; Theorem 4].

Theorem 3.5.2. *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 $\beta_T \psi_T = o_{a.s.}(1)$, 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

$$\limsup_{T\to\infty} \sup_{0\leq t\leq T-a_T} \sup_{0\leq u\leq a_T} \beta_T \left| \int_t^{t+u} f(X_s) ds \right| \leq \frac{\sigma_\xi^2}{\varrho} \tag{3.42}$$

As noted by Berkes et al. [12], the split invariance principle also implies the distributional version of Theorem 3.5.2; with similar adaptations to their argument this would also hold in our case. Since the approximation error ψ_T of Theorem 3.3.5 cannot be guaranteed to be smaller than $\mathcal{O}_{a.s.}(T^{1/4} \log T)$, the fluctuation result given in Theorem 3.5.2 cannot describe the magnitude of increments over slowly growing time intervals a_T.

3.5.1. Application to diffusion processes

Diffusions are an important class of processes for which the strong approximation given in Theorem 3.3.5 and the related fluctuation result given in Theorem 3.5.2 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)

$$\begin{cases} dX_t = b(X_t)dt + \sigma(X_t)dW_t \\ X_0 \sim \mu, \end{cases} \tag{3.43}$$

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 Rogers and Williams [135].

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

$$s(u) = \int_{x_0}^{u} \exp\left[-2\int_{x_0}^{z} \frac{b(y)}{\sigma^2(y)}dy\right] dz \text{ and must satisfy } \lim_{u\to\pm\infty} s(u) = \pm\infty. \tag{3.44}$$

If condition (3.44) 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 $m(u) = (s(u)\sigma^2(u))^{-1}$, must be Lebesque integrable for the diffusion to be positive Harris recurrent. For higher-dimensional diffusion processes Bhattacharya [18] gives conditions that guarantee positive Harris recurrence. The results of Lazić and Sandrić [92; 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 Stramer and Tweedie [144; 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 3.3.1 and Theorem 3.3.5 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 Cattiaux et al. [31; Theorem 8.3 and 8.4] and Stramer and Tweedie [145; Theorem 3.1 and 4.1].

In order for the strong approximation result in Theorem 3.3.5 and the related fluctuation result of Theorem 3.5.2 to hold, the Nummelin splitting scheme of Löcherbach and Loukianova [98] 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 Williams [155; 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; Stroock and Varadhan [146; 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 Rogers and Williams [135; Theorem 38.16].

3.5.2. Discussion and suggestions for further research

We see that Theorem 3.3.1 and Theorem 3.3.5 are applicable to a broad class of diffusions and extend the current results on strong approximations for diffusion processes. In Heunis [79] and Mihalache [109] strong invariance principles are obtained for diffusions and a complementary fluctuation result and change point test respectively. The results of Mihalache [109] yield an explicit approximation error comparable to that of Theorem 3.3.5, but are only applicable to stochastic integrals with respect to Brownian motion, i.e., diffusion processes with no drift. The results of Heunis [79] give an implicit approximation error and hold for singular diffusions. The strong invariance principle of Heunis [79] 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 Csörgö and Hall [41] and their given references. Following the approach of Berkes et al. [12; Proposition 2], Theorem 3.3.5 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.

3.6. Proofs

3.6.1. Theorem 3.3.1

In Kuelbs and Philipp [91] 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 Bradley [28; page 118].

Theorem 3.6.1 (Kuelbs and Philipp [91; 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*

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

such that

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

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 3.3.1.

Lemma 3.6.2 (Douc et al. [53; 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 (2.14).*

Proof.

$$\begin{aligned}
\alpha_Z(s) &= \sup_{\substack{F \in \sigma(Z_0) \\ G \in \sigma(Z_s)}} |\mathbb{P}_\mu(F \cap G) - \mathbb{P}_\mu(F)\mathbb{P}_\mu(G)| \\
&= \sup_{\substack{A \in \mathscr{E} \\ B \in \mathscr{E}}} |\mathbb{P}_\mu(Z_0 \in A; Z_s \in B) - \mathbb{P}_\mu(Z_0 \in A)\mathbb{P}_\mu(Z_s \in B)| \\
&= \sup_{\substack{A \in \mathscr{E} \\ B \in \mathscr{E}}} \left| \int_A P_s(z, B)\mu(dz) - \mu(A)\mu(B) \right| \\
&\leq \sup_{\substack{A \in \mathscr{E} \\ B \in \mathscr{E}}} \int_A \big|P_s(z, B) - \mu(B)\big|\mu(dz) \\
&\leq \int_E \sup_{B \in \mathscr{E}} \big|P_s(z, B) - \mu(B)\big|\mu(dz) \\
&= \int_E \|P_s(z, \cdot) - \mu(\cdot)\|_{TV}\mu(dz) \\
&\leq e^{-\beta s} \int_E V(z)\mu(dz) = e^{-\beta s}\mu(V)
\end{aligned}$$

□

Lemma 3.6.3 (Davydov [47] and Rio [132])**.** *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 3.6.4 (Piterbarg and Rodionov [124; 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 Kuelbs and Philipp [91] also holds for continuous-time ergodic processes.

Proof of Theorem 3.3.1

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., $Y_t = (Y_{1t}, \cdots, Y_{dt})^\top$ and therefore also each ξ_k is a d-dimensional vector, $\xi_k = (\xi_{1k}, \cdots, \xi_{dk})^\top$. 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 3.6.2, 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 $\sigma\left(f(X_t)\right) \subseteq \sigma\left(X_t\right)$ and $\sigma\left(\xi_k\right) \subseteq \sigma\left(X_s : k-1 \leq s \leq k\right)$. 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 3.6.1, 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

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

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

$$\left\| \sum_{k=1}^{n} \xi_k - \int_0^T Y_t dt \right\| = \mathcal{O}_{a.s.}(T^{1/p+\varepsilon}) \text{ a.s. for } T \to \infty, \tag{3.45}$$

$$\|W_T - W_n\| = o_{a.s.}(T^{1/p+\varepsilon}) \text{ a.s. for } T \to \infty, \text{ and} \tag{3.46}$$

$$\Sigma_f = \Sigma_\xi. \tag{3.47}$$

In order to show that (3.45) holds, we note that

$$\left\| \int_0^T Y_s ds - \sum_{k=1}^{n} \xi_k \right\| = \left\| \int_n^T Y_s ds \right\| \leq \int_n^{n+1} \|Y_s\| ds. \tag{3.48}$$

By a Borel-Cantelli argument, it will follow that

$$\int_n^{n+1} \|Y_s\| \, ds = \mathcal{O}_{a.s.}(n^{1/p+\varepsilon}) = \mathcal{O}_{a.s.}(T^{1/p+\varepsilon}) \quad \text{a.s. for } T \to \infty. \tag{3.49}$$

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

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

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 (4.97). A similar Borel-Cantelli argument also shows that (3.46) 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

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

Let W denote a $d-$dimensional Brownian motion and let $\|W_t\|$ denote the corresponding Bessel process, then we have by Lemma 3.6.4 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

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

By a Borel-Cantelli argument we see that

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

Therefore the term (3.46) will be asymptotically negligible. Finally, we see that by Lemma 3.6.3 the asserted asymptotic variance Σ_f is finite, i.e., all entries

$$(\Sigma_f)_{ij} = \int_0^\infty \mathrm{Cov}_\pi(f_i(X_0), f_j(X_s))\,ds + \int_0^\infty \mathrm{Cov}_\pi(f_i(X_s), f_j(X_0))\,ds, \quad (3.50)$$

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 3.6.3 gives us

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

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 (3.50) 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

$$\lim_{T\to\infty} \operatorname*{Var}_{\pi}\left(\frac{1}{\sqrt{T}}\int_0^T Y_t dt\right) = \lim_{T\to\infty} \operatorname*{Var}_{\pi}\left(\frac{1}{\sqrt{T}}\left(\int_0^n Y_t dt + \int_n^T Y_t dt\right)\right) =$$
$$\lim_{T\to\infty} \frac{1}{T}\operatorname*{Var}_{\pi}\left(\int_0^n Y_t dt\right) + \lim_{T\to\infty} \frac{1}{T}\operatorname*{Var}_{\pi}\left(\int_n^T Y_t dt\right) + \tag{3.51}$$
$$\lim_{T\to\infty} \frac{1}{T}\operatorname{Cov}_{\pi}\left(\int_0^n Y_t dt, \int_n^T Y_t dt\right) + \lim_{T\to\infty} \frac{1}{T}\operatorname{Cov}_{\pi}\left(\int_n^T Y_t dt, \int_0^n Y_t dt,\right).$$

Let $\Sigma_{T1}, \Sigma_{T2}, \Sigma_{T3}$ and Σ_{T4} denote the four terms in (3.51). 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 (3.51):

$$(\Sigma_{T1})_{ij} = \frac{1}{T}\int_0^n \int_0^n \operatorname{Cov}_{\pi}(Y_{it}, Y_{js})\ dtds, \tag{3.52}$$
$$(\Sigma_{T2})_{ij} = \frac{1}{T}\int_n^T \int_n^T \operatorname{Cov}_{\pi}(Y_{it}, Y_{js})\ dtds, \tag{3.53}$$
$$(\Sigma_{T3})_{ij} = \frac{1}{T}\int_0^n \int_n^T \operatorname{Cov}_{\pi}(Y_{it}, Y_{js})\ dtds, \tag{3.54}$$
$$(\Sigma_{T4})_{ij} = \frac{1}{T}\int_0^n \int_n^T \operatorname{Cov}_{\pi}(Y_{is}, Y_{jt})\ dtds. \tag{3.55}$$

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

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

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 3.6.2 and Lemma 3.6.3 gives us that

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

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

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

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 Meyn and Tweedie [107; Proposition 17.1.6] it follows that the strong invariance principle holds for every initial distribution. Let

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

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., $h(x) = P_s h(x)$. Indeed, for every x in E and $s \geq 0$ we have

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

By the Markov property and the tower property of conditional expectation, we have that $P_s h(x)$ can be expressed as

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

where the last inequality follows since for all fixed $s \geq 0$ we have, by the same argument of (4.97), 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 $\mathcal{O}_{a.s.}(\psi_T)$ almost surely. By Kallenberg [88; 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

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

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

3.6.2. Proposition 3.3.3

In Berkes et al. [12] 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 Berkes et al. [12]). Their strong invariance principle, stated in Theorem 3.6.5, 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 Berkes et al. [12].

Theorem 3.6.5 (Berkes et al. [12; 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 m-dependent approximation rate function κ, i.e.,*

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

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

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

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

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

$$t_n^2 = \mathcal{O}(n/\log n),$$

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

The m-dependent approximation rate function κ appearing in Theorem 3.6.5 describes how well the process can be approximated by an m–dependent

process in terms of L^p–distance. For a more detailed definition, we refer to Berkes et al. [12; Definition 1]. As noted by Berkes et al. [12], the perturbed time sequences $\{s_n\}$ and $\{t_n\}$ are deterministic and can be explicitly calculated.

Proof of Proposition 3.3.3

Proof. Firstly, assume that the initial distribution of X is equal to ν. By Proposition 3.2.1 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 3.2.2, 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

$$\xi_n := \int_{R_{n-1}}^{R_n} \{f(X_s) - \pi(f)\}\, ds, \quad n \geq 1, \tag{3.56}$$

is stationary under $\mathbb{P}_\nu$. Moreover, by Proposition 3.2.3 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

$$\left| \int_0^T \{f(X_s) - \pi(f)\}\, ds - \sum_{k=1}^{N(T)} \xi_k \right| \leq \int_{R_{N(T)}}^T |f(X_s) - \pi(f)| ds. \tag{3.57}$$

By an argument analogous to the one given in Theorem 3.3.1 for the remainder term defined in (3.48), we will show that

$$\int_{R_{N(T)}}^{T} |f(X_s) - \pi(f)|ds = \mathcal{O}_{a.s.}(T^{1/p}) \quad \text{a.s. for } T \to \infty. \tag{3.58}$$

In order to show that (3.58) holds, we note that

$$\int_{R_{N(T)}}^{T} |f(X_s) - \pi(f)|ds \leq \int_{R_{N(T)}}^{R_{N(T)+1}} |f(X_s) - \pi(f)|ds. \tag{3.59}$$

By a Borel-Cantelli argument it will follow that

$$\int_{R_{N(T)}}^{R_{N(T)+1}} |f(X_s) - \pi(f)|ds = \mathcal{O}_{a.s.}(T^{1/p}) \quad \text{a.s. for } T \to \infty. \tag{3.60}$$

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

$$\int_{R_n}^{R_{n+1}} |f(X_s) - \pi(f)|ds = \mathcal{O}_{a.s.}(n^{1/p}) \quad . \tag{3.61}$$

Moreover, since $N(T)$ is almost surely increasing and $N(T) = \mathcal{O}_{a.s.}(T)$, as shown in (3.65), it follows that

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

Hence proving the claim formulated in (3.60) and as a direct consequence also the bound stated in (3.58). Furthermore, by Proposition 3.2.3, 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 Berkes et al. [12; 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 3.6.5, 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

$$\left| \sum_{k=1}^{n} \xi_k - n\mathbb{E}_\nu \xi_1 - W_1(s_n^2) - W_2(t_n^2) \right| = \mathcal{O}_{a.s.}(n^{1/p} \log^2 n) \tag{3.62}$$

where $\{s_n\}$ and $\{t_n\}$ are increasing deterministic sequences with $s_n^2 = \sigma_\xi^2 n + \mathcal{O}_{a.s.}(n/\log n)$ and $t_n^2 = \mathcal{O}_{a.s.}(n/\log n)$. Note that by Proposition 3.2.4 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 (3.62), 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

$$\frac{1}{n^{1/p} \log^2 n} \left| \sum_{k=1}^{n} \xi_k(\omega) - W_1(s_n^2, \omega) - W_2(t_n^2, \omega) \right| < C(\omega) \tag{3.63}$$

Since we have that $\mathbb{E}_\nu R_1^q < \infty$, by Csörgö and Horváth [38; Theorem 2.4] with $q = \beta - 1$, we can construct a Brownian motion $\tilde{W}$ such that

$$\left| N(T) - \frac{T}{\varrho} - \frac{\mathrm{Var}_\nu(R_1)}{\varrho^{3/2}} \tilde{W}_T \right| = o_{a.s.}(T^{1/q}). \tag{3.64}$$

By the law of iterated logarithm for Brownian motion we obtain

$$N(T) = \frac{T}{\varrho} + \mathcal{O}_{a.s.}(\sqrt{T \log\log T}) \quad \text{a.s.} \tag{3.65}$$

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 (3.63) that

$$\limsup_{T\to\infty} \frac{\left|\sum_{k=1}^{N(T)} \xi_k - W_1(s^2_{N(T)}) - W_2(t^2_{N(T)})\right|}{N(T)^{1/p}\log^2(N(T))} < C \quad \text{a.s.} \,, \tag{3.66}$$

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

$$\begin{aligned} s^2_{N(T)} &= \sigma^2_\xi N(T) + \mathcal{O}_{a.s.}(N(T)/\log N(T)) \\ t^2_{N(T)} &= \mathcal{O}_{a.s.}(N(T)/\log N(T)). \end{aligned}$$

We see that (3.66) can be reformulated as

$$\left|\sum_{k=1}^{N(T)} \xi_k - W_1(s^2_{N(T)}) - W_2(t^2_{N(T)})\right| = \mathcal{O}_{a.s.}(N(T)^{1/p}\log^2 N(T))) \quad \text{a.s.} \tag{3.67}$$

$$= \mathcal{O}_{a.s.}(T^{1/p}\log^2 T)) \quad \text{a.s.} \tag{3.68}$$

Here the second equality follows by (3.65). Furthermore, the asymptotic behaviour of $N(T)$ motivates the introduction of σ^2_T, τ^2_T defined as

$$\begin{aligned} \sigma^2_T &= s^2_n/\varrho, \quad \text{for } T \in [n, n+1), \\ \tau^2_T &= t^2_n/\varrho, \quad \text{for } T \in [n, n+1). \end{aligned}$$

By Theorem 3.5.1 (see also Theorem 1.2.1 of Csörgö and Révész [40]) we see that

$$\left|W_1(s^2_{N(T)}) - W_1(\sigma^2_T)\right| \text{ and } \left|W_2(t^2_{N(T)}) - W_1(\tau^2_T)\right| \text{ are both } \mathcal{O}_{a.s.}(T^{1/4}\log T) \text{ a.s.} \tag{3.69}$$

with

$$\sigma^2_T = \frac{\sigma^2_\xi}{\varrho} T + \mathcal{O}_{a.s.}(T/\log T) \text{ and } \tau^2_T = \mathcal{O}_{a.s.}(T/\log T).$$

Combining results (3.58), (3.67), and (3.69) concludes the proof. By the same arguments given in the proof of Theorem 3.3.1, the strong invariance principle holds for every initial distribution. □

Furthermore, we have by Berkes et al. [12; Proposition 1] that

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

Hence (3.27) immediately follows.

3.6.3. Theorem 3.3.4

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 Meyn and Tweedie [107; Propostion 5.4.5] and Tweedie [148; Theorem 3.1] respectively. Moreover, by Down et al. [54; 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 Nummelin [116]. Following a co-de-initialising argument of Roberts and Rosenthal [133], 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 3.2.1 $(Z^1_{T_n}, Z^2_{T_n})_n$, the jump chain of the first two coordinates of Z, has the same distribution as the split chain of the resolvent. From (3.6) and (3.7) we see that $(Z^1_{T_n}, Z^2_{T_n})_n$ is a Markov chain taking values in $E' := E \times [0,1]$ that moves according to the kernel

$$U'((x,u),(dy,dv)) = \nu(dy)\lambda(dv)\mathbb{1}_{\{u \le \alpha \mathbb{1}_C(x)\}} + W(x,dy)\lambda(dv)\mathbb{1}_{\{u > \alpha \mathbb{1}_C(x)\}}, \tag{3.70}$$

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.,

$$U'((x,u),(dy,dv)) \ge s(x,u)\nu(dy)\lambda(dv), \tag{3.71}$$

where

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

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

Lemma 3.6.6 (Hobert et al. [81; 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:*

$$U(x, dy) \geq \alpha \mathbb{1}_C(x)\nu(dy). \tag{3.72}$$

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

$$\mathbb{E}_\pi|g| \geq c\ \mathbb{E}_\nu|g|, \tag{3.73}$$

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 Hobert et al. [81; Lemma 1]. □

Proof of Theorem 3.3.4

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

$$\begin{aligned}
\mathbb{E}_\pi S_1^q &= \int_0^\infty q t^{q-1} \mathbb{P}_\pi \left(S_1 > t\right) dt \\
&= \int_0^\infty q t^{q-1} \sum_{m=0}^\infty \mathbb{P}_\pi \left(\bar{\tau}_A > m; N_t = m\right) dt \\
&= \int_0^\infty q t^{q-1} \sum_{m=0}^\infty \left(\underbrace{\mathbb{P}_\pi \left(\bar{\tau}_A > m; N_t = m; \bar{Z}_0 \in A\right)}_{=0} + \mathbb{P}_\pi \left(\bar{\tau}_A > m; N_t = m; \bar{Z}_0 \notin A\right) \right) dt \\
&= \int_0^\infty q t^{q-1} \int_{E'} \int_{E'} \sum_{m=0}^\infty \frac{t^m}{m!} e^{-t} \sum_{k=m+1}^\infty \left(U' - \nu \otimes \lambda \otimes s\right)^k (x, dz) \mathbb{1}_A(z) \pi(dx) dt \\
&= \int_{E'} \int_{E'} \sum_{m=0}^\infty \int_0^\infty \frac{t^{m+q-1}}{m!} q e^{-t} \sum_{k=m+1}^\infty \left(U' - \nu \otimes \lambda \otimes s\right)^k (x, dz) \mathbb{1}_A(z) dt \pi(dx) \\
&= \int_{E'} \int_{E'} \sum_{k=1}^\infty \left(U' - \nu \otimes \lambda \otimes s\right)^k (x, dz) \mathbb{1}_A(z) q \sum_{m=0}^{k-1} \frac{\Gamma(m+q)}{m!} \pi(dx) \\
&= \int_{E'} \int_{E'} \sum_{k=1}^\infty \frac{\Gamma(k+q)}{\Gamma(k)} \left(U' - \nu \otimes \lambda \otimes s\right)^k (x, dz) s(z) \pi(dx).
\end{aligned}$$

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 Nummelin and Tuominen [118; 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 3.3.4 we follow the argument of Bednorz and Łatuszyński [10; Theorem 2] with some minor adaptations. We

give the proof for completion.

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

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 3.6.6 concludes the proof. □

Remark 3.6.7. 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. △

Remark 3.6.8. For the exponentially ergodic case we would make use of Nummelin and Tuominen [117; Lemma 2.8] which states that for an exponentially ergodic Markov chain there exists an $r > 1$ such that

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

△

3.6.4. Theorem 3.3.6 and 3.3.7

Lemma 3.6.9 (Merlevède et al. [103; 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

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

Proof. The claim immediately follows from Merlevède et al. [103; Lemma 2.4] and a Borel-Cantelli argument. □

Proof of Theorem 3.3.6

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 3.3.4 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 $\alpha = \mathrm{Cov}_\nu(\xi_1, \tau_1)/\mathrm{Var}_\nu(\tau_1)$, 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 Einmahl [60; Theorem 1] and Csörgö and Horváth [38; Theorem 2.1], there exists two independent Brownian motions B_1 and B_2 such that

$$\left| \sum_{k=1}^{n} \xi_k - \alpha(\sum_{k=1}^{n} \tau_k - \varrho) - \tilde{\sigma} B_1 \right| = o_{a.s.}(\psi_n) \tag{3.74}$$

$$|R_n - n\varrho - \sigma_\tau B_2(n)| = o_{a.s.}(\psi_n) \quad , \tag{3.75}$$

with

$$\psi_n = n^{\max\left(\frac{1}{\beta-1}, \frac{1}{p}\right)}. \tag{3.76}$$

Note that in (3.74) we have that $\mathbb{E}_\nu \xi_1 = 0$ by Theorem 3.2.4 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 $o_{a.s.}(n^{1/p})$. By Komlos [90; 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

$$\left| N(n) - \frac{\varrho}{\gamma} n - \frac{\sigma_\tau}{\gamma} B_2(n) \right| = \mathcal{O}_{a.s.}(\log n) \quad , \tag{3.77}$$

where $\gamma = \sigma_\tau^2/\varrho$ and N is constructed increment-wise from B_2 in a deterministic way and is therefore also independent of B_1. From (3.75) and (3.77) it follows that

$$|R_n - \gamma N(n)| = o_{a.s.}(n^{1/p}) \tag{3.78}$$

We claim that it therefore follows that

$$\left| \int_0^{R_n} [f(X_s) - \pi(f)] ds - \int_0^{\gamma N(n)} [f(X_s) - \pi(f)] ds \right| = O(n^{1/p}) \tag{3.79}$$

Indeed, we have that

$$\left| \int_0^{R_n} [f(X_s) - \pi(f)] ds - \int_0^{\gamma N(n)} [f(X_s) - \pi(f)] ds \right| = \left| \int_{b_n}^{c_n} f(X_s) - \pi(f) ds \right|, \tag{3.80}$$

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 (3.78) it follows that $\alpha_n = o_{a.s.}(n^{1/p})$, 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

$$\begin{aligned}\left|\frac{1}{n^{\frac{1}{p}}}\int_{b_n}^{c_n} f(X_s) - \pi(f)ds\right| &= \frac{1}{n^{\frac{1}{p}}}\left|\int_0^{c_n - b_n} f(X_{b_n+u}) - \pi(f)du\right| \\ &\leq \frac{1}{n^{\frac{1}{p}}}\int_0^{\alpha_n} |f(X_{b_n+u}) - \pi(f)|du \\ &\leq \frac{1}{n^{\frac{1}{p}}}\int_0^{\varepsilon_1 n^{1/p}} |f(X_{b_n+s}) - \pi(f)|ds \\ &\leq \frac{1}{n^{\frac{1}{p}}}\sum_{j=\eta(b_n)}^{\eta(b_n+\varepsilon_1 n^{1/p})} Y_j + \frac{1}{n^{\frac{1}{p}}}\int_{R_{\eta(b_n+\varepsilon_1 n^{1/p})}}^{b_n+\varepsilon_1 n^{1/p}} |f(X_s) - \pi(f)|ds\end{aligned} \tag{3.81}$$

From (3.65) 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 (3.75) and (3.77) we obtain $R_n = n/\varrho + \mathcal{O}_{a.s.}(\sqrt{n \log\log n})$ *a.s.* and $N_n = n/\lambda + \mathcal{O}_{a.s.}(\sqrt{n \log\log n})$ *a.s.* respectively. Hence we have that $b_n = \mathcal{O}_{a.s.}(n)$ *a.s.*, and consequently $\eta(b_n) = \mathcal{O}_{a.s.}(n)$ 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

$$\eta(b_n + \varepsilon_1 n^{1/p}) - \eta(b_n) = \mathcal{O}_{a.s.}(\eta(\varepsilon_1 n^{1/p})) \quad a.s. \tag{3.82}$$

Indeed, since we have that $\mathbb{E}_\nu R_1^q < \infty$, by Csörgö and Horváth [38; Theorem 2.4] we can construct a Brownian motion $\tilde{B}_2$ such that

$$\left|\eta(T) - \frac{T}{\mu_\eta} - \sigma_\eta \tilde{B}_2(T)\right| = o_{a.s.}(T^{1/q}) \quad a.s., \tag{3.83}$$

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 $T \geq T_1(\omega)$ we have that

$$\frac{1}{T^{1/q}}\left|\eta(T) - \frac{T}{\mu_\eta} - \sigma_\eta \tilde{B}_2(T)\right| < \varepsilon. \tag{3.84}$$

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 $\eta(b_n)(\omega) \geq T_1(\omega)$ for all $n \geq N_2(\omega)$ and hence

$$\frac{1}{b_n^{1/q}}\left|\eta(b_n) - \frac{b_n}{\mu_\eta} - \sigma_\eta \tilde{B}_2(b_n)\right| < \varepsilon. \tag{3.85}$$

Since $b_n = \mathcal{O}_{a.s.}(n)$ almost surely, it follows that

$$\left|\eta(b_n) - \frac{b_n}{\mu_\eta} - \sigma_\eta \tilde{B}_2(b_n)\right| = o_{a.s.}(b_n^{1/q}) = o_{a.s.}(n^{1/q}) \tag{3.86}$$

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

$$\eta(b_n + a_n) - \eta(b_n) \leq \left|\eta(b_n + a_n) - (b_n + a_n)/\mu_\eta - \sigma_\eta \tilde{B}_2(\eta(b_n) + a_n)\right| \tag{3.87}$$

$$+ a_n/\mu_\eta + \left|-\eta(b_n) + b_n/\mu_\eta + \sigma_\eta \tilde{B}_2(b_n)\right| \tag{3.88}$$

$$+ \sigma_\eta \left|\tilde{B}_2(b_n + a_n) - \tilde{B}_2(b_n)\right| \tag{3.89}$$

By (3.86) the rhs of (3.87) and the second term in (3.88) are both $o_{a.s.}(n^{1/q})$ and thus $o_{a.s.}(n^{1/p})$. Furthermore, by Csörgö and Révész [39; Theorem 2] we have that

$$\limsup_{n\to\infty} \sup_{0\leq s\leq a_n} \frac{\left|\tilde{B}_2(n+s) - \tilde{B}_2(n)\right|}{[a_n(\log(n/a_n) + \log\log n)]^{1/2}} = 1 \quad a.s. \tag{3.90}$$

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

$$\sup_{0\leq s\leq a_n} \left|\tilde{B}_2(n+s) - \tilde{B}_2(n)\right| = \mathcal{O}_{a.s.}\left(n^{1/2p}\log(n)\right) = o_{a.s.}\left(n^{1/p}\right) \tag{3.91}$$

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

$$\sup_{0\leq s\leq a_n} \left|\tilde{B}_2(\eta(b_n) + s) - \tilde{B}_2(\eta(b_n))\right| = o_{a.s.}\left(\eta(b_n)^{1/p}\right) = o_{a.s.}\left(n^{1/p}\right) \tag{3.92}$$

Hence, we have shown that

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

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 (3.82) 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

$$\left| \sum_{k=0}^{n} Y_k - n\mu_Y - \sigma_Y B_3(n) \right| = o_{a.s.}(n^{1/p}) \quad , \tag{3.93}$$

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

$$\left| \sum_{k=0}^{\eta(b_n)} Y_k - \eta(b_n)\mu_Y - \sigma_Y B_3(\eta(b_n)) \right| = o_{a.s.}(\eta(b_n^{1/p})) = o_{a.s.}(n^{1/p}) \tag{3.94}$$

By the triangle inequality, we obtain

$$\left| \sum_{k=\eta(b_n)}^{\eta(b_n)+\tilde{a}_n} Y_k \right| \leq \left| \sum_{k=0}^{\eta(b_n)+\tilde{a}_n} Y_k - (\eta(b_n) + \tilde{a}_n)\mu_Y - \sigma_Y B_3(\eta(b_n) + \tilde{a}_n)) \right| \tag{3.95}$$

$$+ \tilde{a}_n \mu_Y + \left| - \sum_{k=0}^{\eta(b_n)} Y_k + \eta(b_n)\mu_Y + \sigma_Y B_3(\eta(b_n)) \right| \tag{3.96}$$

$$+ \sigma_y |B_3(\eta(b_n) + \tilde{a}_n) - B_3(\eta(b_n))| \tag{3.97}$$

$$\leq \tilde{a}_n \mu_Y + o_{a.s.}(n^{1/p}) \tag{3.98}$$

The last inequality follows, since by (3.94) both the term in (3.95) and the second term in (3.97) are $o_{a.s.}(n^{1/p})$ almost surely. Furthermore, by (3.92) the last inequality also follows. Hence it follows that

$$\mathbb{P}_\nu \left(\limsup_{n\to\infty} \frac{1}{n^{1/p}} \left| \sum_{k=\eta(b_n)}^{\eta(b_n+a_n)} Y_k \right| \leq K\mu_Y \right) = 1. \tag{3.99}$$

Hence the first term in the upper bound (3.81) is $\mathcal{O}_{a.s.}(1)$ almost surely. For the second term, we see that from (3.61), it follows that

$$Y_n = \int_{R_n}^{R_{n+1}} |f(X_s) - \pi(f)| ds = \mathcal{O}_{a.s.}(n^{1/p}) \tag{3.100}$$

Therefore

$$\begin{aligned}
\int_{R_{\eta(b_n+a_n)}}^{b_n+a_n} |f(X_{R_n+s}) - \pi(f)| ds &\leq \int_{R_{\eta(b_n+a_n)}}^{R_{\eta(b_n+a_n)+1}} |f(X_{R_n+s}) - \pi(f)| ds && (3.101)\\
&= Y_{\eta(b_n+\varepsilon_1 n^{1/p})} && (3.102)\\
&= \mathcal{O}_{a.s.}\left((\eta(b_n + \varepsilon_1 n^{1/p}))^{1/p}\right) && (3.103)\\
&= \mathcal{O}_{a.s.}\left((n + n^{1/p})^{1/p}\right) = \mathcal{O}_{a.s.}\left(n^{1/p}\right) && (3.104)
\end{aligned}$$

Hence our claim (3.81) follows, and consequently we have also shown (3.79). Combining (3.74), (3.78), and (3.79) it follows that

$$\left| \int_0^{\gamma N(n)} [f(X_s) - \pi(f)] ds - \alpha\gamma N(n) + \alpha\varrho n - \tilde{\sigma} B_1(n) \right| = o_{a.s.}\left(n^{\frac{1}{p}}\right) \tag{3.105}$$

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 (3.105) and subsequently making the substitution $n = \gamma n'$, it follows that

$$\begin{aligned}
\left| \int_0^{n} [f(X_s) - \pi(f)] ds - \alpha n + \alpha\varrho\Gamma_{n/\gamma} - \tilde{\sigma} B_1(\Gamma_{n/\gamma}) \right| &= o_{a.s.}\left({\Gamma_n}^{1/p}\right)\\
&= o_{a.s.}\left(n^{1/p}\right) \ , \qquad (3.106)
\end{aligned}$$

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 [90; Theorem 1] that there exists a Brownian motion B_4 such that

$$\left| \Gamma_n - \frac{n}{\lambda} - \frac{1}{\lambda} B_4(n) \right| = \mathcal{O}_{a.s.}(\log n) \tag{3.107}$$

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

3.6.9 with $n = \Gamma_n$ it follows that there exists a standard Brownian motion B_5 independent of N and Γ such that

$$\left| B_1(\Gamma_n) - \frac{1}{\sqrt{\lambda}} B_5(n) \right| = \mathcal{O}_{a.s.}(\log n) \tag{3.108}$$

Applying the obtained approximations given in (3.107) and (3.108) to (3.106) it follows that

$$\left| \int_0^n f(X_s) - \pi(f) ds - \left(\frac{\tilde{\sigma}}{\sqrt{\lambda\gamma}} B_5(n) - \frac{\alpha\varrho}{\lambda\sqrt{\gamma}} B_4(n) \right) \right| = o_{a.s.}\left(n^{1/p}\right) \tag{3.109}$$

Note that since B_4 and B_5 are independent we have that

$$W_n = \frac{1}{\sigma_f} \left(\frac{\tilde{\sigma}}{\sqrt{\lambda\gamma}} B_5(n) - \frac{\alpha\varrho}{\lambda\sqrt{\gamma}} B_4(n) \right) \tag{3.110}$$

is a standard Brownian motion since

$$\frac{\tilde{\sigma}^2}{\gamma\lambda} + \frac{\alpha^2\varrho^2}{\gamma\lambda^2} = \frac{\mathbb{E}_\nu \xi_1^2}{\varrho} = \sigma_f^2. \tag{3.111}$$

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

$$\frac{1}{n^{1/p}} \left| \int_0^n f(X_s(\omega)) ds - T\pi(f) - \sigma_f^2 W_n(\omega) \right| < C(\omega). \tag{3.112}$$

It immediately follows that (3.112) also holds for T sufficiently large and hence carries over for $T \to \infty$. By the same argument given in the proof of Theorem 3.3.1, the strong invariance principle holds for every initial distribution. □

Proof of Theorem 3.3.7

Proof. From Bierkens et al. [21; 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

$$\left\| \int_0^T f(X_t)\, dt - T\pi(f) - \Sigma_f^{1/2} W(T) \right\| \leq$$
$$\sqrt{d} \max_i \left| \int_0^T f_i(X_t^i)\, dt - T\pi_i(f_i) - \sigma_{f_i} W^i(T) \right|,$$

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 3.3.6. For every component $i = 1, \ldots, d$ 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 3.3.4 to obtain that

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

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

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

Note that for every coordinate i we have that

$$\left|\int_0^T f_i(X_t^i)\ dt - T\pi_i(f_i) - \sigma_{f_i}W^i(T)\right| \leq \left|\int_0^{R_1^i} f_i(X_t^i) - \pi_i(f_i)dt\right| \tag{3.113}$$

$$+\left|\int_{R_1^i}^T f_i(X_t^i) - \pi_i(f_i)dt - \sigma_{f_i}W^i(T)\right|. \tag{3.114}$$

By assuming that the process starts at its stationary distribution, it follows by the argument in the proof of Theorem 3.3.4 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 $\alpha_i = \mathrm{Cov}_\nu(\xi_1^i, \tau_1^i)/\mathrm{Var}_\nu(\tau_1^i)$, 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 Einmahl [60; Theorem 1] and Csörgö and Horváth [38; Theorem 2.1] to the sequence of random vectors

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

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

$$\left|\sum_{k=1}^n z_k - \mathbb{E}_\nu z_1 - \tilde{\Sigma}_z B_n\right| = o_{a.s.}\left(n^{1/p}\right)\ , \tag{3.115}$$

where $\tilde{\Sigma}_z = \mathrm{diag}(\mathrm{Var}_\nu(z_1), \cdots, \mathrm{Var}_\nu(z_k))$. 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

$$\left|\sum_{k=1}^n \xi_k^i - \alpha_i\left(\sum_{k=1}^n \tau_k^i - \varrho_i\right) - \tilde{\sigma}_i B_{i1}\right| = o_{a.s.}\left(n^{1/p}\right) \tag{3.116}$$

$$\left|R_n^i - n\varrho_i - \sigma_{\tau_i}B_{i2}(n)\right| = o_{a.s.}\left(n^{1/p}\right) \tag{3.117}$$

Note that in (3.116) we have that $\mathbb{E}_\nu \xi_1^i = 0$ by Theorem 3.2.4 and that $\tilde{\sigma}_i = \mathrm{Var}_\nu(\xi_1^i - \alpha_i(\tau_1^i - \varrho_i))$. By following the argument of the proof of Theorem 3.3.6 for every component, we see that

$$\left| \int_{R_1^i}^n f_i(X_t^i) - \pi_i(f_i)ds - \sigma_{f_i} W_n^i \right| = o_{a.s.}\left(n^{1/p}\right) \quad \text{for } i = 1, \dots, d. \tag{3.118}$$

By combining (3.113), (3.114) and (3.118) the claim follows. By the argument given in the proof of Theorem 3.3.1, the strong invariance principle holds for every initial distribution. □

3.6.5. Proof of Theorem 3.5.2

Proof. Firstly, by Proposition 3.3.3 there exist two standard Brownian motions W_1 and W_2 such that

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

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

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

as $T \to \infty$, where σ_ξ^2 and ϱ are defined in Theorem 3.3.3. An application of our strong invariance principle gives the following

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

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 Berkes et al. [12; 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 $T \geq T_0$

$$\sigma_T^2 \leq \left(\frac{\sigma_\xi^2}{\varrho} + \varepsilon \right) T \quad \text{and} \quad \sup_{u \geq 0} \{ \sigma_{u+a_T}^2 - \sigma_u^2 \} \leq \left(\frac{\sigma_\xi^2}{\varrho} + \varepsilon \right) a_T. \tag{3.119}$$

Since $\sigma_T^2 = \frac{\sigma_\xi^2}{\mu} T + \mathcal{O}(T/\log(T))$, the first required property described in (3.119) follows directly. From the proof of Theorem 3.3.5 we have that

$$\sigma_T^2 = s_n^2/\varrho, \quad \text{for } T \in [n, n+1). \tag{3.120}$$

Note that (3.120) is equivalent to $\sigma_T^2 = s_{\lfloor T \rfloor}^2/\varrho$ and therefore $\limsup_{u\to\infty}(\sigma_{\lfloor u \rfloor+1}^2 - s_{\lfloor u \rfloor}^2) \leq \limsup_{k\to\infty}(s_{k+1}^2 - s_k^2)/\varrho = \sigma_\xi^2/\varrho$, where the last equality follows since by Theorem 3.6.5 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

$$\begin{aligned}
\sup_{u > U_0} \{ \sigma_{u+a_T}^2 - \sigma_u^2 \} &= \sup_{u > U_0} \{ \sigma_{\lfloor u + a_T \rfloor}^2 - \sigma_u^2 \} \\
&\leq \sup_{u > U_0} \{ \sigma_{\lfloor u + a_T \rfloor}^2 - \sigma_{\lfloor u \rfloor}^2 \} \\
&\leq \limsup_{u \to \infty} \sum_{j=1}^{\lfloor a_T \rfloor} (\sigma_{\lfloor u \rfloor + j}^2 - \sigma_{\lfloor u \rfloor + j - 1}^2) \\
&\leq 1/\varrho \sum_{j=1}^{\lfloor a_T \rfloor} \limsup_{k \to \infty} (s_{\lfloor u \rfloor + j}^2 - s_{\lfloor u \rfloor + j - 1}^2) \\
&\leq (\sigma_\xi^2/\varrho + \varepsilon) \lfloor a_T \rfloor \leq (\sigma_\xi^2/\varrho + \varepsilon) a_T,
\end{aligned} \tag{3.121}$$

where the first equality follows from (3.120), the first inequality due to the fact that $(\sigma_u)_{u \geq 0}$ is a non-decreasing sequence. Note that for all $U_0 > 0$ we have that

$$\sup_u \{ \sigma_{u+a_T}^2 - \sigma_u^2 \} = \max \left\{ \sup_{u \leq U_0} \{ \sigma_{u+a_T}^2 - \sigma_u^2 \}, \sup_{u > U_0} \{ \sigma_{u+a_T}^2 - \sigma_k^2 \} \right\} \tag{3.122}$$

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

$$\sup_{u \leq U_0} \{ \sigma_{u+a_T}^2 - \sigma_u^2 \} \leq \sigma_{U_0 + a_T}^2 \leq (\sigma_\xi^2/\varrho + \varepsilon) a_T. \tag{3.123}$$

Combining (3.121) and (3.123) gives (3.122). Consequently, we have also shown that the required properties given in (3.119) hold. Hence for $T \geq T_0$ we obtain almost surely

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

where $\tilde{T}_\varepsilon$ and $\tilde{a}_{T,\varepsilon}$ are defined as $(\sigma^2_\xi/\varrho+\varepsilon)T$ and $(\sigma^2_\xi/\varrho+\varepsilon)a_T$ 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 3.5.1 we have that almost surely

$$\limsup_{T\to\infty}\sup_{0\leq t\leq\tilde{T}_\varepsilon-a_{T,\varepsilon}}\sup_{0\leq u\leq a_{T,\varepsilon}}\tilde{\beta}_{T,\varepsilon}|W(t+u)-W_t|\leq\sigma^2_\xi/\varrho$$

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

□

Chapter 4

Gaussian Approximations for High-dimensional MCMC

The widespread use of Markov Chain Monte Carlo (MCMC) methods for high-dimensional applications has motivated research into the scalability of these algorithms with respect to the dimension of the problem. Despite this, numerous problems concerning output analysis in high-dimensional settings have remained unaddressed. We present novel quantitative Gaussian approximation results for a broad range of MCMC algorithms. Notably, we analyse the dependency of the obtained approximation errors on the dimension of both the target distribution and the feature space. We demonstrate how these Gaussian approximations can be applied for MCMC uncertainty quantification. This includes determining the simulation effort required to guarantee Markov chain central limit theorems and consistent variance estimation in high-dimensional settings. We give quantitative convergence bounds for termination criteria and show that the termination time of a wide class of MCMC algorithms scales polynomially in dimension while ensuring a desired level of precision. Our results offer guidance to practitioners for obtaining appropriate standard errors and deciding the minimum simulation effort of MCMC algorithms in both multivariate and high-dimensional settings.

4.1. Introduction

Markov Chain Monte Carlo (MCMC) methods are widely applied in various high-dimensional settings, such as those encountered in computational Bayesian statistics and machine learning, see for example Brooks et al. [30];

Gawlikowski et al. [68]; Springenberg et al. [143]; Welling and Teh [154]. We consider the problem of sampling from a high-dimensional probability distribution π defined on $E \subseteq \mathbb{R}^N$ and we are interested in estimating $\pi(f)$ for some appropriately integrable function $f : E \to E'$ with $E' \subseteq \mathbb{R}^d$. We will refer to E and E' as the state space and feature space, respectively.

We follow the framework of Qin and Hobert [125]; Yang and Rosenthal [156]; Zhou et al. [161] where a drift-and-minorisation approach is used to obtain quantitative convergence bounds in high-dimensional settings. A drift condition describes how fast the Markov chain moves through the state space, while the minorisation condition controls how fast the Markov chain forgets its past. Our analysis considers both geometric and polynomial drift conditions, which characterise the varying speeds at which the Markov chain can move towards subsets of the state space. We also consider both the one-step and multi-step minorisation case, which covers all widely used MCMC algorithms, see Meyn and Tweedie [107; Proposition 5.4.5] and Orey [119].

We remind the reader that a weak Gaussian approximation holds for $X = (X_t)_{t\in\mathbb{N}}$ if the process can be defined on a probability space, together with a Brownian motion W, such that

$$\lim_{T\to\infty} \mathbb{P}\left(\frac{1}{\Psi_T} \left| \sum_{t=0}^{T} f(X_t) - T\pi(f) - \Sigma_f^{1/2} W_T \right| \leqslant K \bar{\psi}_N \psi_d \right) = 1, \tag{4.1}$$

where $|\cdot|$ denotes the Euclidean norm, K denotes a dimension-independent almost surely finite constant, and $\bar{\psi}_N, \psi_d$, and Ψ_T denote the dependence of the approximation error on the dimension of the state space, the dimension of the feature space, and the sampling time respectively. Similarly, we say that a strong Gaussian approximation holds for $X = (X_t)_{t\in\mathbb{N}}$ if

$$\mathbb{P}\left(\limsup_{T\to\infty} \frac{1}{\Psi_T} \left| \sum_{t=0}^{T} f(X_t) - T\pi(f) - \Sigma_f^{1/2} W_T \right| \leqslant K \bar{\psi}_N \psi_d \right) = 1. \tag{4.2}$$

We use the customary notation for Gaussian approximations, i.e.,

$$\left| \sum_{t=0}^{T} f(X_t) - T\pi(f) - \Sigma_f^{1/2} W_T \right| = \begin{cases} \mathcal{O}_P\left(\bar{\psi}_N \psi_d \Psi_T\right) \\ \mathcal{O}_{a.s.}\left(\bar{\psi}_N \psi_d \Psi_T\right) \end{cases}, \tag{4.3}$$

where $\mathcal{O}_P$ and $\mathcal{O}_{a.s.}$ denote the weak and strong approximation respectively. We will further impose some regularity conditions on the asymptotic covariance matrix Σ_f and mainly consider component-wise moment conditions of the form $\sup_{i\in\{1,\cdots,d\}} \pi(|f_i|^{p+\epsilon}) < \infty$ for given $p > 2$ and $\epsilon > 0$. Table 4.1 summarises the results of Theorems 4.3.1 and 4.3.2 which describe the dependence of the Gaussian approximation rate on the simulation time.

	one-step minorisation	**multi-step minorisation**
geometric drift	$T^{1/p}\log(T)$	$T^{1/4+1/4(p-1)}\log(T)$
polynomial drift	$T^{1/p_0}\log(T)$	$T^{1/4+1/4(p_0-1)}\log(T)$

Table 4.1: Gaussian approximation rate Ψ_T

Here p_0 is specified in Equation (4.8) of Theorem 4.3.1 and depends explicitly on both the moment condition and the degree of polynomial drift. Our Gaussian approximation results cover a larger class of polynomially ergodic Markov chains than aforementioned works and are the first to quantify the influence of the polynomial drift on the approximation rate. Furthermore, we note that besides the result of Merlevède et al. [103], we are the first to obtain the optimal rate in a Markov chain setting. The proof of the Gaussian approximation for the one-step minorisation case builds upon the approach of Merlevède et al. [103]. Merlevède et al. [103] also utilises the regenerative structure obtained from the one-step minorisation condition and subsequently applies the Gaussian approximation of Zaitsev [159]. However, that result is not applicable to our setting since it requires the existence of the moment-generating function. Furthermore, the result of Merlevède et al. [103] assumes geometric ergodicity and bounded one-dimensional functions f. Currently, under the moment condition $\pi(|f|^p)$, the best-known Gaussian approximation error for MCMC algorithms is $\max\left\{T^{1/p}, T^{1/4}\right\}\log(T)$, see the results of Csáki and Csörgő [37]; Flegal and Jones [66]; Jones et al. [87] and Banerjee and Vats [8]; Li and Qin [95] for the one-dimensional and multivariate setting respectively. Under the same moment conditions, we now obtain the optimal Komlós–Major–Tusnády approximation rate, up to a logarithmic factor, for the one-step minorisation case. For the multi-step minorisation case, we do not recover the optimal rate; however, to the best of our knowledge, our rate is the best rate currently available for the class of

processes considered.

Additionally, our Gaussian approximation results provide explicit bounds for $\bar{\psi}_N$ and ψ_d, the dependence of the dimension of the state space and feature space on the approximation rate. For the weak Gaussian approximation, we obtain $\psi_d = d^{3/2}$ in all settings, however, this can be greatly improved under growth conditions on the moments of $|f|$ under π. This is reasonable in settings with sparsity and regularisation. The dependence of the dimension of the state space is expressed in terms of the parameters of the drift and minorisation conditions and properties of the asymptotic covariance matrix. Let q denote the speed of polynomial drift, given stability conditions on the drift and minorisation, our obtained weak Gaussian approximation results guarantee a CLT for Markov processes, provided that the dimension of the feature space grows as given in Table 4.2 for any $\bar{\varepsilon} > 0$.

	one-step minorisation	**multi-step minorisation**
geometric drift	$o\left(T^{1/3-\bar{\varepsilon}}\right)$	$o\left(T^{1/6-\bar{\varepsilon}}\right)$
polynomial drift	$o\left(T^{\frac{q-2}{3q}-\bar{\varepsilon}}\right)$	$o\left(T^{\frac{q-2}{6(q-1)}-\bar{\varepsilon}}\right)$

Table 4.2: Growth rate d such that CLT holds for sufficiently large p

As the speed of polynomial drift tends to the geometric case, which corresponds to $q \to \infty$, we show that the corresponding approximation errors in Table 4.1 coincide and consequently also the growth conditions of the dimension given in Table 4.2.

In order to construct valid confidence ellipsoids for our features of interest, we require the weak Gaussian approximation to hold. However, the usual approach in statistics, where the dimension is allowed to grow with the sample size, which in this context corresponds to the simulation time, is not appropriate. In the MCMC setting, the dimension of the target distribution and the set of features of interest are known prior to the simulation. Consequently, the critical question becomes how large the simulation time needs to be to ensure that the Gaussian approximations remain reliable in the setting where the dimension of the problem grows. In Table 4.3 below, we see how the simulation time is required to grow with the dimension of the

feature space in order to guarantee a CLT. For two sequences (a_n) and (b_n) we write $a_n = \Omega(b_n)$ if $\limsup_{n\to\infty} \frac{a_n}{b_n} > 0$.

	one-step minorisation	**multi-step minorisation**
geometric drift	$\Omega\left(d^{3+\bar{\varepsilon}}\right)$	$\Omega\left(d^{6+\bar{\varepsilon}}\right)$
polynomial drift	$\Omega\left(d^{\frac{3q}{q-2}+\bar{\varepsilon}}\right)$	$\Omega\left(d^{\frac{6(q-1)}{q-2}+\bar{\varepsilon}}\right)$

Table 4.3: Simulation time T such that CLT holds given sufficiently large p

For applications in Bayesian statistics, our results provide a direct link between the statistical model complexity and the computational complexity of the MCMC algorithm. Any theoretical guarantees for the posterior distribution, such as large-sample concentration results and prior induced regularisation, have direct implications on the required running times for sampling algorithms. We note that under additional assumptions on the target distribution, the growth rates presented in Table 4.1 and Table 4.2 can be greatly improved, see Remark 4.3.4 for a more extensive discussion on this result.

By taking the dimensionality into account, multiple insights that do not arise in the finite-dimensional setting are revealed. In the high-dimensional setting, a good initialisation of the Markov chain is crucial, as improper initial states can lead to exponentially increasing bounds. Additionally, also the minorisation volume plays a vital role, since contrary to the low-dimensional setting, a naive minorisation lower bound can cause to the Gaussian approximation error rate to increase exponentially in dimension. For the mixing time of high-dimensional MCMC algorithms, similar observations have been made, see for example Qin and Hobert [125]; Rajaratnam and Sparks [128]; Yang and Rosenthal [156] Furthermore, also the isometry of the asymptotic covariance matrix plays a larger role, since for isotropic target distributions we are able to give better convergence guarantees. Our obtained results, only have an additional $\sqrt{d}$ penalty compared to the currently best-known weak Gaussian approximation results for independent random vectors satisfying only p moments given in Bonis [24]; Eldan et al. [63]; Mies and Steland [108].

Through our obtained Gaussian approximation results, we are able to extend various results concerning output analysis to a wider array of applications. Firstly, we consider the estimation of the asymptotic covariance matrix. Many key aspects of output analysis for MCMC depend on the uncertainty quantification of our sampling algorithm. Estimating the Monte Carlo standard error is essential for ensuring the credibility of our simulation results and for computing many convergence diagnostics. In Vats et al. [151] and Vats et al. [152], the consistency of the spectral variance and batch means estimator are proven, respectively, under the assumption of a Gaussian approximation with an implicit rate. Consequently, our results are immediately applicable and can be used to adapt the tuning parameters of the considered variance estimation methods to high-dimensional settings. In accordance with empirical findings, we observe noteworthy differences between the simulation requirements for the polynomial and geometric drift cases, as well as between low and high-dimensional scenarios. These findings are summarised in Table 4.8 and Table 4.9 respectively.

Our results enable us to study the convergence complexity of a broad range of termination criteria and show that the termination rules introduced in Glynn and Whitt [73] and Vats et al. [152] can be applied to high-dimensional settings. These results also provide us with insights regarding the choice of termination rule. The analysis of these termination rules relies on the FCLT and consistency of variance estimators. Our results allow us to take the error of the Gaussian approximation and the convergence rate of the variance estimator into account and thus generalise the analysis of aforementioned results. Finally, we obtain novel requirements for the minimum simulation threshold that guarantee the validity of variance estimation and termination criteria in high-dimensional settings. We give conditions that guarantee that the termination time of an MCMC algorithm scales polynomially in dimension while ensuring a desired level of precision.

This chapter is organised as follows. In Section 4.2, we review some preliminary results regarding MCMC. In Section 4.3, we give our obtained Gaussian approximation results. In Section 4.4, we apply our results to MCMC output analysis and provide guarantees for the termination time. In Section 4.5, we present the proofs of our results.

Notation: In this chapter, we will denote the Euclidean norm of a vector $x \in \mathbb{R}^d$ with $|x|$. For a matrix A, we denote the Frobenius norm and the spectral norm with $|A|$ and $|A|_*$ respectively. Furthermore, let $\sigma_d(A)$ and $\sigma_1(A)$, denote the largest and smallest eigenvalue respectively.

4.2. Drift and Minorisation conditions

We consider a Markov chain $X = (X_t)_{t\in\mathbb{N}}$ on $(E, \mathcal{E})$ where a $E \subseteq \mathbb{R}^N$ denotes the state-space and $\mathcal{E}$ the corresponding Borel σ-algebra, with m-step transition kernel $P^m(x, \cdot)$. We follow the framework of Qin and Hobert [125]; Yang and Rosenthal [156]; Zhou et al. [161] where a drift-and-minorisation approach is used to obtain quantitative convergence bounds in high-dimensional settings. We say that the Markov chain satisfies a geometric drift condition if Drift Condition 5 holds.

Drift Condition 5. *Let there exist a function $V : E \to \mathbb{R}^+$, some set C, constants $\lambda \in (0, 1)$ and $0 < b, \upsilon_C < \infty$ such that $\upsilon_C = \sup_{x\in C} V(x)$ and*

$$PV(x) = \int_E V(y)P(x, dy) \leq \lambda V(x) + b\mathbb{1}_C(x),$$

for some set $C \in \mathcal{E}$.

In many applications, we can only guarantee that the drift function decays at a polynomial rate while the process is not in C. This corresponds to the following polynomial drift condition.

Drift Condition 6. *Let there exist a function $V : E \to \mathbb{R}^+$, some set C, constants $0 < c, b, \upsilon_C < \infty$, such that $\upsilon_C = \sup_{x\in C} V(x)$ and $\eta \in (0, 1)$ such that*

$$PV(x) \leq V(x) - cV(x)^\eta + b\mathbb{1}_C(x),$$

for some set $C \in \mathcal{E}$ with $\pi(C) > 0$.

It will also be useful to consider the geometric and polynomial drift condition for the m_0-skeleton of the Markov chain.

Drift Condition 7. *Let there exist a function $V : E \to \mathbb{R}^+$, some set C, constants $\lambda \in (0, 1)$ and $0 < b, \upsilon_C < \infty$ such that $\upsilon_C = \sup_{x\in C} V(x)$ and*

$$P^{m_0}V(x) = \int_E V(y)P^{m_0}(x, dy) \leq \lambda V(x) + b\mathbb{1}_C(x).$$

Drift Condition 8. *Let there exist a function $V : E \to \mathbb{R}^+$, some set C, constants $0 < c, b, \upsilon_C < \infty$, such that $\upsilon_C = \sup_{x \in C} V(x)$ and $\eta \in (0, 1)$ such that*

$$P^{m_0} V(x) \leq V(x) - cV(x)^\eta + b\mathbb{1}_C(x).$$

Note that we can obtain a drift condition for the m_0–skeleton by iterating the one-step drift condition. However, in order for the parameters of the drift condition to remain tractable, we give some sufficient conditions. Propositions 4.5.4 and 4.5.6 and Propositions 4.5.5 and 4.5.7 give conditions such that the one-step drift condition implies the desired drift condition for the skeleton chain in case of the geometric and polynomial drift condition respectively.

We say that an associated local m_0-step minorisation condition holds for the Markov chain if the following holds.

Minorisation Condition 2. *Let ν be some probability measure defined on C such that*

$$P^{m_0}(x, \cdot) \geq \alpha \mathbb{1}_C(x)\nu(\cdot),$$

the minorisation volume $\alpha \in (0, 1]$, $m_0 \in \mathbb{N}$ and small set C with $\pi(C) > 0$.

It is known that a multi-step minorisation condition holds for all widely used MCMC algorithms, see for example Meyn and Tweedie [107; Proposition 5.4.5]. Often it can even be shown that the Markov chain satisfies a one-step minorisation condition. For a more detailed explanation regarding the interpretation of the drift and minorisation conditions and how they relate to one another, we refer to Section 4.5 and the given references. In the framework of Qin and Hobert [125]; Yang and Rosenthal [156]; Zhou et al. [161], it is demonstrated that for the drift and minorisation conditions to behave well as the dimension of the state-space increases, it is useful to consider a family of drift functions. Let π_N denote the stationary distribution of the Markov chain and let N be the dimension of the corresponding state-space. We say that a family of non-negative functions, $\{V_N(\cdot)\}_{N \in \mathbb{N}}$, is a family of drift functions if for each N they all satisfy a certain class of drift conditions. We say that a family of geometric drift conditions is stable if $\lambda := \limsup_{N \to \infty} \lambda_N < 1$ and $b := \limsup_{N \to \infty} b_N < \infty$. Similarly, we say that a family of polynomial drift conditions is stable if $\eta := \liminf_{N \to \infty} \eta_N > 0$, $c := \liminf_{N \to \infty} c_N > 0$ and $b := \limsup_{N \to \infty} b_N < \infty$. Furthermore, a minorisation condition is

stable provided that $\alpha := \liminf_{N\to\infty} \alpha_N > 0$. In high-dimensional settings, the drift function can exhibit undesirable concentration behaviour, causing the small set C to grow too fast, which in turn leads the minorisation volume α to tend to zero at an uncontrolled rate. Consequently, as shown in, for example, Rajaratnam and Sparks [128], many existing convergence bounds deteriorate as the dimension of the state space grows large. In the remainder of the paper, we will often simply write π and V instead of π_N and V_N. However, it is important to keep in mind that when we study the behaviour of the chain as the state-space grows, $N \to \infty$, we are considering a family of Markov chains $\{X_N\}_{N\in\mathbb{N}}$ with a corresponding family of target distributions and drift and minorisation conditions. Furthermore, throughout the paper, we will consider all drift and minorisation conditions to be stable.

4.3. Gaussian approximation for MCMC samplers

We consider the following component-wise moment conditions on our features:

Assumption 4. *Let $f : E \to \mathbb{R}^d$ be a measurable function such that either of the following moment conditions holds*

1. $\sup_{i\in\{1,\cdots,d\}} \pi(|f_i|^{p+\varepsilon}) < \infty$ *holds for given $p > 2$ and some $\varepsilon \in (0, 1/p]$,*
2. $\sup_{i\in\{1,\cdots,d\}} \pi(e^{tf_i}) < \infty$ *holds for t in some neighbourhood of 0.*
3. $\sup_{i\in\{1,\cdots,d\}} |f_i|_\infty < \infty$,

where $|\cdot|_\infty$ denotes the usual supremum norm.

Most of our results will rely on Assumption 4.1, which is the most reasonable to assume in practice. Furthermore, we will impose the following regularity conditions on the covariance matrix of our MCMC sampler.

Assumption 5. *Suppose that for every $T \geq 0$ the smallest eigenvalues of*

$$\Sigma_{f_T} = \mathrm{Cov}_\pi \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \{f(X_t) - \pi(f)\} \right)$$

and $\Sigma_f := \lim_{T\to\infty} \Sigma_{f_T}$ are larger than some constant $\sigma_0 > 0$. Furthermore, suppose that

$$\sup_{i\in\{1,\cdots,d\}} \sup_{j\in\{1,\cdots,d\}} \left|\Sigma_{f_{ij}}\right| < \infty,$$

where $\Sigma_{f_{ij}}$ denotes the (i,j)–th entry of the matrix Σ_f for $1 \leq i, j \leq d$.

Assumption 5 guarantees that the asymptotic covariance matrix is well defined, and that both the empirical and asymptotic covariance matrix are non-singular. Firstly, we state our obtained weak Gaussian approximation results for the one-step minorisation case under both a geometric and polynomial drift condition.

Theorem 4.3.1. *Let $(X_t)_{t\in\mathbb{N}}$ be an irreducible aperiodic stationary Markov chain, assume that a one-step minorisation condition (M2 with $m_0 = 1$) holds and that drift condition 5 is satisfied. Then for all functions f for which Assumption 4 and 5 hold we can, on an enriched probability space, define a process that is equal in law to X and a standard d-dimensional Brownian motion W such that*

$$\left|\sum_{t=1}^{T} f(X_t) - T\pi(f) - \Sigma_f^{1/2} W_T\right| = \begin{cases} \mathcal{O}_P\left(d^{3/2}\bar{\psi}_N^{2/p} T^{1/p}\log(T)\right) \\ \mathcal{O}_{a.s.}\left(d^{25/4+\theta_0}\log^*(d)\bar{\psi}_N^{2/p}\left(\frac{\sigma_d}{\sigma_0}\right)^{1/2} T^{1/p}\log(T)\right) \end{cases}, \tag{4.4}$$

where

$$\bar{\psi}_N := \psi(\alpha,\lambda,b) = \alpha^{-1}\left(\frac{b}{\alpha(1-\lambda)}\right)^{1+\varepsilon/p}\left(\frac{p}{\ln(1/\lambda)e}\right)^{p}\sup_{i\in\{1,\ldots,d\}}\pi(|f_i|^{p+\varepsilon}) \tag{4.5}$$

If we assume that drift condition 6 holds instead of 5, then we have that

$$\left|\sum_{t=1}^{T} f(X_t) - T\pi(f) - \Sigma_f^{1/2} W_T\right| = \begin{cases} \mathcal{O}_P\left(d^{3/2}\tilde{\psi}_N^{2/p_0} T^{1/p_0}\log(T)\right) \\ \mathcal{O}_{a.s.}\left(d^{25/4+\theta_0}\log^*(d)\tilde{\psi}_N^{2/p}\left(\frac{\sigma_d}{\sigma_0}\right)^{1/2} T^{1/p_0}\log(T)\right) \end{cases}, \tag{4.6}$$

where

$$\tilde{\psi}_N := \tilde{\psi}_N(\alpha,b,c,\upsilon_C) = \alpha^{-1}\left(1+\frac{b}{c\alpha}+\frac{\upsilon_c - c + b}{1-\alpha}\right)^{\varepsilon/p_0}\sup_{i\in\{1,\ldots,d\}}\pi(|f_i|^{p+\varepsilon}) \tag{4.7}$$

and

$$p_0 = \begin{cases} \frac{pq(\eta)}{p+q(\eta)+\varepsilon}, & \text{if } \frac{2p}{3p-2} < \eta \leq p(p+\varepsilon)/(p(p+\varepsilon)+\varepsilon), \\ p, & \text{if } \eta > p(p+\varepsilon)/(p(p+\varepsilon)+\varepsilon), \\ q(\eta) - \bar{\epsilon}, & \text{if } \eta > 1/2 \quad \text{and A4.2 holds}, \end{cases} \tag{4.8}$$

with $q(\eta) = \eta/(1-\eta)$ *and the entries of* Σ_f *are given by*

$$\Sigma_f = \mathrm{Var}_\pi(f(X_0)) + \sum_{k=1}^{\infty} \mathrm{Cov}_\pi(f(X_0), f(X_k)) + \sum_{k=1}^{\infty} \mathrm{Cov}_\pi(f(X_k), f(X_0)). \tag{4.9}$$

In Theorem 4.3.2, we formulate our obtained Gaussian approximation results for the multi-step minorisation case under both a geometric and polynomial drift condition.

Theorem 4.3.2. *Let* $(X_t)_{t\in\mathbb{N}}$ *be an irreducible aperiodic stationary Markov chain, assume that an* m_0*-step minorisation condition holds and that drift condition 7 is satisfied. Then for all functions* f *for which Assumption 4.1 and 5 hold we can, on an enriched probability space, define a process that is equal in law to* X *and a standard* d*-dimensional Brownian motion* W *such that*

$$\left| \sum_{t=1}^{T} f(X_t) - T\pi(f) - \Sigma_f^{1/2} W_T \right| = \begin{cases} \mathcal{O}_P\left(d^{3/2} \bar{\psi}_N^{2/p} T^{\frac{1}{4}+\frac{1}{4(p-1)}} \right) \\ \mathcal{O}_{a.s.}\left(d^{25/4+\theta_0} \log^*(d) \psi_N^{2/p} \left(\frac{\sigma_d}{\sigma_0}\right)^{1/2} T^{\frac{1}{4}+\frac{1}{4(p-1)}} \right) \end{cases}, \tag{4.10}$$

where

$$\bar{\psi}_N := \psi(\alpha, \lambda, b, m_0) = \alpha^{-1} \left(\frac{b m_0}{\alpha(1-\lambda)} \right)^{\varepsilon/p} \sup_{i\in\{1,\ldots,d\}} \pi(|f_i|^{p+\varepsilon})$$

If we assume that drift condition 8 holds instead of 7, then we have that

$$\left| \sum_{t=1}^{T} f(X_t) - T\pi(f) - \Sigma_f^{1/2} W_T \right| = \begin{cases} \mathcal{O}_P\left(d^{3/2} \tilde{\psi}_N^{2/p} \left(\frac{\sigma_d}{\sigma_0}\right) T^{\frac{1}{4}+\frac{1}{4(p_0-1)}} \right), \\ \mathcal{O}_{a.s.}\left(d^{25/4+\theta_0} \log^*(d) \tilde{\psi}_N^{2/p} \left(\frac{\sigma_d}{\sigma_0}\right)^{1/2} T^{\frac{1}{4}+\frac{1}{4(p_0-1)}} \right), \end{cases} \tag{4.11}$$

where

$$\tilde{\psi}_N := \tilde{\psi}(\alpha, b, c, \upsilon_c, m_0) = \alpha^{-1} m_0^{q(\eta)/p_0^2} \left(1 + \frac{b}{c\alpha} + \frac{\upsilon_c - c + b}{1-\alpha}\right)^{(p-p_0+\varepsilon)/p} \sup_{i\in\{1,\ldots,d\}} \pi(|f_i|^{p+\varepsilon})$$

$$p_0 = \begin{cases} \frac{pq(\eta)}{p+q(\eta)+\varepsilon}, & \text{if } \frac{2p}{3p-2} < \eta \leq p(p+\varepsilon)/(p(p+\varepsilon)+\varepsilon), \\ p, & \text{if } \eta > p(p+\varepsilon)/(p(p+\varepsilon)+\varepsilon), \\ q(\eta) - \bar{\epsilon}, & \text{if } \eta > 1/2 \quad \text{and A4.2 holds}, \end{cases}$$

with $q(\eta) = \eta/(1-\eta)$ *and the entries of* Σ_f *are given by*

$$\Sigma_f = \mathrm{Var}_\pi(f(X_0)) + \sum_{k=1}^{\infty} \mathrm{Cov}_\pi(f(X_0), f(X_k)) + \sum_{k=1}^{\infty} \mathrm{Cov}_\pi(f(X_k), f(X_0)). \tag{4.12}$$

Furthermore, we note that Gaussian approximation results for the continuous-time setting can be obtained by a similar argument to the discrete-time multi-step minorisation case. Note that the approach of Theorem 4.3.1 is not applicable, since there is in general no equivalent regenerative structure for continuous-time processess as implied by the one-step minorisaiton condition in the discrete-time case. Therefore we formulate the following result. Note that both the weak and strong approximation hold.

Theorem 4.3.3. *Let $X = (X_t)_{t\geq 0}$ be an irreducible aperiodic stationary Markov processs. Assume that a small set condition holds for some set C accompanied with drift condition 5 or 6 for the corresponding P_{t_0}–skeleton chain. Then for all functions f for which Assumption 4.1 and 5 hold we can, on an enriched probability space, define a process that is equal in law to X and a standard d-dimensional Brownian motion W such that*

$$\left| \int_0^T f(X_t)dt - T\pi(f) - \Sigma_f^{1/2} W_T \right| = \mathcal{O}_P\left(d^{3/2} \bar{\psi}_N^{2/p} T^{\frac{1}{4}+\frac{1}{4(p-1)}} \right) \tag{4.13}$$

where

$$\bar{\psi}_N := \psi(\alpha, \lambda, b, t_0) = \alpha^{-1} \left(\frac{bt_0}{\alpha(1-\lambda)} \right)^{\varepsilon/p} \sup_{i\in\{1,\ldots,d\}} \pi(|f_i|^{p+\varepsilon})$$

If we assume that drift condition 6 holds instead of 5, then we have that

$$\left| \int_0^T f(X_t)dt - T\pi(f) - \Sigma_f^{1/2} W_T \right| = \mathcal{O}_P\left(d^{3/2} \tilde{\psi}_N^{2/p} \left(\frac{\sigma_d}{\sigma_0} \right) T^{\frac{1}{4}+\frac{1}{4(p_0-1)}} \right), \tag{4.14}$$

where

$$\tilde{\psi}_N := \tilde{\psi}(\alpha, b, c, \upsilon_c) = \alpha^{-1} t_0^{q(\eta)/p_0^2} \left(1 + \frac{b}{c\alpha} + \frac{\upsilon_c - c + b}{1-\alpha} \right)^{(p-p_0+\varepsilon)/p} \sup_{i\in\{1,\ldots,d\}} \pi(|f_i|^{p+\varepsilon})$$

where p_0 is the same is in the aforemention results and the entries of Σ_f are given by

$$\Sigma_f = \int_0^\infty \mathrm{Cov}_\pi(f(X_0), f(X_t))\ dt + \int_0^\infty \mathrm{Cov}_\pi(f(X_t), f(X_0))\ dt, \tag{4.15}$$

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

Remark 4.3.4. Firstly, our obtained weak Gaussian approximation results given in Theorems 4.3.1 and 4.3.2 guarantee a central limit theorem for Markov chains, provided that the dimension grows at most as described in Table 4.4 below for any $\bar{\varepsilon} > 0$.

	one-step minorisation	**multi-step minorisation**
geometric drift	$o\left(T^{\frac{p-2}{3p}-\bar{\varepsilon}}\right)$	$o\left(T^{\frac{p-2}{6(p-1)}-\bar{\varepsilon}}\right)$
polynomial drift	$o\left(T^{\frac{p_0-2}{3p_0}-\bar{\varepsilon}}\right)$	$o\left(T^{\frac{p_0-2}{6(p_0-1)}-\bar{\varepsilon}}\right)$

Table 4.4: Growth rate d such that CLT holds

It is important to note that the growth rates given in Table 4.4 are the general case. However, in most applications of MCMC, there are more structural properties present such that the growth rate of the dimension can be greatly improved. From the proofs of Theorems 4.3.1 and 4.3.2 it follows that the dependence of the dimension of the feature space is equal to

$$\sqrt{d}\pi(|f|^{p_0})^{2/p_0}.$$

In general, we will have that $\pi(|f|^{p_0})^{1/p_0} \leq \sqrt{d}\sup_i \pi(|f_i|^{p_0})^{1/p_0}$, which gives us $\psi_d = d^{3/2}$. However, in many applications of MCMC this term can be smaller. For example, in Bayesian statistics the prior is often used to induce sparsity in the posterior or to provide regularisation, which will reduce $\pi(|f|^p)$ and, in the most favourable case, allow growth rates of $d = o(T)$ and $d = o(\sqrt{T})$ for the one-step and multi-step minorisation respectively. Similarly, the dimension dependence of the strong approximation can be improved for specific settings. Moreover, under Assumption 4.3, it can be shown that the rate of Theorem 4.3.1 improves to $\mathcal{O}_P\left(d\log^2(C_N)\log^2(T)\right)$ for the weak approximation case.

Finally, in order for a central limit theorem to hold in the high-dimensional setting, we see that the simulation time of our MCMC algorithm should scale with the dimension of the state-space and feature space as detailed in Table 4.5 below, for any $\bar{\varepsilon} > 0$.

Hence we see that for large p the simulation time of our sampling algorithm should scale as ψ_d^2 and ψ_d^4 in the one-step and multistep minorisation

	one-step minorisation	**multi-step minorisation**
geometric drift	$\Omega\left(d^{\frac{3p}{p-2}+\bar{\varepsilon}}\ \bar{\psi}_N^{4/(p-2)+\bar{\varepsilon}}\right)$	$\Omega\left(d^{\frac{6(p-1)}{p-2}+\bar{\varepsilon}}\ \bar{\psi}_N^{8(p-1)/(p^2-2p)+\bar{\varepsilon}}\right)$
polynomial drift	$\Omega\left(d^{\frac{3p_0}{p_0-2}+\bar{\varepsilon}}\ \tilde{\psi}_N^{4/(p_0-2)+\bar{\varepsilon}}\right)$	$\Omega\left(d^{\frac{6(p_0-1)}{p_0-2}+\bar{\varepsilon}}\ \tilde{\psi}_N^{8(p_0-1)/(p_0^2-2p_0)+\bar{\varepsilon}}\right)$

Table 4.5: Simulation time T such that CLT holds

case respectively. We note that the simulation requirements for consistent estimation of the asymptotic variance and obtaining precision guarantees, are given by the results in Section 4.4.

△

Remark 4.3.5. We note that for the Weak Gaussian approximation error we can obtain a better dimension dependence in our bounds. The argument is given in detail in Pengel et al. [122] on page 36, 44 for the one-step and multi-step minorisation case respectively. Note that it is only a slight deviation of the proof given in this thesis. These results imply that the weak approximation errors in Theorems 4.3.1, 4.3.2,4.3.3 hold with $\psi_d = \sqrt{d}\pi(|f|^p)^{1/p} = \mathcal{O}(d)$, which implies the following growth rate and simulation times such that the CLT holds in high-dimensional settings.

	one-step minorisation	**multi-step minorisation**
geometric drift	$o\left(T^{\frac{p-2}{2p}-\bar{\varepsilon}}\right)$	$o\left(T^{\frac{p-2}{4(p-1)}-\bar{\varepsilon}}\right)$
polynomial drift	$o\left(T^{\frac{p_0-2}{2p_0}-\bar{\varepsilon}}\right)$	$o\left(T^{\frac{p_0-2}{4(p_0-1)}-\bar{\varepsilon}}\right)$

Table 4.6: Growth rate d such that CLT holds

△

	one-step minorisation	**multi-step minorisation**
geometric drift	$\Omega\left(d^{\frac{3p}{p-2}+\bar{\varepsilon}}\ C_N^{4/(p-2)+\bar{\varepsilon}}\right)$	$\Omega\left(d^{\frac{6(p-1)}{p-2}+\bar{\varepsilon}}\ C_N^{8(p-1)/(p^2-2p)+\bar{\varepsilon}}\right)$
polynomial drift	$\Omega\left(d^{\frac{3p_0}{p_0-2}+\bar{\varepsilon}}\ C_N^{4/(p_0-2)+\bar{\varepsilon}}\right)$	$\Omega\left(d^{\frac{6(p_0-1)}{p_0-2}+\bar{\varepsilon}}\ C_N^{8(p_0-1)/(p_0^2-2p_0)+\bar{\varepsilon}}\right)$

Table 4.7: Simulation time T such that CLT holds

Remark 4.3.6. Theorem 4.3.1 gives the first Gaussian approximation results that attain the optimal Komlós–Major–Tusnády approximation rate, up to a logarithmic factor, for MCMC samplers in the multivariate setting. We note that these results are applicable to the high-dimensional settings considered in Qin and Hobert [125]; Yang and Rosenthal [156] and Zhou et al. [161], where an asymptotically stable one-step minorisation condition and family of exponential drift conditions are shown to hold. Moreover, we note that Theorem 4.3.2 can also be proven for continuous-time processes under a petite set condition and similar drift conditions. Furthermore, we note that all obtained weak approximation results can also be formulated as high probability deviation bounds.

In Gouëzel [76] a strong Gaussian approximation is obtained with rate $o_{a.s.}(T^\theta)$ for any $\theta > \frac{1}{4} + \frac{1}{4(p-1)}$, under assumptions on the dependence decay of the process through its spectral properties. These conditions on the characteristic function could be challenging to verify in MCMC settings. While the obtained approximation error Ψ_T with respect to the sampling time is independent of dimension, contrary to previously obtained results mentioned in the paper, the dimensionality of both the state-space and the feature-space can still influence the overall approximation error. Lu et al. [100] gives a strong Gaussian approximation with rate $o_{a.s.}\left(T^{1/3+2/3(3p-2)}\right)$ for Hilbert space valued stochastic processes whose dependence is controlled through exponentially decaying β–coefficients. The results of Lu et al. [100] are therefore applicable for Markov chains satisfying a geometric drift condition. Applying their Gaussian approximation result to $\mathbb{R}^d$ results in the approximation rate of Gouëzel [76] with a multiplicative factor $d^8 \log(d)$ introduced to the approximation error. However, since their result requires exponential decay of β–mixing coefficients, they are not applicable to the polynomial drift condition case.

Finally, we note that there has been quite some work on extending dimension-dependent Berry-Esseen type results to settings with dependence, see for example Chang et al. [33] and the given references. However, we have not taken this approach since for most applications in MCMC output analysis we require the FCLT and its refinements. △

Remark 4.3.7. Contrary to previously obtained Gaussian approximation results for MCMC, see Flegal and Jones [66]; Banerjee and Vats [8]; Li and Qin [95]; Merlevède et al. [103]; Pengel and Bierkens [121]; Vats et al. [151], in the high-dimensional setting, Gaussian approximation cannot be expected to hold for arbitrary initial distributions. It is well known that regularity conditions on the initial state are required in high-dimensional settings, since, for example, a cold initialisation can easily lead to mixing time bounds that are exponentially increasing in dimension, as demonstrated by Bandeira et al. [7]. The harmonic function argument of Meyn and Tweedie [107; Proposition 17.1.6] that is often used to generalise these types of results to an arbitrary initial distribution is not applicable to the high-dimensional setting. Despite this, stationarity is not required for any of our obtained results to hold. By the Comparison theorem, Meyn and Tweedie [107; Theorem 14.2.2], it can be shown that our result can be formulated for any initial condition x, as detailed in Remark 4.5.20. However, the drift at the initial value, i.e., $V(x)$ would enter the obtained bounds, see (4.159) and (4.160) for the geometric and polynomial drift case respectively. Therefore, some regularity conditions, which would be similar in nature to a warm start condition, would be required to ensure that the initialisation does not dominate our obtained dimension-dependent bounds. △

Remark 4.3.8. In our approach, we control the decay of the dependence of our process through both the drift and minorisation conditions. In alignment with expectations, we see that if a polynomial drift condition holds, the dimension of the state-space introduces a larger penalty in our approximation rate when compared to the case where a geometric drift condition holds. Note that for higher rates of polynomial drift, the approximation rate T^{1/p_0} tends to the rate with a geometric drift.

We note that for our results to be applicable in high-dimensional settings, we require the drift conditions to be asymptotically stable. Additionally, we see that in contrast to the low-dimensional case where the minorisation volume is negligible, see for example Flegal and Jones [66; Remark 2], in the high-dimensional case it plays a vital role since an improper minorisation

lower bound can cause to the Gaussian approximation error rate to increase exponentially in dimension. Hence, approaches like the ones given in Qin and Hobert [125]; Yang and Rosenthal [156] are critical to ensure asymptotically stable drift and minorisation conditions.

Furthermore, we see that the spectral condition number, which is defined as the quotient of the largest and the smallest eigenvalue of the asymptotic covariance matrix, appears in our strong approximation errors. It is known that better dimension dependence in Gaussian approximations can be obtained for isotropic targets, see for example Fathi [64]. From the specification of the asymptotic covariance matrix given in (4.9), we see that if the distribution of the features under the target measure is isotropic, then the spectral condition number of the first term $\mathrm{Var}_\pi(f(X_0))$ will be smaller. Furthermore, if the auto-covariance of the process decays faster, then the second and third terms of (4.9) will be smaller. By Rio [132; Theorem 1.1] and Douc et al. [53; Theorem F.3.3], we see that we can control the decay of the auto-covariance through the drift conditions. Consequently, for the geometric drift case, we can guarantee the condition number of the asymptotic covariance matrix to be smaller than the polynomial case. △

Remark 4.3.9. We note that Theorem 4.3.3 directly assumes a drift condition for the resolvent chain or the skeleton chain. In practice, drift conditions of the form 3 and 4 will be proven for the process directly. We need to show that these imply drift conditions for the derived chains. The results of Down et al. [54; Theorem 5.1] and Douc et al. [52; Theorem 4.9] show that this indeed follows for the exponential and polynomial drift, respectively. Furthermore, we note that Theorem 4.3.3 can also be shown by showing that the resolvent chain satisfies drift condition 5 or 6. However, instead of the coordinate-wise application of Bednorz and Łatuszyński [10; Lemma 2] to prove Lemma 4.5.2 and 4.5.3, we would have to make use of a coordinate-wise application of Theorem 3.3.4. △

4.4. High-dimensional MCMC Output Analysis

In order to assess the accuracy of our sampling method, we need to estimate the asymptotic variance appearing in the central limit theorem. In the high-dimensional setting, this gives us additional requirements for the simulation time of our algorithms. Estimation of the asymptotic auto-covariance matrix Σ_f plays a central role in MCMC output analysis, specifically for comput-

ing many convergence diagnostics, and implementing termination criteria. Through our obtained Gaussian approximation results, we are able to extend the results on variance estimation and termination criteria for MCMC algorithms. We illustrate the applications of our obtained Gaussian approximation results to the batch means method considered in Vats et al. [152]. We use our results to adapt the tuning parameters of the considered variance estimation methods to take slower convergence rates into account due to polynomial drift conditions or high dimensionality. Finally, our quantitative convergence bounds for termination criteria allow us to analyse the influence of the ergodicity and dimensionality on the appropriate running time of our MCMC algorithms.

4.4.1. Estimation of the Asymptotic Variance

We first consider the multivariate batch estimator since it enjoys computational advantages over other variance estimation methods in MCMC settings. The batch means method divides the simulation output into k_T batches of length ℓ_T such that $k_T = \lfloor T/\ell_T \rfloor$. The batch means estimator is then given by

$$\hat{\Sigma}_T^{BM} = \frac{\ell_T}{k_T - 1} \sum_{i=1}^{k_T} \left(\bar{Z}_i(\ell_T) - \frac{1}{k_T} \sum_{i=1}^{k_T} \bar{Z}_i(\ell_T) \right) \left(\bar{Z}_i(\ell_T) - \frac{1}{k_T} \sum_{i=1}^{k_T} \bar{Z}_i(\ell_T) \right)^T, \tag{4.16}$$

where $\bar{Z}_i(\ell_T)$ denotes the sample average of each obtained batch, i.e.,

$$\bar{Z}_i(\ell_T) := \frac{1}{\ell_T} \sum_{s=(i-1)\ell_T}^{i\ell_T} f(X_s), \quad i = 1, \ldots, k_T. \tag{4.17}$$

We impose the following conditions on the batch size.

Assumption 6. *Let ℓ_T be an integer sequence such that $\ell_T \to \infty$ and $n/\ell_T \to \infty$ as $n \to \infty$ where ℓ_T and n/ℓ_T are non-decreasing. Moreover, assume that there exists a constant $c \geq 2$ such that $\sum_T (\ell_T/T)^c < \infty$, $(\ell_T/T)\log(T) = o(1)$, $\ell_T^{-1} \log T = o(1)$, and $T > 2\ell_T$.*

Applying our Gaussian approximations to the results on the batch means estimator of Vats et al. [152] gives us the following theorem.

Theorem 4.4.1. *Suppose that $f : \mathbb{R}^N \to \mathbb{R}^d$, with $\sup_{i\in\{1,\cdots,d\}} \pi(|f_i|^{p+\varepsilon}) < \infty$ for some $p \geq 4$ and let X satisfies a weak Gaussian approximation with approximation error $\bar{\psi}_N \psi_d \Psi_T \log(T)$ with $\psi_d = d^a$ for some $a > 0$. Assume that Assumption 6 holds, and that*

$$\frac{\bar{\psi}_N d\psi_d \Psi_T \log(T)}{\ell_{T,d}^{1/2}} = o(1) \text{ and } \frac{\bar{\psi}_N^2 d\psi_d^2 \Psi_T^2 \log(T)}{T} = o(1), \tag{4.18}$$

then we have that $\hat{\Sigma}_T^{BM} \to \Sigma_f$ with probability 1 as $T \to \infty$. Moreover, if we assume that

$$\Psi_T = \begin{cases} \Psi_T^{(1)} := T^{1/p_0} \log(T), \\ \Psi_T^{(2)} := T^{1/4+1/4(p_0-1)} \log(T), \end{cases} \tag{4.19}$$

for some $p_0 > 2$ and we choose the simulation time

$$T = \begin{cases} \Omega\left(\left(\bar{\psi}_N d\psi_d\right)^{\frac{2p_0}{(p_0-2)}(1+\bar{\delta})}\right), & \text{under rate } \Psi_T^{(1)} \\ \Omega\left(\left(\psi_N d^{1/4}\psi_d\right)^{\frac{p_0-1}{p_0-2}4(1+\bar{\delta})}\right), & \text{under rate } \Psi_T^{(2)}, \end{cases} \tag{4.20}$$

for any $\bar{\delta} > 1/(1+a)$ then the choice of batch size $\ell_T = d^{-(p_0-2)/(2p_0(1+\bar{\delta}))}\lfloor T^{\alpha} \rfloor$ with

$$\alpha = \begin{cases} \frac{1}{2} + \frac{p_0-2}{2p_0(1+\delta)} + \frac{1}{p_0}, & \text{under rate } \Psi_T^{(1)} \\ \frac{3}{4} + \frac{1}{4(p_0-1)} + \frac{(p_0-2)}{4(p_0-1)(1+\bar{\delta})}, & \text{under rate } \Psi_T^{(2)}, \end{cases} \tag{4.21}$$

optimises the given convergence rate for $T \to \infty$.

Remark 4.4.2. While the result of Theorem 4.4.1 is formulated in a high-dimensional setting, note that in an application where we can assume the influence of the dimension to be negligible, we obtain as an immediate corollary following the choice of bath size from our strong approximation results. △

It has empirically been observed that for many practical problems, where slower convergence rates to stationarity are expected, larger batch sizes and truncation windows are required when applying the batch means and spectral variance methods for MCMC simulation output. Consistency of the batch

	one-step minorisation	**multi-step minorisation**
exponential drift	$T^{\frac{1}{2}+\frac{1}{p_0}}$	$T^{\frac{3}{4}+\frac{1}{4(p_0-1)}}$
polynomial drift	$T^{\frac{1}{2}+\frac{1}{p_0}}$	$T^{\frac{3}{4}+\frac{1}{4(p-1)}}$

Table 4.8: Batch size ℓ_T multivariate setting

means estimator requires that each batch gives an accurate representation of the dependence structure of the process. Naturally, in situations with slower convergence rates, a larger batch size will be required. An immediate consequence of slower mixing, is the slower decay of the autocovariance function. Hence also spectral variance estimators will require larger truncation points for consistent estimation of the asymptotic variance. But these corrections for either polynomial convergence rates to stationarity or the dimension of the problem have been done in heuristic ways. In Table 4.9, we give the appropriate batch sizes for the batch means estimator, which guarantees consistency in the polynomial drift as well as in the high-dimensional setting.

	one-step minorisation	**multi-step minorisation**
exponential drift	$T^{\frac{1}{2}+\frac{p-2}{2p(1+\delta)}+\frac{1}{p}}$	$T^{\frac{3}{4}+\frac{1}{4(p-1)}+\frac{(p-2)}{4(p-1)(1+\delta)}}$
polynomial drift	$T^{\frac{1}{2}+\frac{p_0-2}{2p_0(1+\delta)}+\frac{1}{p_0}}$	$T^{\frac{3}{4}+\frac{1}{4(p_0-1)}+\frac{(p_0-2)}{4(p_0-1)(1+\delta)}}$

Table 4.9: Batch size ℓ_T high-dimensional setting

Furthermore, we note that our results are also applicable to the multivariate spectral variance estimator. Let $\hat{\gamma}_T(s)$ denote the sample autocovariance function at lag s, i.e.,

$$\hat{\gamma}_T(s) = \frac{1}{T}\sum_{t=0}^{T-s}(f(X_t) - \hat{\pi}_T(f))(f(X_{t+s}) - \hat{\pi}_T(f))^T.$$

The spectral variance estimator is then given by

$$\hat{\Sigma}_T^{SV} = \sum_{s=0}^{\ell_T - 1} w_T(s)\hat{\gamma}_T(s) + \sum_{s=0}^{\ell_T - 1} w_T(s)\hat{\gamma}_T^T(s), \tag{4.22}$$

where w_T and ℓ_T denote the lag window and truncation point respectively. Note that the spectral variance estimator estimates the spectral density of our process evaluated at frequency zero. It is well known that the sample analogue of the spectral density is inconsistent due to undersmoothing. The use of a lag window and a truncation point ensures that more weight is given to the lower frequencies, which can be estimated with higher precision. Whereas the higher frequencies, which are observed with a lower signal-to-noise ratio, are assigned less weight in the estimation procedure. In [151], it is shown how the tuning parameters of the spectral variance estimator can be determined from the Gaussian approximation rate. Following their approach, we impose the following conditions on the lag window.

Assumption 7. *The lag window $w_T(\cdot)$ is an even function and $|w_T(s)| \leq 1$ for all T and s, $w_T(0) = 1$ for all T, and $w_T(s) = 0$ for all $|s| \geq \ell_T$.*

We impose the following technical conditions on the truncation point.

Furthermore, the first and second difference of the lag window are defined as

$$\Delta_1 w_T(s) = w_T(s-1) - w_T(s)$$

and

$$\Delta_2 w_T(s) = w_T(s-1) - 2w_T(s) + w_T(s+1)$$

respectively. [66]: It has empirically been found that high-dimensional problems require larger batches/ windows for batch means and spectral variance methods in MCMC than theory suggests. Given our dimension-dependent approximation results in Theorem 4.3.1 and Theorem 4.3.2, we can adapt the tuning parameters of the spectral variance estimator to the dimension of both the state-space of the chain and the dimension of the vector of functionals we want to estimate.

Theorem 4.4.3. *Suppose that* $f : \mathbb{R}^N \to \mathbb{R}^d$, *with* $\sup_{i\in\{1,\cdots,d\}} \pi(|f_i|^{p+\varepsilon}) < \infty$ *for some* $p \geq 4$ *and let* X *satisfy a strong Gaussian approximation with approximation error* $\bar{\psi}_N \psi_d \Psi_T$. *Assume that Assumptions 6 and 7 hold, and that*

(i) $\dfrac{\ell_T}{T} \sum_{s=1}^{\ell_T} s|\Delta_1 w_T(s)| = o(1),$

(ii) $\ell_T \bar{\psi}_N^2 d\psi_d^2 \Psi_T^2 \log(T) \left(\sum_{s=1}^{\ell_T} |\Delta_2 w_T(s)| \right)^2 = o(1),$

(iii) $\bar{\psi}_N^2 d\psi_d^2 \Psi_T^2 \sum_{s=1}^{\ell_T} |\Delta_2 w_T(s)| = o(1),$

(iv) $\ell_T^{-1} \bar{\psi}_N d\psi_d \Psi_{\sqrt{T}} = o(1)$, *and*

(v) $\ell_T^{-1} \bar{\psi}_N d\psi_d \Psi_T = o(1).$

then $\widehat{\Sigma}_T^{SV} \to \Sigma_f$ *with probability 1 as* $T \to \infty$.

Note that only the approximation error of Theorem 4.3.1 is tight enough to ensure the consistency of the spectral variance estimator. Although, the spectral variance estimator is better than the batch means estimator in the mean squared error sense, the batch means estimator enjoys a computation advantage and requires less stringent conditions on the Gaussian approximation error. Moreover, it can be easily applied to continuous time sampling algorithms.

4.4.2. MCMC Termination Criteria

Sequential termination rules are the standard practice for determining the appropriate running time of an MCMC algorithm. The Fixed Volume Stopping Rule (FVSR) allows termination of the simulation when the volume of a confidence region for the parameter of interest is below some predetermined tolerance level. Firstly, note that due to the obtained weak Gaussian approximation results, we can construct a confidence interval for $\pi(f)$, namely,

$$C(T) = \left\{ x \in \mathbb{R}^d : T(\hat{\pi}_T(f) - x)^\top \hat{\Sigma}_T^{-1} (\hat{\pi}_T(f) - x) < q_\alpha \right\}, \tag{4.23}$$

where $\hat{\pi}_T(f)$ denotes the empirical average of f over the simulation output, and $\hat{\Sigma}_T$ denotes some estimator of the asymptotic covariance matrix, which is evaluated using simulation output until time T, and q_α denotes the $(1-\alpha)$–quantile of the χ^2 distribution with d degrees of freedom. Given some user-specified tolerance level ε, the FVSR defines the time of termination $T(\varepsilon)$ for our simulation experiment as

$$T(\varepsilon) = \inf\{t > 0 : \mathrm{Vol}(C(t))^{1/d} + \Lambda(t) \leq \varepsilon\}. \tag{4.24}$$

Here $\mathrm{Vol}(\cdot)$ denotes the standard volume element and $\Lambda(t)$ is some positive sequence tending to zero. The role of $\Lambda(t)$ is to prevent early termination due to an inaccurate estimate of the covariance matrix or unreliability of the CLT from an insufficient sample size. A common choice is $\Lambda(t) = \mathbb{1}_{\{t<T^*\}} + o(t^{-1/2})$, for some appropriate threshold T^*.

All termination criteria, see for example Glynn and Whitt [73]; Jones et al. [87]; Vats et al. [152] and Gong and Flegal [75], make use of some sort of minimum simulation threshold in order to prevent early termination due to an inaccurate estimate of the covariance matrix or unreliability of the FCLT due to the insufficient sample size. It is often mentioned that this minimum simulation effort should take the complexity of the problem into account. However, the choice of this simulation threshold has always been done in heuristic ways. Our approach for determining the simulation threshold guarantees that the approximation rate between the estimated confidence ellipsoid and its limiting quadratic form is of a smaller asymptotic magnitude than the desired precision level ε. This results in both the error of the Gaussian approximation and the covariance matrix estimation procedure being of a smaller magnitude than the desired precision level. In the following theorem, we generalise Glynn and Whitt [73; Theorem 1] by giving quantitative convergence bounds for the FVSR.

Theorem 4.4.4. *Suppose that X satisfies the following strong Gaussian approximation*

$$\left| \sum_{t=1}^{T} f(X_t) - T\pi(f) - \Sigma_f^{1/2} W_T \right| = \mathcal{O}_{a.s.}\left(\bar{\psi}_N \psi_d \Psi_T\right), \tag{4.25}$$

with approximation error $\Psi_T = T^{1/p_0}\log(T)$ for some $p_0 > 4$ and $\psi_d = d^a$ for some $a > 0$. Let $T_1(\varepsilon)$ be given by

$$T_1(\varepsilon) = \inf\{t > 0 : \mathrm{Vol}(C(t))^{1/d} + \varepsilon\Lambda(t) < \varepsilon\}, \tag{4.26}$$

with $C(t)$ the confidence ellipsoid given in (4.23) and $\Lambda(t) = \mathbb{1}_{\{t<T^*(\varepsilon,d,N)\}} + t^{-1}$, with

$$T^*(\varepsilon, d, N) = \left(\bar{\psi}_N \left(\frac{\mathrm{tr}(\Sigma_f)}{\sigma_0}\right)^2 d^3\psi_d\right)^{\frac{2p_0}{(p_0-2)}(1+\bar{\delta}_1)} \left(\frac{1}{\varepsilon}\right)^{\frac{4p_0}{(p_0-2)}(1+\bar{\delta}_2)} \vee e^{\frac{10p_0}{p_0-2}} \tag{4.27}$$

for any $\bar{\delta}_1 > 3/(3+a)$ and $\bar{\delta}_2 > 0$. Let $\widehat{\Sigma}_T$ in (4.23) denote the batch means estimator defined in (4.16), with batch size ℓ_T set as

$$\ell_T = \bar{\psi}_N \psi_d T^{1/2+1/p_0}. \tag{4.28}$$

Suppose that Assumptions 4 and 5 hold. Then we have as $\varepsilon \downarrow 0$ the following:

1. *The asymptotic behaviour of the termination time $T_1(\varepsilon)$ is characterised by*

$$\frac{\varepsilon^2 T_1(\varepsilon)}{c_{\alpha,d}^{2/d} \det(\Sigma_f)^{1/d}} = 1 + o_{a.s.}\left(\log^2(\bar{\psi}_N d^3\psi_d)\bar{\psi}_N^{-\bar{\delta}_1/2} d^{-1/2}\varepsilon\right), \tag{4.29}$$

   where $c_{\alpha,d}$ denotes the product of $q_\alpha^{d/2}$ and the volume of a standard d-dimensional hypersphere.

2. *Asymptotic validity of the resulting confidence set*

$$\mathbb{P}_\pi\left(C(T_1(\varepsilon)) \ni \pi(f)\right) \to 1 - \alpha. \tag{4.30}$$

By choosing an appropriate simulation threshold T^*, the results of Theorem 4.4.4 can also be guaranteed hold for the approximation rate obtained for the multi-step minorisation case.

Theorem 4.4.5. *Suppose that X satisfies the following strong Gaussian approximation*

$$\left|\sum_{t=1}^{T} f(X_t) - T\pi(f) - \Sigma_f^{1/2} W_T\right| = \mathcal{O}_{a.s.}\left(\bar{\psi}_N \psi_d \Psi_T\right), \tag{4.31}$$

with approximation error $\Psi_T = T^{1/4p_0+1/4(p_0-1)}\log(T)$ for some $p_0 > 2$ and $\psi_d = d^a$ for some $a > 0$. Let $T_1(\varepsilon)$ be defined in (4.24) with $\Lambda(t) =$

$\mathbb{1}_{\{t<T^*(\varepsilon,d,N)\}} + t^{-1}$, *where*

$$T^*(\varepsilon, d, N) = \left(\bar{\psi}_N \left(\frac{\mathrm{tr}(\Sigma_f)}{\sigma_0}\right)^2 d^3 \psi_d\right)^{\frac{4(p_0-1)}{(p_0-2)}(1+\bar{\delta}_1)} \left(\frac{1}{\varepsilon}\right)^{\frac{8(p_0-1)}{(p_0-2)}(1+\bar{\delta}_2)} \vee e^{\frac{16(p_0-1)}{(p_0-2)}} \tag{4.32}$$

for any $\bar{\delta}_1 > 3/(3+a)$ *and* $\bar{\delta}_2 > 0$. *Let* $\widehat{\Sigma}_T$ *in* (4.23) *denote the batch means estimator defined in* (4.16), *with batch size* ℓ_T *set as*

$$\ell_T = \bar{\psi}_N \psi_d T^{\frac{3}{4}+\frac{1}{4(p-1)}} \log^{\bar{\delta}_3}(T), \tag{4.33}$$

for any $\bar{\delta}_3 > 0$. *Suppose that Assumptions 4 and 5 hold. Then we have as* $\varepsilon \downarrow 0$ *that the conclusions of Theorem 4.4.4 regarding the asymptotic behaviour of the termination time and the asymptotic validity of the resulting confidence set hold.*

The quantitative convergence bounds obtained in Theorems 4.4.4 and 4.4.5 offer guidelines for the implementation of the FVSR in a wide array of settings.

Remark 4.4.6. While the result of Theorem 4.4.4 is formulated in a high-dimensional setting, note that in an application where we can assume the influence of the dimension to be negligible, we obtain as an immediate corollary the following appropriate minimum simulation thresholds. △

	one-step minorisation	**multi-step minorisation**
geometric drift	$\left(\frac{1}{\varepsilon}\right)^{\frac{4p}{(p-2)}(1+\bar{\delta}_2)} \vee e^{\frac{10p}{p-2}}$	$\left(\frac{1}{\varepsilon}\right)^{\frac{8(p-1)}{(p-2)}(1+\bar{\delta}_2)} \vee e^{\frac{16(p-1)}{(p-2)}}$
polynomial drift	$\left(\frac{1}{\varepsilon}\right)^{\frac{4p_0}{(p_0-2)}(1+\bar{\delta}_2)} \vee e^{\frac{10p_0}{p_0-2}}$	$\left(\frac{1}{\varepsilon}\right)^{\frac{8(p_0-1)}{(p_0-2)}(1+\bar{\delta}_2)} \vee e^{\frac{16(p_0-1)}{(p_0-2)}}$

Table 4.10: Dependence of minimum simulation threshold T^* on precision ε

Our results also enable us to study the convergence complexity of the FVSR and guarantee its validity in high-dimensional settings. Moreover, it can be guaranteed that the termination time scales polynomially in dimension while ensuring a desired level of precision. In the high-dimensional setting, we need to impose an additional multiplicative factor on the simulation threshold T^*. These factors are detailed in the table below.

	one-step minorisation	**multi-step minorisation**
geometric drift	$\left(\bar{\psi}_N \left(\frac{\mathrm{tr}(\Sigma_f)}{\sigma_0}\right)^2 d^3 \psi_d\right)^{\frac{2p}{(p-2)}(1+\bar{\delta}_1)}$	$\left(\bar{\psi}_N \left(\frac{\mathrm{tr}(\Sigma_f)}{\sigma_0}\right)^2 d^3 \psi_d\right)^{\frac{4(p-1)}{(p-2)}(1+\bar{\delta}_1)}$
polynomial drift	$\left(\bar{\psi}_N \left(\frac{\mathrm{tr}(\Sigma_f)}{\sigma_0}\right)^2 d^3 \psi_d\right)^{\frac{2p_0}{(p_0-2)}(1+\bar{\delta}_1)}$	$\left(\bar{\psi}_N \left(\frac{\mathrm{tr}(\Sigma_f)}{\sigma_0}\right)^2 d^3 \psi_d\right)^{\frac{4(p_0-1)}{(p_0-2)}(1+\bar{\delta}_1)}$

Table 4.11: Dimension dependence of the minimum simulation threshold T^*

4.5. Proofs of Main Results

4.5.1. Preliminary results on drift and moment conditions

In this section, we discuss the preliminary results that are required for our Gaussian approximation results. More specifically, we show how they follow from the assumed drift and minorisation conditions. We briefly the splitting procedure of Harris chains based on Asmussen [5], Meyn and Tweedie [107; Chapter 17.3], and Sigman and Wolff [141] and we discuss the implications of the assumed drift and minorisation conditions. Let X be an irreducible, aperiodic, positive Harris recurrent Markov chain taking values in Polish state space. From Meyn and Tweedie [107; Proposition 5.4.5] we know that X satisfies the following minorisation condition

$$P^{m_0}(x, dy) \geq \alpha \mathbb{1}_C(x) \nu(dy), \tag{4.34}$$

for some $\alpha \in (0, 1)$, $m_0 \in \mathbb{N}$, measurable set C with $\pi(C) > 0$, and probability measure ν that is equivalent to $\pi|_C$. Note that from (4.34) it follows that the transition kernel of the so-called m_0-skeleton chain, defined as $(X_{km_0})_{k \in \mathbb{N}}$, can be interpreted as a mixture of two transition kernels, namely

$$P^{m_0}(x, dy) = s(x)\nu(dy) + (1 - s(x))R(x, dy), \tag{4.35}$$

where $s(x) = \alpha \mathbb{1}_C(x)$ and the so-called residual kernel $R(x, dy)$ is defined as

$$R(x, dy) = \frac{P^{m_0}(x, dy) - s(x)\nu(dy)}{1 - s(x)}. \tag{4.36}$$

Given that the skeleton 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 R. Since the m_0-skeleton chain is also positive Harris recurrent, it will hit C infinitely often. By a Borel–Cantelli argument it follows that the chain will transition according to ν infinitely often. Let X' denote the split chain of the m_0-skeleton of X, i.e. for $n \in \mathbb{Z}_+$ we define $X'_n := (X_{nm_0}, \delta_n)$, where δ_n is a Bernoulli random variable that describes the distribution of the next point of the skeleton chain. The split chain has state space $\mathscr{X} \times \{0,1\}$ and has transition kernel

$$P'((x,\delta),(dy,d\delta')) = \begin{cases} P(x,dy)s(y)^{\delta'}(1-s(y))^{1-\delta'}, & x \notin C, \\ \nu(dy)s(y)^{\delta'}(1-s(y))^{1-\delta'} \quad , & x \in C;\ \delta = 1, \\ R(x,dy)s(y)^{\delta'}(1-s(y))^{1-\delta'}, & x \in C;\ \delta = 0. \end{cases} \tag{4.37}$$

Hence if $\delta_n = 1$, the next point of the skeleton chain has law ν and otherwise its law is described by the residual kernel. Let R_k denote the k-th time that the m_0-skeleton chain moves according to ν. The randomised stopping times $(R_k)_k$ serve as regeneration epochs for the skeleton chain, whereas they will be semi-regeneration epochs for the process X. We say that a process is semi-regenerative if there exists (by enlarging the probability space if necessary) a sequence of independent and identically distributed random variables (ρ_k) that define a renewal process (R_n) with $R_n = \sum_{k=1}^{n} \rho_k$ such that for each $n \geq 0$ the post-R_n process

$$\{(X_{R_n+k})_{k\geq 0}, R_{n+1}, R_{n+2}, \cdots\}$$

is independent of $R_0, \cdots, R_n$ and its distribution does not depend on n. Note that this implies that the process can be split into identically distributed cycles, where the lengths of the cycle are described by a renewal process. The classically regenerative definition would also impose the cycles to be independent. If the chain X satisfies a one-step minorization condition, i.e., (4.34) holds with $m_0 = 1$, that the process inherits a classical regenerative structure, whereas for the general case where $m_0 > 1$ the process has a semi-regenerative structure with one-dependent cycles. In order to see this, we first show how the state space can be enlarged to support the semi-regeneration times of the process. Let (δ_n) again denote a sequence of Bernoulli random variables which will describe the distribution of the m_0-skeleton points. Let $(\mathcal{F}_t^X)_t, (\mathcal{F}_t^\delta)_t$ denote the natural filtration of the process and the auxiliary

Bernoulli variables respectively. Consider the joint law of the chain in blocks of size m_0;

$$\begin{aligned}
&\mathbb{P}\left(\delta_n = 1, X_{nm_0+1} \in dx_1, \cdots, X_{(n+1)m_0-1} \in dx_{m_0-1}, , X_{(n+1)m_0} \in dy \mid \mathcal{F}^X_{nm_0}, \mathcal{F}^\delta_{n-1}; X_{nm_0}\right) \\
&= \mathbb{P}\left(\delta_n = 1, X_{nm_0+1} \in dx_1, \cdots, X_{(n+1)m_0-1} \in dx_{m_0-1}, X_{(n+1)m_0} \in dy \mid X_{nm_0}\right) \\
&= \alpha r(X_{nm_0}, y) P(X_{nm_0}, dx_1) \cdots P(x_{m_0-1}, dy), \qquad (4.38)
\end{aligned}$$

where all equalities hold almost surely and r denotes the Radon-Nykodym derivative

$$r(x, y) = \mathbb{1}_C(x) \frac{\nu(dy)}{P^{m_0}(x, dy)}. \qquad (4.39)$$

Note that we also have

$$\begin{aligned}
&\mathbb{P}(\delta_n = 1, X_{(n+1)m_0} \in dy \mid \mathcal{F}^X_{nm_0}, \mathcal{F}^\delta_{n-1}; X_{nm_0}) \\
&= \int_{x_1, \cdots, x_{m_0-1}} \mathbb{P}\left(\delta_n = 1, X_{nm_0+1} \in dx_1, \cdots, X_{(n+1)m_0-1} \in dx_{m_0-1}, X_{(n+1)m_0} \in dy \mid X_{nm_0}\right) \\
&= \int_{x_1, \cdots, x_{m_0-1}} \alpha r(X_{nm_0}, y) P(X_{nm_0}, dx_1) \cdots P(x_{m_0-1}, dy) \\
&= \alpha \mathbb{1}_C(X_{nm_0}) \frac{\nu(dy)}{P^{m_0}(X_{nm_0}, dy)} P^{m_0}(X_{nm_0}, dy) \\
&= \alpha \mathbb{1}_C(X_{nm_0}) \nu(dy),
\end{aligned}$$

where the third equality follows from the Chapman–Kolmogorov equations.

It easily follows that we also have

$$\mathbb{P}(\delta_n = 1 \mid \mathcal{F}^X_{nm_0}, \mathcal{F}^\delta_{n-1}; X_{nm_0}) = \alpha \mathbb{1}_C(X_{nm_0}) \qquad (4.40)$$

$$\mathbb{P}(X_{(n+1)m_0} \in dy \mid \mathcal{F}^X_{nm_0}, \mathcal{F}^\delta_{n-1}; X_{nm_0}, \delta_n = 1) = \nu(dy) \qquad (4.41)$$

From (4.40) and (4.41) we see that given $\delta_n = 1$ we have that

$\{X_k, \delta_i : k \leq nm_0, i \leq n\}$ is independent of $\{X_k, \delta_i : k \geq (n+1)m_0, i \geq n+1\}$.

Furthermore, we also have that the process $\{X_k, \delta_i : k \geq (n+1)m_0, i \geq n+1\}$ is equal in distribution to $\{(X_k, \delta_k) : k \geq 0\}$ with initial distribution

$$\mathbb{P}(X_0 \in dx_0, \delta_0 \in d\delta) = \nu(dx_0) \,\text{Bernoulli}(s(x))$$

Hence we see that the process X can be embedded in a richer process, which admits a recurrent atom $A := C \times \{1\}$ in the sense of the following proposition.

Proposition 4.5.1. *Let (S_n, R_n) be a sequence of stopping times defined as $S_0 = R_0 := 0$ and*

$$S_{n+1} := \inf\{km_0 > R_n : (X_{km_0}, \delta_k) \in C \times \{1\}\} \quad \text{and} \quad R_{n+1} := S_{n+1} + m_0.$$

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

$$(X_{R_n}, \delta_{R_n}) \sim \nu(dx)\,\text{Bernoulli}(s(x)) \quad \text{for all } n \geq 1,$$

the process is semi-regenerative, the cycles lengths ρ_k are independent and identically distributed, and the cycles

$$\{X_k : R_{n-1} \leq k < R_n\}$$

are identically distributed and one-dependent for all $n \in \mathbb{N}$. If $m_0 = 1$, then the cycles are in fact independent.

Proof. Sigman and Wolff [141; Theorem 4.2] and Asmussen [5; Theorem]. □

The stopping times $\{S_n\}_n$ thus denote the hitting times of the recurrent atom A and $\{R_n\}_n$ denote the implied regeneration epochs of the chain. As a direct consequence, of the semi-regenerative structure and the fact that R_n forms a renewal process, we obtain the following characterisation of the stationary measure. Moreover, introduce the following stopping time for the m_0–skeleton:

$$\bar{\tau}_{C\times\{1\}} = \inf\{k : (X_{km_0}, \delta_k) \in C \times \{1\}\} \tag{4.42}$$

It is well-known that the drift inequalities are closely related to the moments of hitting times of the process. For processes satisfying a geometric drift condition, the moment bounds for the hitting time $\bar{\tau}_{C\times\{1\}}$ given in Baxendale [9] are to the best of our knowledge the tightest bounds that are currently available.

Lemma 4.5.2 (Baxendale [9; Proposition 4.4])**.** *Let $(X_t)_{t\in\mathbb{N}}$ be an irreducible aperiodic Markov chain, assume that an m_0–step minorisation condition and drift condition 7 are satisfied. Then, for $1 < r \leq \lambda^{-1}$,*

$$\mathbb{E}_x[r^{\bar{\tau}_{C\times\{1\}}}] \leq \frac{\alpha G(r,x)}{1-(1-\alpha)r^a} \text{ and } \mathbb{E}_\nu[r^{\bar{\tau}_{C\times\{1\}}}] \leq \frac{\pi(V)}{(1-(1-\alpha)r^a)\alpha\pi(C)}, \tag{4.43}$$

where

$$a = 1 + \left(\log \frac{\lambda \upsilon_V + b - \alpha}{1 - \alpha} \right) / (\log(\lambda^{-1})),$$

and

$$G(r, x) \leq \begin{cases} V(x), & \text{for } x \in C \\ r(\lambda \upsilon_C + b), & \text{for } x \notin C. \end{cases} \tag{4.44}$$

Moreover, we have that

$$\pi(V) \leq \frac{b}{1 - \lambda} \pi(C). \tag{4.45}$$

Proof. The claim (4.43) is a combination of Baxendale [9; Proposition 4.4] and Lemma 4.5.12. The claim (4.45) is given in in Meyn and Tweedie [106; Proposition 4.3].

□

For chains satisfying a polynomial drift condition, we formulate the following result which gives us bounds on the moments of the semi-regeneration times.

Lemma 4.5.3. *Let $(X_t)_{t \in \mathbb{N}}$ be an irreducible aperiodic Markov chain, assume that an m_0–step minorisation condition and drift condition 8 are satisfied. Then, provided that $\eta > 1/2$, we can assume that drift condition 8 holds for some drift function V with $\pi(V) \leq \pi(C)b/c$ such that*

$$\mathbb{E}_x[\bar{\tau}^q_{C \times \{1\}}] \leq V(x) + (1 - \alpha)^{-1} (\upsilon_C - c + b) \, \mathbb{1}_C(x) \tag{4.46}$$

and

$$\mathbb{E}_\nu[\bar{\tau}^q_{C \times \{1\}}] \leq \frac{\pi(V)}{\alpha \pi(C)} + \frac{\upsilon_C - c + b}{1 - \alpha}, \tag{4.47}$$

with $q = \frac{\eta}{1 - \eta}$. Moreover,

$$\mathbb{E}_x[\tau_C] \leq \frac{1}{(1 - \eta)c} \left(V^{1-\eta}(x) + (b^\eta + b_0) \mathbb{1}_C(x) \right) \quad \text{and} \tag{4.48}$$

$$\mathbb{E}_\nu[\tau_C] \leq \frac{\pi(V)^{1-\eta}}{(1 - \eta)\alpha c \pi(C)} + \frac{(b^\eta + b_0)}{(1 - \eta)\alpha c}. \tag{4.49}$$

Proof. Firstly, by the Comparison theorem, Meyn and Tweedie [107; Theorem 14.2.2], we have that $\pi(V^\eta) \leq b\pi(C)/c$. From Jarner and Roberts [84; Lemma 3.5] we see that V^η is also a Lyapunov function satisfying drift condition

$$PV^\eta \leq V^\eta - c\eta V^{2\eta-1} + (b^\eta + b_0)\mathbb{1}_C,$$

for some $b_0 > 0$. Hence provided that $\eta > 1/2$, the first assertion follows. In order to use the hitting time bounds implied by a polynomial drift condition, we must first show that the split-chain of the m_0–skeleton also satisfies a polynomial drift condition. Define a Lyapunov function on the extended state space as follows: $V(x,0) = V(x,1) = V(x)$. Since the transition kernel of the split chain of the m_0–skeleton is given by (4.37), we have that for $x \notin C$, the Lyapunov condition is already satisfied. For $x \in C$ and $\delta = 1$ it immediately follows that

$$\begin{aligned}\hat{P}^{m_0}V((x,1)) &= \int_C V(y)\nu(dy) \\ &\leq V(x) - cV^\eta(x) + c\upsilon_C^\eta + \upsilon_C\end{aligned}$$

For $x \in C$ and $\delta = 0$ we have that

$$\begin{aligned}\hat{P}^{m_0}V((x,0)) &= \int_E V(y)R(x,dy) \\ &\leq (1-\alpha)^{-1}\int_E V(y)P^{m_0}(x,dy) \\ &\leq (1-\alpha)^{-1}\left(V(x) - cV^\eta(x) + b\right) \\ &\leq V(x) - cV(x)^\eta + c\upsilon_C^\eta + (1-\alpha)^{-1}\left(\upsilon_C - c + b\right)\end{aligned}$$

By Douc et al. [51; Proposition 2.2] we have that

$$\mathbb{E}_x[\bar{\tau}_C^q] \leq V(x) + (1-\alpha)^{-1}\left(\upsilon_C - c + b\right)\mathbb{1}_C(x)$$

By Lemma 4.5.12 it follows that

$$\begin{aligned}\mathbb{E}_\nu[\bar{\tau}_C^q] &= \frac{1}{\alpha\pi(C)}\mathbb{E}_\pi[\bar{\tau}_C^q] \\ &= \frac{1}{\alpha\pi(C)}\left(\pi(V) + (1-\alpha)^{-1}\left(\upsilon_C - c + b\right)\pi(C)\right)\end{aligned}$$

In order to show (4.48), we note that by Jarner and Roberts [84; Lemma 3.5]] $V^{1-\eta}$ is also a Lyapunov function satisfying drift condition

$$PV^{1-\eta} \leq V^{1-\eta} - c(1-\eta) + (b^{\eta} + b_0)\mathbb{1}_C.$$

By the Comparison theorem, Meyn and Tweedie [107; Theorem 14.2.2], we have that the first claim follows. The second part again follows from Lemma 4.5.12 and Jensen's inequality. □

Note that the bounds obtained in Lemma 4.5.3 are quite general and therefore for specific situations tighter bounds could be obtained, see for example Andrieu et al. [3] obtain quantitative bounds for subgeometric markov chains under the assumptions of a lower bound on the Lyapunov function V outside of C and a minorisation condition for all skeleton chains of the process. Andrieu et al. [3] show that if we assume a lower bound on the Lyapunov function: $\inf_{x \notin C} V(x) \geq b(1-\varepsilon)^{-1}$, for some $\varepsilon > 0$ and that a one-step minorisation condition holds for the original chain and all its skeletons, then the bounds presented in Lemma 4.5.3 can be improved.

Note that we can obtain a drift condition for the m_0–skeleton by iterating the one-step drift condition, namely,

$$P^{m_0}V(x) \leq \lambda^{m_0}V(x) + b\sum_{i=0}^{m_0-1} P^i\mathbb{1}_C(x) \leq \lambda V(x) + bm_0\mathbb{1}_{C(m_0)}(x),$$

where $C(m_0)$ is a small set for the skeleton chain and $C \subseteq C(m_0)$. Under additional regularity conditions, we can obtain a drift condition for the skeleton chain towards the set C.

Proposition 4.5.4. *Let $(X_t)_{t\in\mathbb{N}}$ be an irreducible aperiodic Markov chain, assume that an m_0–step minorisation condition holds and that drift condition 7 holds for some $C_0 \supsetneq C$ such that either there exists some $\bar{\alpha} > 0$ such that*

$$P^{m_0}(x, C) \geq \bar{\alpha}\mathbb{1}_{C_0\setminus C}(x) \quad \text{or} \quad \inf_{x\in C_0\setminus C} V(x) > b(1-\lambda)^{-1}. \tag{4.50}$$

Then there exists a function $\hat{V} : E \to \mathbb{R}^+$ and $\hat{\lambda} \in (0,1)$ and $\hat{b} > 0$ such that

$$P^{m_0}\hat{V}(x) \leq \hat{\lambda}\hat{V}(x) + \hat{b}\mathbb{1}_C(x).$$

Proof. Let $z_1 := \inf_{x \in C_0 \backslash C} V(x) > b(1-\lambda)^{-1}$. Then for $x \in C_0 \backslash C$ we have that

$$\begin{aligned} PV(x) &\leq V(x) - cV(x)^\eta + b \\ &\leq \hat{\lambda} V(x), \end{aligned}$$

provided that we choose $\hat{\lambda} = \frac{b + z_1 \lambda}{z_1}$. Note that since $\lambda < \hat{\lambda} < 1$, it follows immediately that for $x \in C$ we have $PV \leq \lambda V + b \leq \hat{\lambda} V + b$. Hence the desired claim immediately follows. Under the assumption $P^{m_0}(x, C) \geq \bar{\alpha} \mathbb{1}_{C_0 \backslash C}(x)$, the claim follows completely analogously to the proof of Meyn and Tweedie [106; Theorem 6.1]. □

Proposition 4.5.5. *Let $(X_t)_{t \in \mathbb{N}}$ be an irreducible aperiodic Markov chain, assume that an m_0–step minorisation condition holds and that drift condition 8 holds for some $C_0 \supsetneq C$ such that either there exists some $\bar{\alpha} > 0$ such that*

$$P^{m_0}(x, C) \geq \bar{\alpha} \mathbb{1}_{C_0 \backslash C}(x) \quad \text{or} \quad \inf_{x \in C_0 \backslash C} V(x) > (1 + b/c)^{1/\eta}. \tag{4.51}$$

Then there exists a function $\hat{V} : E \to \mathbb{R}^+$ and $\hat{\eta} \in (0, \eta)$ and $\hat{c} > 0$ such that

$$P^{m_0} \hat{V}(x) \leq \hat{V}(x) - \hat{c} \hat{V}^{\hat{\eta}}(x) + \hat{b} \mathbb{1}_C(x).$$

Proof. Under the assumption $P^{m_0}(x, C) \geq \bar{\alpha} \mathbb{1}_{C_0 \backslash C}(x)$, the claim follows completely analogously to the proof of Proposition 4.5.7. For the second case, we will take $\hat{c} = c$ and $\hat{V} = V$. Let $z_1 := \inf_{x \in C_0 \backslash C} V(x) > (1 + b/c)^{1/\eta}$ and $\hat{\eta} = \ln(z_1^\eta - b/c) / \ln(z_1)$. Then for $x \in C_0 \backslash C$ we have that

$$\begin{aligned} PV(x) &\leq V(x) - cV^\eta(x) + b \\ &\leq V(x) - cV^{\hat{\eta}(x)}, \end{aligned}$$

since we have that $\hat{\eta} < \eta$ and therefore

$$\frac{c}{V^{\eta - \hat{\eta}}(x)} + \frac{b}{V^\eta(x)} \leq \frac{c}{z_1^{\eta - \hat{\eta}}} + \frac{b}{z_1^\eta} \leq c,$$

which is equivalent to $cV^{\hat{\eta}}(x) + b \leq cV^\eta(x)$. □

Hence, Propositions 4.5.4 and 4.5.5 give conditions such that the one-step drift condition implies the desired drift condition for the skeleton chain. Furthermore, we note that in order for a drift condition towards C to hold, it is sufficient to show that it holds on some appropriate subset of C. This might be a useful property for T-chains, where every closed set is petite and hence a small set.

Proposition 4.5.6 (Meyn and Tweedie [106; Theorem 6.1])**.** *Let $(X_t)_{t\in\mathbb{N}}$ be an irreducible aperiodic Markov chain, assume that an m_0–step minorisation condition holds and that drift condition 7 holds for some $C_0 \subsetneq C$ with $\pi(C_0) > 0$. Then there exists a function $\hat{V} : E \to \mathbb{R}^+$ with $V \leq \hat{V} \leq V + b/\alpha\nu(C_0)$ such that*

$$P^{m_0}\hat{V}(x) \leq \hat{\lambda}\hat{V}(x) + \hat{b}\mathbb{1}_C(x),$$

with

$$\hat{\lambda} = \frac{\lambda\alpha\nu(C_0) + b}{\alpha\nu(C_0) + b} \text{ and } \hat{b} = b + b/\alpha\nu(C_0).$$

We can easily prove an analogous lemma for chains satisfying a polynomial drift condition.

Proposition 4.5.7. *Let $(X_t)_{t\in\mathbb{N}}$ be an irreducible aperiodic Markov chain, assume that an m_0–step minorisation condition holds and drift condition 8 is satisfied for some $C_0 \subseteq C$ with $\pi(C_0) > 0$, then there exist a function $\hat{V} : E \to \mathbb{R}^+$ with $V \leq \hat{V} \leq V + (1 \wedge b)/\alpha\nu(C)$ such that*

$$P^{m_0}\hat{V}(x) \leq \hat{V}(x) - \hat{c}V(x)^{\hat{\eta}} + \hat{b}\mathbb{1}_C(x),$$

with

$$\hat{c} \in \left(\frac{c}{(1 \wedge b)/\alpha\nu(C)^{\eta} + 1}, \frac{c}{2}\right),$$

$$\hat{b} = b + \frac{(1 \wedge b)}{\alpha\nu(C)}(1 - \alpha\nu(C)) \text{ and } \hat{\eta} = \frac{\ln\left(\frac{c-\hat{c}}{\hat{c}}\right)}{\ln((1 \wedge b)/\alpha\nu(C))}$$

Proof. For some B to be determined at a later point in the proof, let

$$\hat{V}(x) = \begin{cases} V(x), & x \in C \\ V(x) + B, & x \notin C. \end{cases}$$

Now we see that

$$\begin{aligned} P^{m_0}\hat{V}(x) &= \int_E \hat{V}(y)P^{m_0}(x, dy) = \int_C V(y)P^{m_0}(x, dy) + \int_{C^c}(V(y) + B)P^{m_0}(x, dy) \\ &= P^{m_0}V(x) + BP^{m_0}(x, C^c) \end{aligned}$$

For $x \notin C$ we have that

$$P^{m_0}\hat{V}(x) \leq V(x) - cV(x)^{\eta} + B$$

Hence if $B > 1$, $\hat{c} \in \left(\frac{c}{B^{\eta}+1}, \frac{c}{2}\right)$, and $\hat{\eta} \in \left(0, \ln\left(\frac{c-\hat{c}}{\hat{c}}\right)/\ln(B)\right]$, we see that $\hat{\eta} < \eta$ and thus

$$V^{\hat{\eta}-\eta} + \frac{B^{\hat{\eta}}}{V^{\eta}} \leq 1 + B^{\hat{\eta}} \leq \frac{c}{\hat{c}}.$$

Therefore it also follows that

$$(V+B)^{\hat{\eta}} \leq V^{\hat{\eta}} + B^{\hat{\eta}} \leq \frac{c}{\hat{c}} V^{\eta}.$$

Note that $1 + B^{\hat{\eta}} \leq c/\hat{c}$ gives $\hat{\eta}$ and the restriction $0 < \hat{\eta} < \eta$ gives the restriction on $\hat{c}$. It follows that

$$P^{m_0}\hat{V}(x) \leq \hat{V}(x) - \hat{c}\hat{V}(x)^{\hat{\eta}},$$

as desired. For $x \in C^c \cap C'$, we have that $P^{m_0}(x, C) \geq \alpha\nu(C)$ and hence

$$\begin{aligned} P^{m_0}\hat{V}(x) &= P^{m_0}V(x) + B(1 - P^{m_0}(x, C)) \\ &\leq V(x) - cV(x)^{\eta} + b + B - \alpha\nu(C)B \\ &\leq \hat{V}(x) - cV(x)^{\eta} \\ &\leq \hat{V}(x) - \hat{c}\hat{V}(x)^{\hat{\eta}}, \end{aligned}$$

given that we choose $B = (1 \wedge b)/\alpha\nu(C)$. For $x \in C$ we see that

$$\begin{aligned} P^{m_0}\hat{V}(x) &\leq V(x) - cV(x)^{\eta} + b + B(1 - \alpha\nu(C)) \\ &\leq \hat{V}(x) - \hat{c}\hat{V}(x)^{\hat{\eta}} + b + B(1 - \alpha\nu(C)) \end{aligned}$$

□

In Jones et al. [87] and Bednorz and Łatuszyński [10] it is shown that geometric ergodicity and moment conditions with respect to the stationary measure are sufficient to guarantee moment conditions of functionals over regenerative cycles. These results easily carry over to the continuous-time and multivariate setting as seen in Pengel and Bierkens [121] and Banerjee and Vats [8] respectively. We generalise these results using the explicit bounds given in Lemma 4.5.2 and 4.5.3. We note that the proofs essentially follow the argument of Bednorz and Łatuszyński [10; Lemma 2] coordinate-wise. These results are of independent interest, see for example Bertail and Ciołek [14], who assume that explicit bounds for moments of hitting times and blocks are given. Lemma 4.5.8 and 4.5.9 could for example be used to obtain dimension-dependent Bernstein inequalities, see Bertail and Ciołek [14]. Note that all the constants C_0, C_1 and C_3 in Lemmas 4.5.8 and 4.5.9 can be given explicitly.

Lemma 4.5.8. *Let $(X_t)_{t\in\mathbb{N}}$ be an irreducible aperiodic Markov chain, assume that an m_0-step minorisation condition holds and that drift condition 7 is satisfied. Then for any t with $|t| \leq \ln(1/\lambda)/m_0$ we have that*

$$\mathbb{E}_\nu[e^{tR_1}] \lesssim \frac{b}{\alpha\lambda(1-\lambda)} \tag{4.52}$$

Moreover, under Assumption A4.1, we have that

$$\sup_{i\in\{1,\dots,d\}} \mathbb{E}_\nu\left[\left(\sum_{t=0}^{R_1}|f_i(X_t)|\right)^p\right] \lesssim \alpha^{-1}\left(\frac{b}{\alpha\lambda(1-\lambda)}\right)^{\varepsilon/p} \sup_{i\in\{1,\dots,d\}} \pi(|f_i|^{p+\varepsilon}), \tag{4.53}$$

and

$$\mathbb{E}_\nu\left[\left|\sum_{t=0}^{R_1} f(X_t)\right|^p\right] \lesssim \alpha^{-1} d^{p/2}\left(\mathbb{E}_\nu e^{tR_1}\right)^{\varepsilon/p} \sup_{i\in\{1,\dots,d\}} \pi(|f_i|^{p+\varepsilon}), \tag{4.54}$$

If we can additionally assume that Assumption A4.3 holds, we also have that

$$\sup_{i\in\{1,\dots,d\}} \mathbb{E}_\nu\left[\exp\left(\sum_{t=0}^{R_1}|f_i(X_t)|\right)\right] \lesssim \mathbb{E}_\nu[e^{tR_1}]. \tag{4.55}$$

Lemma 4.5.9. *Let $(X_t)_{t\in\mathbb{N}}$ be an irreducible aperiodic Markov chain, assume that an m_0–step minorisation condition holds and that drift condition 8 is satisfied for some Lyapunov function. Then*

$$\mathbb{E}_\nu[R_1^q] \lesssim 2^{q-1} m_0^q\left(1 + \frac{b}{c\alpha} + \frac{v_c - c + b}{1-\alpha}\right). \tag{4.56}$$

where $q(\eta) = \frac{\eta}{1-\eta}$. Moreover, for all $f : E \to \mathbb{R}^d$ such that Assumption A4.1 holds, we have that

$$\sup_{i\in\{1,\dots,d\}} \mathbb{E}_\nu\left[\left(\sum_{t=0}^{R_1}|f_i(X_t)|\right)^{p_0}\right] \lesssim \alpha^{-1}\left(\mathbb{E}_\nu R_1^q\right)^{\varepsilon/p_0} \sup_{i\in\{1,\dots,d\}} \pi(|f_i|^{p+\varepsilon}), \tag{4.57}$$

where

$$p_0 = \begin{cases} \frac{pq(\eta)}{p+q(\eta)+\varepsilon}, & \text{if } \frac{p}{2p-1} < \eta \leq p(p+\varepsilon)/(p(p+\varepsilon)+\varepsilon), \\ p, & \text{if } \eta > p(p+\varepsilon)/(p(p+\varepsilon)+\varepsilon), \\ q(\eta) - \bar{\epsilon}, & \text{if } \eta > 1/2 \quad \text{and A4.2 holds}, \end{cases} \tag{4.58}$$

for any fixed $\bar{\epsilon} \in \left(0, \min\{\frac{1}{2}, \frac{2\eta-1}{1-\eta}\}\right)$.

Proof of Lemma 4.5.8

Proof. Firstly, we show that R_1 admits an exponential moment in some neighborhood of zero. Note that from Proposition 4.5.1 we see that $R_1 = m_0 + m_0\bar{\tau}_{C\times\{1\}}$. Hence for any t with $|t| < \frac{\ln(1/\lambda)}{m_0}$ we have that

$$\begin{aligned}\mathbb{E}_\nu[e^{tR_1}] &= \mathbb{E}_\nu[e^{t(m_0+m_0\bar{\tau}_{C\times\{1\}}))}] \\ &\leq \mathbb{E}_\nu[e^{\frac{\ln(1/\lambda)}{m_0}(m_0+m_0\bar{\tau}_{C\times\{1\}}))}] \leq \frac{1}{\lambda}\mathbb{E}_\nu\left[\left(\frac{1}{\lambda}\right)^{\bar{\tau}_{C\times\{1\}}}\right]\end{aligned}$$

By an application of Lemma 4.5.2 we obtain that

$$\mathbb{E}_\nu[e^{tR_1}] \leq \frac{b}{\alpha\lambda(1-\lambda)}$$

Note that

$$\sum_{t=0}^{R_1} f(X_t) = \left(\sum_{t=0}^{R_1} f_1(X_t), \ldots, \sum_{t=0}^{R_1} f_d(X_t)\right)^T,$$

with $f_i : E \to \mathbb{R}$ for $i = 1, \ldots, d$. By Lemma 4.5.12 and a coordinate-wise application of Bednorz and Łatuszyński [10; Lemma 2], we see for all $\imath = 1, \ldots, d$ we have that

$$\begin{aligned}\mathbb{E}_\nu\left[\left(\sum_{t=0}^{R_1}|f_i(X_t)|\right)^p\right] &\leq (\alpha\pi(C))^{-1}\mathbb{E}_\pi\left[\left(\sum_{t=0}^{R_1}|f_i(X_t)|\right)^p\right] \\ &\leq (\alpha\pi(C))^{-1}\pi(|f_i|^{p+\varepsilon})^{\frac{p}{p+\varepsilon}}\left(\sum_{k=1}^{\infty}\mathbb{P}_\pi\left(R_1 > k\right)^{\frac{\varepsilon}{p(p+\varepsilon)}}\right)^p\end{aligned}$$

By an application of Markov's inequality we have that

$$\begin{aligned}&\leq (\alpha\pi(C))^{-1}\pi(|f_i|^{p+\varepsilon})^{\frac{p}{p_0+\varepsilon}}\left(\sum_{k=1}^{\infty}e^{-\frac{t\varepsilon}{p(p+\varepsilon)}k}\right)^p\left(\mathbb{E}_\pi e^{tR_1}\right)^{\varepsilon/(p+\varepsilon)} \\ &\leq (\alpha\pi(C))^{-1}\pi(|f_i|^{p+\varepsilon})^{\frac{p}{p+\varepsilon}}\left(e^{t\varepsilon/p(p+\varepsilon)}-1\right)^{-p}\left(\frac{b}{\alpha\lambda(1-\lambda)}\right)^{\varepsilon/p}.\end{aligned}$$

The second claim follows directly from the equivalence of norms,

$$\begin{aligned}
\mathbb{E}_\nu\left[\left|\sum_{t=0}^{R_1} f(X_t)\right|^p\right] &= \mathbb{E}_\nu\left[\left(\sum_{i=1}^{d}\left(\sum_{t=0}^{R_1}|f_i(X_t)|\right)^2\right)^{p/2}\right] \\
&\leq \mathbb{E}_\nu\left[\left(d^{1/2-1/p}\left(\sum_{i=1}^{d}\left|\sum_{t=0}^{R_1}|f_i(X_t)|\right|^p\right)^{1/p}\right)^p\right] \\
&= d^{p/2-1}\mathbb{E}_\nu\left[\sum_{i=1}^{d}\left|\sum_{t=0}^{R_1}|f_i(X_t)|\right|^p\right] = d^{p/2-1}\left[\sum_{i=1}^{d}\mathbb{E}_\nu\left|\sum_{t=0}^{R_1}|f_i(X_t)|\right|^p\right] \\
&\leq d^{p/2}\sup_{i\in\{1,\ldots,d\}}\mathbb{E}_\nu\left|\sum_{t=0}^{R_1}|f_i(X_t)|\right|^p,
\end{aligned}$$

which gives (4.60).

□

Proof of Lemma 4.5.9

Proof. Firstly, we show that R_1 has an exponential moment. From Proposition 4.5.1 we see that

$$\mathbb{E}_\nu R_1^q = \mathbb{E}_\nu[(m_0 + m_0\bar{\tau}_{C\times\{1\}})^q] \leq 2^{q-1}(m_0^q + m_0^q\mathbb{E}_\nu\bar{\tau}_{C\times\{1\}}^q),$$

where the first equality follows since $S_1 = m_0\bar{\tau}_{C\times\{1\}}$. The first claim (4.56) now follows for drift conditions 7 and 8 by an application of Lemma 4.5.2 and Lemma 4.5.3 respectively. Note that

$$\sum_{t=0}^{R_1} f(X_t) = \left(\sum_{t=0}^{R_1} f_1(X_t), \ldots, \sum_{t=0}^{R_1} f_d(X_t)\right)^T,$$

with $f_i : E \to \mathbb{R}$ for $i = 1, \ldots, d$. By Lemma 4.5.12 we see that for any $p_0 \leq p$ and for all $i = 1, \ldots, d$ we have that

$$\mathbb{E}_\nu\left[\left(\sum_{t=0}^{R_1}|f_i(X_t)|\right)^{p_0}\right] \leq (\alpha\pi(C))^{-1}\mathbb{E}_\pi\left[\left(\sum_{t=0}^{R_1}|f_i(X_t)|\right)^{p_0}\right]$$

Following the argument of Bednorz and Łatuszyński [10; Lemma 2] for every coordinate, we see that

$$\mathbb{E}_\pi\left[\left(\sum_{t=0}^{R_1}|f_i(X_t)|\right)^{p_0}\right] \leq \pi(|f_i|^{p_0+\varepsilon})^{\frac{p_0}{p+\varepsilon}}\left(\sum_{k=1}^{\infty}\mathbb{P}_\pi\left(R_1>k\right)^{(1-\frac{p_0}{p+\varepsilon})}\right)^{p_0}$$
$$\leq \pi(|f_i|^{p_0+\varepsilon})^{\frac{p_0}{p_0+\varepsilon}}\left(\sum_{k=1}^{\infty}k^{-\frac{q(\eta)}{p_0}\left(1-\frac{p_0}{p+\varepsilon}\right)}\right)^{p_0}\left(\mathbb{E}_\nu R_1^{q(\eta)}\right)^{\left(1-\frac{p_0}{p+\varepsilon}\right)}.$$

Note that in each of the following cases the series $\sum_{k=1}^{\infty}k^{-\frac{q(\eta)}{p_0}\left(1-\frac{p_0}{p+\varepsilon}\right)}$ converges. Firstly, consider $\frac{p}{2p-1}<\eta\leq p(p+\varepsilon)/(p(p+\varepsilon)+\varepsilon)$ we have that $q(\eta)\leq p(p+\varepsilon)/\varepsilon$ and hence for a given value of p, the largest value p_0 such that the series converges is $p_0=\frac{pq(\eta)}{p+q(\eta)+\varepsilon}$. Note that impose $\eta>\frac{p}{2p-1}$ in order to have $p_0>1$. We also see that if $\eta>p(p+\varepsilon)/(p(p+\varepsilon)+\varepsilon)$, then $q(\eta)>p(p+\varepsilon)/\varepsilon$ and we can take $p_0=p$. Finally, we see that under Assumption 4.2, we can take $p>q(q-\bar{\epsilon})/\bar{\epsilon}$, since f has moments of all orders with respect to π, and hence for any $\bar{\epsilon}>0$ such that $q(\eta)-\bar{\epsilon}>1$ we can take $p_0=q(\eta)-\bar{\epsilon}$. The remaining claims follow completely analogously to Lemma 4.5.8.

□

Lemma 4.5.10. *(Marcinkiewicz—Zygmund inequality; De la Pena and Giné [48; Lemma 1.4.13.]) Suppose $z_1,\ldots,z_d$ are one-dimensional independent zero-mean random variables with finite p-th moment, then there exist constants c_p and C_p such that*

$$c_p\mathbb{E}\left[\left(\sum_i z_i^2\right)^{p/2}\right]\leq\mathbb{E}\left[\left|\sum_i z_i\right|^p\right]\leq C_p\mathbb{E}\left[\left(\sum_i z_i^2\right)^{p/2}\right]$$

By the Marcinkiewicz—Zygmund inequality, we can describe the growth of the moments of sums in terms of the bounds given in Lemmata 4.5.8 and 4.5.9.

Lemma 4.5.11. *Let $\{\xi_k\}_{k\in\mathbb{N}}$ be d-dimensional one-dependent zero-mean random vectors with*

$$\sup_{k\in\mathbb{N}}\sup_{i\in\{1,\cdots,d\}}\mathbb{E}[|\xi_{k,i}|^p]\leq C_\xi,$$

then we have that

$$\mathbb{E}\left[\left|\sum_{k=0}^{\lfloor m/2\rfloor}\xi_{2k}\right|^p\right] \lesssim C_\xi\ d^{p/2}m^{p/2}.$$

Proof. Let $Z := (z_1,\dots,z_d)^T := \sum_{k=0}^{\lfloor m/2\rfloor}\xi_{2k}$ which is a d-dimensional random vector. Note that

$$|Z| := \sqrt{\sum_{k=1}^{d} z_i^2}$$

By the equivalence of norms we see that

$$\begin{aligned}\mathbb{E}\left[|Z|^p\right] &= \mathbb{E}\left[\left(\sum_{i=1}^{d} z_i^2\right)^{p/2}\right] = \mathbb{E}\left[\left(\left(\sum_{i=1}^{d} z_i^2\right)^{1/2}\right)^{p}\right]\\ &\le \mathbb{E}\left[\left(d^{1/2-1/p}\left(\sum_{i=1}^{d}|z_i|^p\right)^{1/p}\right)^{p}\right] = d^{p/2-1}\mathbb{E}\left[\sum_{i=1}^{d}|z_i|^p\right] \le d^{p/2}\sup_{\{i\in 1,\cdots,d\}}\mathbb{E}[|z_i|^p]\end{aligned}$$

Now note that $z_i = \sum_{k=0}^{\lfloor m/2\rfloor}\xi_{2k,i}$ is a sum of one-dimensional independent random variables. From an application of the Marcinkiewicz—Zygmund inequality, see Lemma 4.5.10, and the equivalence of norms we obtain

$$\begin{aligned}\mathbb{E}[|z_i|^p] &= \mathbb{E}\left[\left|\sum_{k=0}^{\lfloor m/2\rfloor}\xi_{2k,i}\right|^p\right] \le C_p\mathbb{E}\left[\left(\sum_{k=0}^{\lfloor m/2\rfloor}\xi_{2k,i}^2\right)^{p/2}\right]\\ &= C_p\mathbb{E}\left[\left(\left(\sum_{k=0}^{\lfloor m/2\rfloor}\xi_{2k,i}^2\right)^{1/2}\right)^{p}\right] \le C_p\mathbb{E}\left[\left(m^{1/2-1/p}\left(\sum_{k=0}^{\lfloor m/2\rfloor}|\xi_{2k,i}|^p\right)^{1/p}\right)^{p}\right]\\ &= C_p m^{p/2-1}\mathbb{E}\left[\sum_{k}|\xi_{2k,i}|^p\right] \le C_p m^{p/2}\sup_{k}\mathbb{E}[|\xi_{2k,i}|^p].\end{aligned}$$

Overall, we have that

$$\sup_{i\in\{1,\cdots,d\}}\mathbb{E}[|z_i|^p] \lesssim m^{p/2}C_\xi,$$

and hence the claim follows. □

4.5.2. Auxiliary results

Note that an immediate consequence of Minorisation condition 2 is that the small measure ν is absolutely continuous with respect to the stationary measure π. Furthermore, from the minorisation condition it also follows that we can bound the expectations with respect to the measures as formulated in Hobert et al. [81; Lemma 1].

Lemma 4.5.12 (Hobert et al. [81; Lemma 1]). *Let $(X_t)_{t\in T}$ be a positive Harris recurrent Markov process with invariant distribution π. Then for any π-integrable function $g : E^T \to \mathbb{R}^d$ we have the following inequality holds*

$$\mathbb{E}_\pi|g| \geq \alpha\pi(C)\ \mathbb{E}_\nu|g|, \tag{4.59}$$

where α and C are defined in (4.34).

Note that Lemma 4.5.12 was stated by Hobert et al. [81; Lemma 1] in a one-dimensional setting, and a one-step minorisation condition. However, the claim can mutatis mutandis be stated for the general multidimensional m_0–step local minorisation case.

4.5.3. Preliminary results on Gaussian approximation

Firstly, we will consider the following weak Gaussian approximation for bounded independent random vectors.

Lemma 4.5.13 (Eldan et al. [63; Theorem 1]). *Suppose $x_1, \ldots, x_n \in \mathbb{R}^d$ are independent identically distributed mean-zero bounded random vectors, i.e., $|x_i| \leq \tau$, for $i = 1, \ldots, n$ for some $\tau > 0$. Let σ_d denote the maximal eigenvalue of $\mathrm{Cov}(\frac{1}{\sqrt{n}}\sum_i x_i)$. Then one can construct independent random vectors $(x_i^c)_i$ and $(y_i)_i$ in a richer probability space such that $x_i^c \overset{D}{=} x_i$ and $Y_n \sim \mathcal{N}_d(0, \mathrm{Cov}(\sum x_i))$ such that for all $\theta_0 > 0$ we have that,*

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\sum_{i=1}^n x_i - Y_n\right| \leq \tau d^{1/2}(32 + \log(n))\right) = 1,$$

Note that this lemma is an immediate consequence of Eldan et al. [63; Theorem 1] and the fact that convergence in the Wasserstein metric implies convergence of moments, see for example Villani et al. [153; Theorem 6.9]. Furthermore, note that we also have the following weak coupling inequality, which is a corollary of either Zaitsev [158; Theorem 1.1] or Zaitsev [157; Theorem 1.1].

Lemma 4.5.14 (Einmahl and Mason [61; Fact 2.2]). *Suppose $x_1, \ldots, x_n \in \mathbb{R}^d$ are independent random vectors that $\mathbb{E}[x_i] = 0$ and $x_i = (x_{i,1}, \ldots, x_{i,d})^T$, in which $|x_i|$ are bounded such that $|x_i| \leq \tau, i = 1, \ldots, n$. Let $X_n = \sum_{k=1}^n x_k$, then one can construct X_n^c and Y_n^c in a richer probability space such that $x_n^c \stackrel{D}{=} x_n$ and for all $\epsilon \geq 0$,*

$$\mathbb{P}\left(|X_n^c - Y_n^c| > \epsilon\right) \leq C_2 d^2 \exp\left(-\frac{\epsilon}{C_3 d^2 \tau}\right),$$

where Y_n^c is sum of n i.i.d Gaussian vectors with the same mean and covariance matrix as X_n^c and C_2, C_3 are positive dimension-independent constants.

Note that Zaitsev [158; Theorem 1.1] and Zaitsev [157; Theorem 1.1] give us a dimension dependence of $d^{5/2}$ and d^2 respectively in Lemma 4.5.14. As shown in Einmahl and Mason [61; Fact 2.2], both results apply to bounded vectors. Furthermore, we will also consider the following strong Gaussian approximation for bounded independent random vectors.

Lemma 4.5.15 (Zaitsev [160; Corollary 3]). *Suppose $x_1, \ldots, x_n \in \mathbb{R}^d$ are independent mean-zero bounded random vectors, i.e., $|x_i| \leq \tau$, for $i = 1, \ldots, n$ for some $\tau \geq 1$. Let σ_d denote the maximal eigenvalue of $\mathrm{Cov}(\frac{1}{\sqrt{n}} \sum_i x_i)$. Then one can construct independent random vectors $(x_i^c)_i$ and $(y_i)_i$ in a richer probability space such that $x_i^c \stackrel{D}{=} x_i$ and $y_i \sim \mathcal{N}_d(0, \mathrm{Cov}(x_i))$ such that for all $\theta_0 > 0$ we have that,*

$$\mathbb{P}\left(\limsup_{n\to\infty} \frac{1}{\log(n)} \left|\sum_{i=1}^n x_i^c - \sum_{i=1}^n y_i\right| \leq K_{\theta_0} \tau d^{23/4+\theta_0} \sqrt{\sigma_d} \log^*(d)\right) = 1,$$

where K_{θ_0} depends only on θ_0 and $\log^*(d) := \max(1, \log(d))$.

Lemma 4.5.16 (Merlevède et al. [103; Lemma 2.4]). *Let B be a standard Brownian motion and L be a Poisson process with intensity λ, independent of B. Then there exists a standard Brownian motion W that is also independent of L such that*

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

Proof. The claim immediately follows from Merlevède et al. [103; Lemma 2.4] and a Borel–Cantelli argument. □

Lemma 4.5.17 (Ando–van Hemmen's inequality; van Hemmen and Ando [150; Corollary 4.2]). *Let Σ_1, Σ_2 be two positive definite matrices with the smallest eigenvalue bounded below by $\sigma_0 > 0$. Then for every $0 < r \leq 1$ we have that*

$$|\Sigma_1^r - \Sigma_2^r| \leq \left(\frac{1}{\sigma_0}\right)^{1-r} |\Sigma_1 - \Sigma_2|.$$

We will show that the strong approximation for bounded vectors given in Lemma 4.5.15 can be extended to independent vectors with only p finite moments. Furthermore, we obtain as an immediate consequence also the corresponding weak Gaussian approximation. The weak Gaussian approximation result for independent random vectors is comparable to the result of [108], which they obtain through martingale embeddings.

Theorem 4.5.18. *Let $\{\xi_k\}$ be a sequence of independent and identically distributed mean zero random vectors in $\mathbb{R}^d$ such that $\sup_{k\in\mathbb{N}} \sup_{i\in\{1,\cdots,d\}} \mathbb{E}[|\xi_{k,i}|^p] \leq C_\xi$, for some finite C_ξ, and for some $p > 2$ and such that the smallest eigenvalue of $\mathrm{Cov}(\frac{1}{\sqrt{n}} \sum_{k=1}^n \xi_k)$ is larger than some constant $\sigma_0 > 0$. Then*

$$\left|\sum_{k=1}^n \xi_k - \sum_{k=1}^n Y_k\right| = \begin{cases} \mathcal{O}_P\left(d^{3/2} C_\xi^{2/p} n^{1/p} \log(n)\right) \\ \\ \mathcal{O}_{a.s.}\left(\left(d^{25/4+\theta_0} \log^*(d) C_\xi^{1/p} \vee d^{3/2} C_\xi^{2/p}\right) \left(\frac{\sigma_d}{\sigma_0}\right)^{1/2} n^{1/p} \log(n)\right) \end{cases},$$

where $\sum_{k=1}^n Y_k$ has a Gaussian distribution with the same mean and covariance matrix as $\sum_{k=1}^n \xi_k$.

Proof. We will first prove the almost sure bound. Let $\tau_n := n^{1/p}, \delta = 1/2$ and define the truncated sequence $\tilde{\xi}_k$ as follows

$$\tilde{\xi}_k := \xi_k \mathbb{1}_{\{|\xi_k| \leq d^\delta C_\xi^{1/p} \tau_n\}} - \mathbb{E}[\xi_k \mathbb{1}_{\{|\xi_k| \leq d^\delta C_\xi^{1/p} \tau_n\}}].$$

Clearly, we have $|\tilde{\xi}_k| \leq 2d^{1/2}\tau_n$ and $\mathbb{E}[\tilde{\xi}_k] = 0$. Then by triangle inequality, we have

$$\left|\sum_{j=1}^n \xi_j - \sum_{j=1}^n \tilde{\xi}_j\right| \leq \left|\sum_{j=1}^n \xi_j \mathbb{1}_{\{|\xi_j| > d^\delta C_\xi^{1/p} \tau_n)\}}\right| + \left|\sum_{j=1}^n \mathbb{E}[\xi_j \mathbb{1}_{\{|\xi_j| > d^\delta C_\xi^{1/p} \tau_n\}}]\right|.$$

Since $\sup_k \sup_i \mathbb{E}[|\xi_{k,i}|^p]] \leq C_\xi$ it follows from the equivalence of vector norms that

$$\sup_k \mathbb{E}[|\xi_k|^p] \lesssim d^{p/2} \sup_k \sup_i \mathbb{E}[|\xi_{k,i}|^p] \lesssim d^{p/2} C_\xi. \tag{4.60}$$

For $r \in [1, p)$ we have that

$$\begin{aligned}
\mathbb{E}|\xi_j|^r \mathbb{1}_{|\xi_j| > d^\delta C_\xi^{1/p} \tau_n} &= r \int_{d^\delta C_\xi^{1/p} \tau_n}^\infty s^{r-1} \mathbb{P}(|\xi_j| > s) ds + d^{\delta r} C_\xi^{r/p} \tau_n^r \mathbb{P}(|\xi_j| > d^\delta C_\xi^{1/p} \tau_n) \\
&\leq r \int_{d^\delta C_\xi^{1/p} \tau_n}^\infty \frac{\mathbb{E}|\xi_j|^p}{s^{p-r+1}} ds + d^{\delta r} C_\xi^{r/p} \tau_n^r \frac{\mathbb{E}|\xi_j|^p}{d^{p\delta} \tau_n^p C_\xi} \\
&= \frac{r}{(p-r)} \frac{\mathbb{E}|\xi_j|^p}{(d^\delta C_\xi^{1/p} \tau_n)^{(p-r)}} + d^{\delta r} C_\xi^{r/p} \tau_n^r \frac{\mathbb{E}|\xi_j|^p}{d^{p\delta} \tau_n^p C_\xi} \\
&\lesssim d^{\delta r} \tau_n^{r-p} C_\xi^{r/p}.
\end{aligned} \tag{4.61}$$

Taking $r = 1$ in (4.61) it follows that $\left(|\xi_j| \mathbb{1}_{\{|\xi_j| > d^\delta C_\xi^{1/p} \tau_n\}} \right)_{j \in \mathbb{N}}$ is an i.i.d. sequence with mean $\mu_{n,d} > 0$ such that $\mu_{n,d} \lesssim d^\delta \tau_n^{1-p} C_\xi^{r/p}$. We claim that

$$\sum_{j=1}^n |\xi_j| \mathbb{1}_{\{|\xi_j| > d^\delta C_\xi^{1/p} \tau_n)\}} = \mathcal{O}_{a.s.}(d^\delta C_\xi^{1/p} n \tau_n^{1-p}) \tag{4.62}$$

Note that since $d^{-\delta} C_\xi^{-1/p} \mu_{n,d}$ tends to zero for any τ_n that tends to infinity, the claim does not follow by a standard law of large numbers for triangular arrays. We will prove (4.62) by following a truncation and sub-sequence argument. Firstly, note that for any $\theta > 1/p$ we have that

$$\sum_{j=1}^n |\xi_j| \mathbb{1}_{\{|\xi_j| > d^\delta C_\xi^{1/p} \tau_n)\}} = \sum_{j=1}^n |\xi_j| \mathbb{1}_{\{d^\delta C_\xi^{1/p} \tau_n \leq |\xi_j| \leq d^\delta C_\xi^{1/p} n^\theta)\}} + \sum_{j=1}^n |\xi_j| \mathbb{1}_{\{|\xi_j| > d^\delta C_\xi^{1/p} n^\theta)\}}. \tag{4.63}$$

For the second term in (4.63), we see that if we choose

$$\theta > \frac{2p-1}{p(p-1)}, \tag{4.64}$$

then for every $\varepsilon > 0$ we have that

$$\sum_{n=1}^{\infty} \mathbb{P}\left(\left| \sum_{j=1}^{n} |\xi_j| \mathbb{1}_{\{|\xi_j| > d^{\delta} C_{\xi}^{1/p} n^{\theta}\}} \right| > \varepsilon d^{\delta} C_{\xi}^{1/p} n^{1/p} \right) \leq \sum_{n=1}^{\infty} \frac{n \mathbb{E}[|\xi_j| \mathbb{1}_{\{|\xi_j| > d^{\delta} C_{\xi}^{1/p} n^{\theta}\}}]}{\varepsilon d^{\delta} C_{\xi}^{1/p} n^{1/p}}$$
$$\lesssim \frac{1}{\varepsilon} \sum_{n=1}^{\infty} n^{\theta(1-p)+1-1/p} < \infty,$$

since we have mutatis mutandis to (4.61) that $\mathbb{E}[|\xi_j| \mathbb{1}_{\{|\xi_j| > d^{\delta} C_{\xi}^{1/p} n^{\theta}\}}] \lesssim d^{\delta} C_{\xi}^{1/p} n^{\theta(1-p)}$ for $p > 2$. Hence by the Borel–Cantelli Lemma we have that

$$\sum_{j=1}^{n} |\xi_j| \mathbb{1}_{\{|\xi_j| > d^{\delta} C_{\xi}^{1/p} n^{\theta})\}} = \mathcal{O}_{a.s.}(d^{\delta} C_{\xi}^{1/p} n^{1/p}). \tag{4.65}$$

Note that the first term in the RHS of (4.63) is bounded above by

$$\left| \sum_{j=1}^{n} \left(|\xi_j| \mathbb{1}_{\{d^{\delta} C_{\xi}^{1/p} \tau_n \leq |\xi_j| \leq d^{\delta} C_{\xi}^{1/p} n^{\theta})\}} - \mathbb{E}\left[|\xi_j| \mathbb{1}_{\{d^{\delta} C_{\xi}^{1/p} \tau_n \leq |\xi_j| \leq d^{\delta} C_{\xi}^{1/p} n^{\theta})\}} \right] \right) \right|$$
$$+ \sum_{j=1}^{n} \mathbb{E}\left[|\xi_j| \mathbb{1}_{\{d^{\delta} C_{\xi}^{1/p} \tau_n \leq |\xi_j| \leq d^{\delta} C_{\xi}^{1/p} n^{\theta})\}} \right] \tag{4.66}$$

From (4.61) it also follows that

$$\sum_{j=1}^{n} \mathbb{E}\left[|\xi_j| \mathbb{1}_{\{d^{\delta} C_{\xi}^{1/p} \tau_n \leq |\xi_j| \leq d^{\delta} C_{\xi}^{1/p} n^{\theta})\}} \right] = \mathcal{O}(d^{\delta} C_{\xi}^{1/p} n \tau_n^{1-p}) \tag{4.67}$$

Now we will show that

$$\left| \sum_{j=1}^{n} \left(|\xi_j| \mathbb{1}_{\{d^{\delta} C_{\xi}^{1/p} \tau_n \leq |\xi_j| \leq d^{\delta} C_{\xi}^{1/p} n^{\theta})\}} - \mathbb{E}\left[|\xi_j| \mathbb{1}_{\{d^{\delta} C_{\xi}^{1/p} \tau_n \leq |\xi_j| \leq d^{\delta} C_{\xi}^{1/p} n^{\theta})\}} \right] \right) \right| = o_{a.s.}(d^{\delta} C_{\xi}^{1/p} n^{1/p}). \tag{4.68}$$

Which is equivalent to showing that

$$\left| \frac{1}{n^{1/p} C_{\xi}^{1/p} d^{\delta}} \sum_{j=1}^{n} \left(|\xi_j| \mathbb{1}_{\{d^{\delta} C_{\xi}^{1/p} \tau_n \leq |\xi_j| \leq d^{\delta} C_{\xi}^{1/p} n^{\theta})\}} - \mathbb{E}\left[|\xi_j| \mathbb{1}_{\{d^{\delta} C_{\xi}^{1/p} \tau_n \leq |\xi_j| \leq d^{\delta} C_{\xi}^{1/p} n^{\theta})\}} \right] \right) \right| \tag{4.69}$$

tends to zero almost surely as $n \to \infty$. We will first show that (4.69) holds for a subsequence. Fix any $\lambda > 1$ and define $n(\ell) = \lfloor \lambda^\ell \rfloor$. Then, we claim that as $\ell \to \infty$, we have that

$$\left| \frac{1}{n(\ell)^{1/p} C_{\xi,d}^{1/p} d^\delta} \sum_{j=1}^{n(\ell)} \left(|\xi_j| \mathbb{1}_{\{d^\delta C_{\xi,d}^{1/p} \tau_{n(\ell)} \le |\xi_j| \le d^\delta C_{\xi,d}^{1/p} n(\ell)^\theta)\}} - \mathbb{E}\left[|\xi_j| \mathbb{1}_{\{d^\delta C_{\xi,d}^{1/p} \tau_{n(\ell)} \le |\xi_j| \le d^\delta C_{\xi,d}^{1/p} n(\ell)^\theta)\}} \right] \right) \right| \tag{4.70}$$

tends to zero almost surely. For any $\varepsilon > 0$ let $B := \frac{1}{n(\ell)^{1/p} C_{\xi,d}^{1/p} d^\delta}$ then we have that

$$\begin{aligned}
&\sum_{\ell=1}^{\infty} \mathbb{P}\left(\left| B \sum_{j=1}^{n(\ell)} \left(|\xi_j| \mathbb{1}_{\{d^\delta C_{\xi,d}^{1/p} \tau_{n(\ell)} \le |\xi_j| \le d^\delta C_{\xi,d}^{1/p} n(\ell)^\theta)\}} - \mathbb{E}\left[|\xi_j| \mathbb{1}_{\{d^\delta C_{\xi,d}^{1/p} \tau_{n(\ell)} \le |\xi_j| \le d^\delta C_{\xi,d}^{1/p} n(\ell)^\theta)\}} \right] \right) \right| > \varepsilon \right) \\
&\le \frac{1}{\varepsilon^2} \sum_{\ell=1}^{\infty} \frac{1}{n(\ell)^{2/p} C_{\xi,d}^{2/p} d^{2\delta}} \sum_{j=1}^{n(\ell)} \mathrm{Var}\left(|\xi_j| \mathbb{1}_{\{d^\delta C_{\xi,d}^{1/p} \tau_{n(\ell)} \le |\xi_j| \le d^\delta C_{\xi,d}^{1/p} n(\ell)^\theta)\}} \right) \\
&\le \frac{1}{\varepsilon^2} \sum_{\ell=1}^{\infty} \frac{n(\ell)^{1-2/p}}{d^{2\delta} C_{\xi,d}^{2/p}} \mathbb{E}\left[|\xi_1|^2 \mathbb{1}_{\{d^\delta C_{\xi,d}^{1/p} \tau_{n(\ell)} \le |\xi_1| \le d^\delta C_{\xi,d}^{1/p} n(\ell)^\theta)\}} \right].
\end{aligned}$$

By Tonelli's theorem, we have that

$$\begin{aligned}
&\sum_{\ell=1}^{\infty} \frac{n(\ell)^{1-2/p}}{d^{2\delta} C_{\xi,d}^{2/p}} \mathbb{E}\left[|\xi_1|^2 \mathbb{1}_{\{d^\delta C_{\xi,d}^{1/p} \tau_{n(\ell)} \le |\xi_1| \le d^\delta C_{\xi,d}^{1/p} n(\ell)^\theta)\}} \right] \\
&= \mathbb{E}\left[|\xi_1|^2 \sum_{\ell=1}^{\infty} \frac{n(\ell)^{1-2/p}}{d^{2\delta} C_{\xi,d}^{2/p}} \mathbb{1}_{\{d^\delta C_{\xi,d}^{1/p} \tau_{n(\ell)} \le |\xi_1| \le d^\delta C_{\xi,d}^{1/p} n(\ell)^\theta)\}} \right]
\end{aligned}$$

Since we have that

$$\begin{aligned}
&\sum_{k=1}^{\infty} n(\ell)^{1-2/p} \mathbb{1}_{\{d^\delta C_{\xi,d}^{1/p} n(\ell)^{1/p} \le |\xi_1| \le d^\delta C_{\xi,d}^{1/p} n(\ell)^\theta)\}} \\
&= \sum_{k=1}^{\infty} n(\ell)^{1-2/p} \mathbb{1}_{\{d^{-\delta/\theta} C_{\xi,d}^{-1/p\theta} |\xi_1|^{1/\theta} \le n(\ell) \le\ d^{-\delta p} C_{\xi,d}^{-1} |\xi_1|^p\}} \\
&\le \sum_{k=1}^{\infty} \lambda^{(1-2/p)\ell} \mathbb{1}_{\{d^{-\delta/\theta} C_{\xi,d}^{-1/p\theta} |\xi_1|^{\frac{1}{\theta}} \le \lambda^\ell \le\ d^{-\delta p} C_{\xi,d}^{-1} |\xi_1|^p + 1\}}.
\end{aligned}$$

This gives

$$
\begin{aligned}
&\sum_{\ell=1}^{\infty} \lambda^{(1-2/p)\ell} \mathbb{1}_{\{d^{-\delta/\theta} C_{\xi,d}^{-1/p\theta} |\xi_1|^{\frac{1}{\theta}} \le \lambda^{\ell} \le\ d^{-\delta p} C_{\xi,d}^{-1} |\xi_1|^p + 1\}} \\
&\le \sum_{\ell=1}^{\infty} \lambda^{(1-2/p)\ell} \mathbb{1}_{\{\lfloor \ln\left(d^{-\frac{\delta}{\theta}} C_{\xi,d}^{-1/p\theta} |\xi_1|^{\frac{1}{\theta}}\right)/\ln(\lambda) \rfloor \le \ell \le\ \lceil \ln\left(d^{-\frac{p}{2}} C_{\xi,d}^{-1} |\xi_1|^p + 1\right)/\ln(\lambda) \rceil\}} \\
&= \sum_{\ell=\lfloor \ln\left(d^{-\delta/\theta} C_{\xi,d}^{-1/p\theta} |\xi_1|^{1/\theta}\right)/\ln(\lambda) \rfloor}^{\lceil \ln\left(d^{-\delta p} C_{\xi,d}^{-1} |\xi_1|^p + 1\right)/\ln(\lambda) \rceil} e^{(1-2/p)\ell \ln(\lambda)} \\
&\le \int_{\ln\left(d^{-\delta/\theta} C_{\xi,d}^{-1/p\theta} |\xi_1|^{1/\theta}\right)/\ln(\lambda)}^{\ln\left(d^{-\delta p} C_{\xi,d}^{-1} |\xi_1|^p + 1\right)/\ln(\lambda) + 2} e^{(1-2/p)x \ln(\lambda)} dx \\
&= \frac{1}{\ln(\lambda)} \int_{\ln\left(d^{-\delta/\theta} C_{\xi,d}^{-1/p\theta} |\xi_1|^{1/\theta}\right)}^{\ln\left(d^{-\delta p} C_{\xi,d}^{-1} |\xi_1|^p + 1\right) + 2\ln(\lambda)} e^{(1-2/p)x} dx \\
&\le C_1 (d^{-\delta p} C_{\xi,d}^{-1} |\xi_1|^p + 1)^{1-2/p} - C_2 (d^{-\delta/\theta} C_{\xi,d}^{-1/p\theta} |\xi_1|^{1/\theta})^{(1-2/p)} \\
&\le 2 C_1 d^{\delta(2-p)} C_{\xi,d}^{2/p-1} |\xi_1|^{p-2} - C_2 d^{-(1-2/p)\delta/\theta} C_{\xi,d}^{-(p-2)/p^2\theta} |\xi_1|^{(1-2/p)/\theta},
\end{aligned}
$$

with $C_1 = \frac{p\lambda^{2(1-2/p)}}{(p-2)\ln(\lambda)}$ and $C_2 = (\ln \lambda)^{-1} p/(p-2)$. Here the last inequality follows since on the event $\{|\xi_1| \ge d^{\delta} C_{\xi,d}^{1/p} \tau_{n(\ell)}\}$, we have that $|\xi_1|^p C_{\xi,d}^{-1} d^{-\delta p} \ge 1$ and hence $(d^{-\delta p} C_{\xi,d}^{-1} |\xi_1|^p + 1)^{\alpha} \le 2 d^{-\alpha\delta p} C_{\xi,d}^{-\alpha} |\xi_1|^{\alpha p}$ for any $\alpha > 0$. From (4.61) it follows that for $p > 2$ and θ satisfying (4.64) we have that

$$
\begin{aligned}
\sum_{\ell=1}^{\infty} \frac{n(\ell)^{1-2/p}}{C_{\xi,d}^{2/p} d^{2\delta}} \mathbb{E}\left[|\xi_1|^2 \mathbb{1}_{\{d^{\delta} C_{\xi,d}^{1/p} \tau_{n(\ell)} \le |\xi_1| \le d^{\delta} C_{\xi,d}^{1/p} n(\ell)^{\theta})\}} \right] &\le 2 C_1 d^{-\delta p} C_{\xi,d}^{-1} \mathbb{E}|\xi_1|^p \\
&\quad + C_2 d^{-(1-2/p)\delta/\theta - 2\delta} C_{\xi,d}^{-(p-2)/p^2\theta - 2/p} \mathbb{E}|\xi_1|^{2+(1-2/p)/\theta} < \bar{C},
\end{aligned}
$$

where $\bar{C}$ is some universal constant, since by (4.60) the first term can be bounded by a universal constant and since for θ satisfying (4.64) and $p > 2$ we have that $1 < 2 + (1 - 2/p)/\theta < p$, and thus by Jensen's inequality we can also bound the second term with a universal constant. Consequently, the claim (4.70) follows from the Borel–Cantelli lemma. Moreover, from (4.67)

we also have that

$$\begin{aligned}
&\left|\sum_{j=1}^{n(\ell)}\left(|\xi_j|\mathbb{1}_{\{d^\delta C_{\xi,d}^{1/p}\tau_{n(\ell)}\le|\xi_j|\le d^\delta C_{\xi,d}^{1/p}n(\ell)^\theta)\}}\right)\right| \\
&= o_{a.s.}(d^\delta C_{\xi,d}^{1/p}n(\ell)^{1/p}) + \sum_{j=1}^{n(\ell)}\mathbb{E}\left[|\xi_j|\mathbb{1}_{\{d^\delta C_{\xi,d}^{1/p}\tau_{n(\ell)}\le|\xi_j|\le d^\delta C_{\xi,d}^{1/p}n(\ell)^\theta)\}}\right] \\
&= o_{a.s.}(d^\delta C_{\xi,d}^{1/p}n(\ell)^{1/p}) + \mathcal{O}_{a.s.}(d^\delta C_{\xi,d}^{1/p}n(\ell)\tau_{n(\ell)}^{1-p}) \\
&= \mathcal{O}_{a.s.}(d^\delta C_{\xi,d}^{1/p}n(\ell)^{1/p}) = \mathcal{O}_{a.s.}(d^\delta C_{\xi,d}^{1/p}\lambda^{\ell/p}).
\end{aligned}$$

Since for any n there exists an ℓ such that $\lambda^\ell < n \le \lambda^{\ell+1}$, we have that

$$n^{-1/p}\sum_{j=1}^{n}|\xi_j|\mathbb{1}_{\{d^\delta C_{\xi,d}^{1/p}n^{1/p}\le|\xi_j|\le d^\delta C_{\xi,d}^{1/p}n^\theta)\}} \le \lambda^{-\ell/p}\sum_{j=1}^{\lambda^{\ell+1}}|\xi_j|\mathbb{1}_{\{d^\delta C_{\xi,d}^{1/p}\lambda^{(\ell+1)/p}\le|\xi_j|\le d^\delta C_{\xi,d}^{1/p}\lambda^{(\ell+1)\theta}\}}$$

$$\begin{aligned}
\limsup_{n\to\infty} n^{-1/p}\sum_{j=1}^{n}|\xi_j|\mathbb{1}_{\{d^\delta C_{\xi,d}^{1/p}n^{1/p}\le|\xi_j|\le d^\delta C_{\xi,d}^{1/p}n^\theta)\}} &\le \frac{\lambda^{1/p}}{\lambda^{(\ell+1)/p}}\sum_{j=1}^{\lambda^{\ell+1}}|\xi_j|\mathbb{1}_{\{d^\delta C_{\xi,d}^{1/p}\lambda^{(\ell+1)/p}\le|\xi_j|\le d^\delta C_{\xi,d}^{1/p}\lambda^{\ell\theta}\}} \\
&\lesssim_{a.s.} d^\delta C_{\xi,d}^{1/p}.
\end{aligned}$$

Thus we have proven the claim (4.68) and consequently also (4.62) immediately follows. Putting all the obtained results together, we have that

$$\begin{aligned}
\left|\sum_{j=1}^{n}\xi_j - \sum_{j=1}^{n}\tilde{\xi}_j\right| &= \mathcal{O}_{a.s.}(d^\delta C_{\xi,d}^{1/p}n\tau_n^{1-p}) \\
&= \mathcal{O}_{a.s.}(d^{1/2}C_{\xi,d}^{1/p}n^{1/p}).
\end{aligned} \tag{4.71}$$

Next, since $\mathbb{E}[\tilde{\xi}_j] = 0$ and $\left|\tilde{\xi}_j\right| \le 2d^{1/2}C_{\xi,d}^{1/p}\tau_n$, by Lemma 4.5.15 there exists a sequence of Gaussian random variables $\tilde{y}_1,\dots,\tilde{y}_n$ with $\tilde{y}_j \sim \mathcal{N}_d(0,\mathrm{Cov}(\tilde{\xi}_i))$ such that for all $\theta_0 > 0$ we have that

$$\begin{aligned}
\left|\sum_{j=1}^{n}\tilde{\xi}_j - \sum_{j=1}^{n}\tilde{y}_j\right| &= \mathcal{O}_{a.s.}((d^{23/4+\delta+\theta_0}\log^* d)\sqrt{\sigma_d}C_\xi^{1/p}\tau_n\log n)) \\
&= \mathcal{O}_{a.s.}((d^{25/4+\theta_0}\log^* d)\sqrt{\sigma_d}C_\xi^{1/p}n^{1/p}\log n)),
\end{aligned} \tag{4.72}$$

for any $\theta_0 > 0$. Hence by (4.71) and (4.72) it follows that

$$\begin{aligned}\left|\sum_{k=1}^{n}\xi_k - \tilde{Y}_n\right| &\leq \left|\sum_{k=1}^{n}\xi_k - \sum_{j=1}^{n}\tilde{\xi}_j\right| + \left|\sum_{j=1}^{n}\tilde{\xi}_j - \tilde{Y}_n\right| \\ &= \mathcal{O}_{a.s.}(d^{1/2}C_{\xi,d}^{1/p}n^{1/p}) + \mathcal{O}_{a.s.}((d^{25/4+\theta_0}\log^* d)\sqrt{\sigma_d}C_{\xi}^{1/p}n^{1/p}\log n)),\end{aligned} \tag{4.73}$$

where $\tilde{Y}_n := \sum_{j=1}^{n}\tilde{y}_j$ has the same covariance matrix as $\sum_{j=1}^{n}\tilde{\xi}_j$. Note that we can write $\tilde{y}_j$ in terms of a standard multidimensional Gaussian as

$$\tilde{y}_j = \left(\mathrm{Cov}(\tilde{\xi}_1)\right)^{1/2} z_j \quad \text{for } j = 1, \ldots, n,$$

where $z_1, \ldots, z_n \sim \mathcal{N}_d(0, I_d)$ are all independent and identically distributed. Similarly, using the same Gaussian sequence $z_1, \ldots, z_n$, we can define for $j = 1, \ldots, n$ the following

$$y_j := (\mathrm{Cov}(\xi_1))^{1/2} z_j.$$

Now it follows that

$$\begin{aligned}\left|\sum_{j=1}^{n}\tilde{y}_j - \sum_{j=1}^{n}y_j\right| &= \left|\mathrm{Cov}\,(\xi_1)^{1/2} - \mathrm{Cov}(\tilde{\xi}_1)^{1/2}\right|\left|\sum_{j=1}^{n}z_j\right| \\ &\leq \frac{1}{\sqrt{\sigma_0}}\left|\mathrm{Cov}\,(\xi_1) - \mathrm{Cov}(\tilde{\xi}_1)\right|\left|\sum_{j=1}^{n}z_j\right| \\ &\leq \frac{\left|\sum_{j=1}^{n}z_j\right|}{\sqrt{\sigma_0}}\left[2\sqrt{\mathbb{E}\left[|\xi_1|^2\right]}\sqrt{\mathbb{E}\left[\left|\xi_1 - \tilde{\xi}_1\right|^2\right]} + \mathbb{E}\left[\left|\xi_1 - \tilde{\xi}_1\right|^2\right]\right],\end{aligned}$$

where the first inequality follows by Lemma 4.5.17; Ando–van Hemmen's inequality.
By a component-wise application of the law of iterated logarithm we have that

$$\left|\sum_{j=1}^{n}z_j\right| = \mathcal{O}_{a.s.}\left(\sqrt{dn\log\log n}\right).$$

Note that $\mathbb{E}[|\xi_1|^2] \leq \mathrm{tr}(\Sigma) \leq d\sigma_d$. Furthermore, from (4.61) we see that

$$\mathbb{E}\left[\left|\xi_1 - \tilde{\xi}_1\right|^2\right] \leq \mathbb{E}\left[|\xi_1|^2 \mathbb{1}_{\{|\xi_1| > d^\delta C_{\xi,d}^{1/p} \tau_n\}}\right] \lesssim d^{2\delta} C_{\xi,d}^{2/p} \tau_n^{2-p}.$$

We see that

$$\begin{aligned}\left|\sum_{j=1}^n \tilde{y}_j - \sum_{j=1}^n y_j\right| &= \mathcal{O}_{a.s.}\left(d^{(1+\delta)\wedge(2\delta+1/2)} \left(\frac{\sigma_d}{\sigma_0}\right)^{1/2} C_{\xi,d}^{2/p} \tau_n^{2-p} \sqrt{n \log\log n}\right) \\ &= \mathcal{O}_{a.s.}\left(d^{3/2} \left(\frac{\sigma_d}{\sigma_0}\right)^{1/2} C_{\xi,d}^{2/p} n^{2/p-1} \sqrt{\log\log n}\right), \qquad (4.74)\end{aligned}$$

where the last equality follows from the choice $\tau_n = n^{1/p}$ and $\delta = 1/2$. From (4.74) and (4.76) it follows that

$$\begin{aligned}\left|\sum_{j=1}^n \xi_j - \sum_{j=1}^n y_j\right| &\leq \left|\sum_{j=1}^n \xi_j - \sum_{j=1}^n \tilde{y}_j\right| + \left|\sum_{j=1}^n y_j - Y_n\right| \\ &= \mathcal{O}_{a.s.}\left(\left(d^{25/4+\theta_0} \log^*(d) C_\xi^{1/p} \vee d^{3/2} C_\xi^{2/p}\right) \left(\frac{\sigma_d}{\sigma_0}\right)^{1/2} n^{1/p} \log(n)\right).\end{aligned}$$

This finishes the proof for the strong Gaussian approximation case. For the weak Gaussian approximation case the proof follows completely analogously up to some minor differences. Firstly, we consider an application of Lemma 4.5.13 instead of Lemma 4.5.15 for approximating the truncated sequence $(\tilde{\xi}_k)_{k\in\mathbb{N}}$. This gives us

$$\begin{aligned}\left|\sum_{j=1}^n \tilde{\xi}_j - \tilde{Y}_n\right| &= \mathcal{O}_P(d^{\delta+1/2} C_\xi^{1/p} \tau_n \log n) \\ &= \mathcal{O}_P(d C_\xi^{1/p} n^{1/p} \log n), \qquad (4.75)\end{aligned}$$

where $\tilde{Y}_n$ has a mean-zero Gaussian distribution with covariance matrix $\mathrm{Cov}(\sum_{j=1}^n \tilde{\xi}_j)$. Furthermore, the difference between the Gaussian approximation for the truncated sums and the Gaussian approximation for the true sums can now be obtained with the following argument. $Z \sim \mathcal{N}(0, I_d)$ as follows:

$$\tilde{Y}_n := \mathrm{Cov}\left(\sum_{j=1}^n \tilde{\xi}_j\right)^{1/2} Z.$$

Similarly, using the same Z, we define

$$Y_n := \operatorname{Cov}\left(\sum_{k=1}^{n} \xi_k\right)^{1/2} Z,$$

We see that

$$\begin{aligned}
|Y - \tilde{Y}_n| &= \left| \left(\operatorname{Cov}\left(\sum_{k=1}^{n} \xi_k\right)^{1/2} - \operatorname{Cov}\left(\sum_{k=1}^{n} \tilde{\xi}_k\right)^{1/2} \right) Z \right| \\
&\leq \left| \operatorname{Cov}\left(\sum_{k=1}^{n} \xi_k\right)^{1/2} - \operatorname{Cov}\left(\sum_{k=1}^{n} \tilde{\xi}_j\right)^{1/2} \right| |Z| \\
&\leq \frac{|Z|}{(\sigma_0 n)^{1/2}} \left| \operatorname{Cov}\left(\sum_{k=1}^{n} \xi_k\right) - \operatorname{Cov}\left(\sum_{k=1}^{n} \tilde{\xi}_k\right) \right| \\
&\leq \frac{|Z|}{(\sigma_0 n)^{1/2}} \left[2 \sqrt{\mathbb{E}\left[\left| \sum_{k=1}^{n} \xi_k \right|^2 \right]} \sqrt{\mathbb{E}\left[\left| \sum_{k=1}^{n} \xi_k - \sum_{k=1}^{n} \tilde{\xi}_k \right|^2 \right]} + \mathbb{E}\left[\left| \sum_{k=1}^{n} \xi_k - \sum_{k=1}^{n} \tilde{\xi}_k \right|^2 \right] \right]
\end{aligned}$$

Here the first inequality follows since $|Ax| \leq |A||x|$, the second by Ando–van Hemmen's inequality using the assumption that the smallest eigenvalue of $\frac{1}{\sqrt{n}} \sum_{k=1}^{n} \xi_k$ is bounded below by σ_0.

Furthermore, since $Z \sim \mathcal{N}(0, I_d)$ we have by the Komlós–Major–Tusnády approximation [90; Theorem 1] that there exists a Brownian motion B_1 such that $\left| \sum_{k=1}^{d} Z_i^2 - d - \sqrt{2} B_1(d) \right| = \mathcal{O}_{a.s.}(\log d)$. By the law of iterated logarithm, we see that $|Z|^2 = d + \mathcal{O}_{a.s.}(\sqrt{d \log\log d})$ and consequently $|Z| = \mathcal{O}_{a.s.}(\sqrt{d})$. By following the argument of Lemma 4.5.11 and an application of Jensen's inequality it follows that

$$\sqrt{\mathbb{E}\left[\left| \sum_{k=1}^{n} \xi_k \right|^2 \right]} \lesssim d^{1/2} C_{\xi,d}^{1/p} n^{1/2}.$$

Furthermore, by (4.61) we also have that

$$\mathbb{E}\left[\left|\sum_{k=1}^{n}\xi_k-\sum_{k=1}^{n}\tilde{\xi}_k\right|^2\right]\leq\sum_{k=1}^{n}\mathbb{E}[|\xi_k|^2\mathbb{1}_{\{|\xi_k|>d^{\delta}C_{\xi,d}^{1/p}\tau_n\}}]$$
$$\lesssim d^{2\delta}C_{\xi,d}^{2/p}n\tau_n^{2-p}$$

This gives

$$|Y_n-\tilde{Y}_n|\leq\frac{|Z|}{(\sigma_0 n)^{1/2}}\left[2\sqrt{\mathbb{E}\left[\left|\sum_{k=1}^{n}\xi_k\right|^2\right]}\sqrt{\mathbb{E}\left[\left|\sum_{k=1}^{n}(\xi_k-\tilde{\xi}_k)\right|^2\right]}+\mathbb{E}\left[\left|\sum_{k=1}^{n}(\xi_k-\tilde{\xi}_k)\right|^2\right]\right]$$
$$=\frac{O_{a.s.}(\sqrt{d})}{(\sigma_0 n)^{1/2}}\left[\mathcal{O}(d^{1/2}C_{\xi,d}^{1/p}n^{1/2})\sqrt{\mathcal{O}(d^{2\delta}C_{\xi,d}^{2/p}n\tau_n^{2-p})}+\mathcal{O}(d^{2\delta}C_{\xi,d}^{2/p}n\tau_n^{2-p})\right]$$

Now since $\tau_n=n^{1/p}$ and $\delta=1/2$ we see that

$$\mathbb{E}\left[\left|\sum_{k=1}^{n}(\xi_k-\tilde{\xi}_k)\right|^2\right]=\mathcal{O}(dC_{\xi,d}^{2/p}n^{2/p}),$$

and consequently

$$|Y_n-\tilde{Y}_n|=\mathcal{O}_P(d^{3/2}C_{\xi,d}^{2/p}n^{1/p})+\mathcal{O}_P(d^{3/2}C_{\xi,d}^{2/p}n^{2/p-1/2})=\mathcal{O}_P(d^{3/2}C_{\xi,d}^{2/p}n^{1/p}),\tag{4.76}$$

where the last equality holds since $p>2$. From (4.73) and (4.76) it follows that

$$\left|\sum_{k=1}^{n}\xi_k-Y_n\right|\leq\left|\sum_{k=1}^{n}\xi_k-\tilde{Y}_n\right|+|Y-\tilde{Y}_n|=\mathcal{O}_P\left(d^{3/2}C_{\xi}^{2/p}n^{1/p}\log(n)\right).$$

Finally, writing Y_n as $\sum_{k=1}^{n}y_k$ completes the proof. Note that we can always, on a possibly extended probability space, do this since all conditional distributions of $(y_1,\ldots,y_n,Y_n)$ are known.

□

Theorem 4.5.19. *Let $\{\xi_k\}$ be a one-dependent identically distributed zero-mean sequence such that finite p-th moment for each element* $\sup_k\sup_i\mathbb{E}[|\xi_{k,i}|^p]\leq$

C_ξ for some $p > 2$ and that the smallest eigenvalue of $\Sigma_\xi := \mathrm{Var}(\xi) + \mathrm{Cov}(\xi_1, \xi_2) + \mathrm{Cov}(\xi_1, \xi_2)^T$ is larger than some constant $\sigma_0 > 0$ and $\sup_{i,j} \left|\Sigma_{\xi_{ij}}\right| < \infty$ then there exists a Brownian motion W such that

$$\left|\sum_{k=1}^{n} \xi_k - \Sigma_\xi^{1/2} W(n)\right| = \begin{cases} \mathcal{O}_P\left(d^{3/2} C_\xi^{2/p} n^{\frac{1}{4}+\frac{1}{4(p-1)}} \log(n)\right) \\ \mathcal{O}_{a.s.}\left(\left(d^{25/4+\theta_0} C_\xi^{1/p} \vee d^{3/2} C_\xi^{2/p}\right)\left(\frac{\sigma_d}{\sigma_0}\right)^{1/2} n^{\frac{1}{4}+\frac{1}{4(p-1)}} \log(n)\right) \end{cases}.$$

Proof. Let m be an integer such that $m > 1$ and $m < n$. We define

$$x_j := \sum_{i=1+(j-1)m}^{jm-1} \xi_i, \quad j = 1, \ldots, \lfloor n/m \rfloor.$$

We further define

$$x_{\lfloor n/m \rfloor + 1} := \mathbb{1}_{\{m\lfloor n/m \rfloor < n\}} \left(\sum_{i=1+m\lfloor n/m \rfloor}^{n} \xi_i\right).$$

Clearly, if $m\lfloor n/m \rfloor = n$ we have $x_{\lfloor n/m \rfloor + 1} = 0$. Note that $\{x_j\}$ is an independent sequence.

Then it is easy to verify that

$$\sum_{k=1}^{n} \xi_k = \sum_{j=1}^{\lfloor n/m \rfloor} x_j + x_{\lfloor n/m \rfloor + 1} + \sum_{j=1}^{\lfloor n/m \rfloor} \xi_{jm}.$$

By triangle inequality

$$\left|\sum_{k=1}^{n} \xi_k - \sum_{j=1}^{\lfloor n/m \rfloor} x_j\right| \leq \mathbb{1}_{\{m\lfloor n/m \rfloor < n\}} \left|\sum_{i=1+m\lfloor n/m \rfloor}^{n} \xi_i\right| + \left|\sum_{j=1}^{\lfloor n/m \rfloor} \xi_{jm}\right|$$

Note that $(\xi_{jm})_{j \in \mathbb{N}}$ are a zero mean independent sequence with

$$\sup_k \sup_i \mathbb{E}[|\xi_{k,i}|^p] < C_\xi,$$

hence by Proposition 4.5.18 there exists a Brownian motion B such that

$$\left|\sum_{j=1}^{\lfloor n/m \rfloor} \xi_{jm} - \Sigma_1^{1/2} B(\lfloor n/m \rfloor)\right| = \mathcal{O}_{a.s.}\left(d^\theta \log^*(d) C_\xi^{2/p} \left(\frac{\sigma_d}{\sigma_0}\right)^{1/2} \lfloor n/m \rfloor)^{1/p} \log(\lfloor n/m \rfloor)\right),$$

where $\theta = 25/4 + \theta_0$ for some $\theta_0 > 0$ in the strong approximation case, Σ_1 is the covariance matrix of ξ_m. By applying the law of iterated logarithm to every coordinate of B, it follows that

$$\begin{aligned}\left|\sum_{j=1}^{\lfloor n/m\rfloor} \xi_{jm}\right| &\le \left|\Sigma_1^{1/2} B(\lfloor n/m\rfloor))\right| + \mathcal{O}_{a.s.}(d^\theta C_\xi^{2/p}(\log d)(\lfloor n/m\rfloor)^{1/p}(\log\lfloor n/m\rfloor)) \\ &\le \left|\Sigma_1^{1/2}\right| \left(\sum_{i=1}^{d} B_i^2(\lfloor n/m\rfloor))\right)^{1/2} + \mathcal{O}_{a.s.}(d^\theta C_\xi^{2/p}(\log d)(\lfloor n/m\rfloor)^{1/p}(\log\lfloor n/m\rfloor)) \\ &= \mathcal{O}_{a.s.}(d^{1/2}\sqrt{\mathrm{tr}(\Sigma_1)\lfloor n/m\rfloor}\log\log\lfloor n/m\rfloor) \\ &+ \mathcal{O}_{a.s.}(d^\theta C_\xi^{2/p}(\log d)(\lfloor n/m\rfloor)^{1/p}(\log\lfloor n/m\rfloor)).\end{aligned}$$

Note that since $\sqrt{\mathrm{tr}(\mathrm{Var}(\xi))} = \sqrt{\mathbb{E}|\xi|^2} \le (\mathbb{E}|\xi|^p)^{1/p} \lesssim d^{1/2} C_\xi^{1/p}$, we have that

$$\left|\sum_{j=1}^{\lfloor n/m\rfloor} \xi_{jm}\right| = \mathcal{O}_{a.s.}(d^\theta \log^* d C_\xi^{2/p}(n/m)^{1/2}\log(n/m)). \tag{4.77}$$

Similarly, we can use the same argument for the term $\mathbb{1}_{\{m\lfloor n/m\rfloor<n\}}\left|\sum_{i=1+m\lfloor n/m\rfloor}^{n}\xi_i\right|$. The only additional step is to split the sequence of ξ_i by its odd and even indices and use triangle inequality to divide the sum into two sums of independent sequences, each having $\mathcal{O}(n/m)$ terms. Then by applying the same argument, we obtain

$$\left|\mathbb{1}_{\{m\lfloor n/m\rfloor<n\}}\left(\sum_{i=1+m\lfloor n/m\rfloor}^{n}\xi_i\right)\right| = \mathcal{O}_{a.s.}(d^\theta C_\xi^{2/p}(\log d)(n/m)^{1/2}\log(n/m)).$$

Therefore, we have shown

$$\left|\sum_{k=1}^{n}\xi_k - \sum_{j=1}^{\lfloor n/m\rfloor} x_j\right| = \mathcal{O}_{a.s.}(d^\theta C_\xi^{2/p}(\log d)(n/m)^{1/2}\log(n/m)).$$

Similarly, by an application of the weak Gaussian approximation, we can show

$$\left|\sum_{k=1}^{n}\xi_k - \sum_{j=1}^{\lfloor n/m\rfloor} x_j\right| = \mathcal{O}_P(d^{1/2}C_\xi^{1/p}(n/m)^{1/2}\log(n/m)). \tag{4.78}$$

Let $\tau_n = n^\gamma \log(n)$ and $m := m_n = \lfloor n^\alpha \rfloor$, for some $\alpha, \gamma > 0$, that will be determined later in the proof. Define

$$\tilde{x}_j := x_j \mathbb{1}_{\{|x_j| \le d^{1/2}C_\xi^{1/p}\tau_n\}} - \mathbb{E}[x_j \mathbb{1}_{\{|x_j| \le d^{1/2}C_\xi^{1/p}\tau_n\}}].$$

Clearly, we have $|\tilde{x}_j| \le 2d^{1/2}C_\xi^{2/p}\tau_n$ and $\mathbb{E}[\tilde{x}_j] = 0$. Then by triangle inequality, we have

$$\left|\sum_{j=1}^{\lfloor n/m\rfloor} x_j - \sum_{j=1}^{\lfloor n/m\rfloor} \tilde{x}_j\right| \le \left|\sum_{j=1}^{\lfloor n/m\rfloor} x_j \mathbb{1}_{\{|x_j| > d^{1/2}C_\xi^{1/p}\tau_n)\}}\right| + \left|\sum_{j=1}^{\lfloor n/m\rfloor} \mathbb{E}[x_j \mathbb{1}_{\{|x_j| > d^{1/2}C_\xi^{1/p}\tau_n\}}]\right|.$$

Since we can decompose x_j into sums of sub-sequences of even and odd indices, this implies that by Lemma 4.5.11 we have for all j that $\mathbb{E}[|x_j|^p] \lesssim d^{p/2}C_\xi m^{p/2}$. Then we have

$$\begin{aligned}
\left|\sum_{j=1}^{\lfloor n/m\rfloor} \mathbb{E}[x_j \mathbb{1}_{\{|x_j| > d^{1/2}C_\xi^{1/p}\tau_n\}}]\right| &\le d^{1/2}C_\xi^{1/p}\tau_n \sum_{j=1}^{\lfloor n/m\rfloor} \mathbb{E}\left[\frac{|x_j|}{d^{1/2}C_\xi^{1/p}\tau_n}\mathbb{1}_{\{|x_j| > d^{1/2}C_\xi^{1/p}\tau_n\}}\right] \\
&\le d^{1/2}C_\xi^{1/p}\tau_n \sum_{j=1}^{\lfloor n/m\rfloor} \mathbb{E}\left[\frac{|x_j|^p}{d^{p/2}C_\xi\tau_n^p}\mathbb{1}_{\{|x_j| > d^{1/2}C_\xi^{1/p}\tau\}}\right] \\
&\lesssim d^{1/2}C_\xi^{1/p} n m^{p/2-1}\tau_n^{1-p}.
\end{aligned}$$

Note that $(|x_j|\mathbb{1}_{|x_j| > d^{1/2}C_\xi^{1/p}\tau_n})_{j\in\mathbb{N}}$ is an i.i.d sequence with mean $\mu_{n,d}$ with $0 < \mu_{n,d} \lesssim d^{1/2}C_\xi^{1/p}m^{p/2}\tau_n^{1-p}$. Let $s_n = \lfloor n/m \rfloor$, we claim that

$$\sum_{j=1}^{s_n} |x_j| \mathbb{1}_{\{|x_j| > d^{1/2}C_\xi^{1/p}\tau_n)\}} = \mathcal{O}_{a.s.}(d^{1/2}C_\xi^{1/p} n m^{p/2-1} n^{\gamma(1-p)}). \tag{4.79}$$

This is equivalent to showing that

$$\left| \frac{n^{\gamma(p-1)} m^{1-p/2}}{n d^{1/2} C_\xi^{1/p}} \sum_{j=1}^{s_n} |x_j| \mathbb{1}_{\{|x_j| \geq d^{1/2} C_\xi^{1/p} \tau_n)\}} \right| \xrightarrow{a.s.} 0 \text{ as } n \to \infty. \tag{4.80}$$

We will first show that (4.80) holds for a subsequence. Fix any $\lambda > 1$ and define $n(\ell) = \lfloor \lambda^\ell \rfloor$. Then, we claim that

$$\left| \frac{n(\ell)^{\gamma(p-1)} m_{n(\ell)}^{1-p/2}}{n(\ell) C_\xi^{1/p} d^{1/2}} \sum_{j=1}^{\lfloor s_{n(\ell)} \rfloor} |x_j| \mathbb{1}_{\{|x_j| \geq d^{1/2} C_\xi^{1/p} \tau_{n(\ell))}\}} \right| \xrightarrow{a.s.} 0 \text{ as } \ell \to \infty. \tag{4.81}$$

For every $\varepsilon > 0$ we have that

$$\begin{aligned}
&\sum_{\ell=1}^{\infty} \mathbb{P}\left(\left| \sum_{j=1}^{\lfloor s_{n(\ell)} \rfloor} |x_j| \mathbb{1}_{\{|x_j| > d^{1/2} C_\xi^{1/p} \tau_{n(\ell)}\}} \right| > \varepsilon d^{1/2} C_\xi^{1/p} n(\ell) m_{n(\ell)}^{p/2-1} n(\ell)^{\gamma(1-p)} \right) \\
&\leq \sum_{\ell=1}^{\infty} \frac{\mathbb{E}[|x_1| \mathbb{1}_{\{|x_1| > d^{1/2} C_\xi^{1/p} \tau_{n(\ell)}\}}]}{\varepsilon d^{1/2} C_\xi^{1/p} m_{n(\ell)}^{p/2} n(\ell)^{\gamma(1-p)}} \\
&\lesssim \frac{1}{\varepsilon} \sum_{\ell=1}^{\infty} \frac{\tau_{n(\ell)}^{(1-p)}}{n(\ell)^{\gamma(1-p)}} \\
&\lesssim \frac{1}{\varepsilon} \sum_{\ell=1}^{\infty} \frac{1}{\ell^{(p-1)}} < \infty,
\end{aligned}$$

since $\mathbb{E}[|x_1| \mathbb{1}_{\{|x_1| > d^{1/2} \tau_n\}}] \lesssim d^{1/2} m^{p/2} \tau_n^{(1-p)}$ and $p > 2$. Consequently, the claim (4.81) follows from the Borel–Cantelli lemma. We will choose α and γ such that

$$1 + \alpha(p/2 - 1) + \gamma(1 - p) > 0. \tag{4.82}$$

Since for any n there exists a ℓ such that $\lambda^\ell < n \leq \lambda^{\ell+1}$, we have that

$$\begin{aligned}
&n^{-1+\alpha(1-p/2)+\gamma(p-1)} \sum_{j=1}^{\lfloor n^{1-\alpha} \rfloor} |x_j| \mathbb{1}_{\{|x_j| \geq d^{1/2} C_\xi^{1/p} \tau_n)\}} \\
&\leq \lambda^{\ell(-1+\alpha(1-p/2)+\gamma(p-1)} \sum_{j=1}^{\lfloor \lambda^{(\ell+1)(1-\alpha)} \rfloor} |x_j| \mathbb{1}_{\{|x_j| \geq d^{1/2} C_\xi^{1/p} \lambda^{\gamma(\ell+1)} \ell \ln(\lambda)\}}
\end{aligned}$$

Therefore it immediately follows that

$$\limsup_{n\to\infty} n^{-1+\alpha(1-p/2)+\gamma(p-1)} \sum_{j=1}^{n} |x_j| \mathbb{1}_{\{|x_j| \geq d^{1/2} C_\xi^{1/p} \tau_n)\}}$$
$$\leq \frac{\lambda^{-1+\alpha(1-p/2)+\gamma(p-1)}}{\lambda^{(\ell+1)(-1+\alpha(1-p/2)+\gamma(p-1))}} \sum_{j=1}^{\lambda^{\ell+1}} |x_j| \mathbb{1}_{\{|x_j| \geq d^{1/2} C_\xi^{1/p} \lambda^{\ell\gamma}\}} \lesssim_{a.s.} d^{1/2} C_\xi^{1/p}.$$

Thus we have proven the claim (4.79). Consequently, we have

$$\left| \sum_{j=1}^{s_n} x_j - \sum_{j=1}^{s_n} \tilde{x}_j \right| = \mathcal{O}_{a.s.}(d^{1/2} C_\xi^{1/p} \lambda^{\ell\gamma} n m^{p/2-1} \tau_n^{1-p})$$

Overall, we have shown that if τ_n and m are chosen such that $m^{p/2}\tau_n^{1-p} \to 0$ then

$$\left| \sum_{j=1}^{n} \xi_j - \sum_{j=1}^{s_n} \tilde{x}_j \right| = \mathcal{O}_{a.s.}(d^\theta C_\xi^{2/p} (\log d)(n/m)^{1/2} \log(n/m)) + \mathcal{O}_{a.s.}(d^{1/2} C_\xi^{1/p} n m^{p/2-1} \tau_n^{1-p}).$$

Similarly, for the weak approximation we have that

$$\left| \sum_{j=1}^{n} \xi_j - \sum_{j=1}^{s_n} \tilde{x}_j \right| = \mathcal{O}_P(d^{1/2} C_\xi^{1/p} (n/m)^{1/2} \log(n/m)) + \mathcal{O}_P(d^{1/2} C_\xi^{1/p} n m^{p/2-1} \tau_n^{1-p}). \tag{4.83}$$

Next, since $\mathbb{E}[\tilde{x}_j] = 0$ and $|\tilde{x}_j| \leq 2d^{1/2} C_\xi^{1/p} \tau_n$, by Lemma 4.5.14 there exists a sequence of independent Gaussians $(\tilde{y}_j)_j$ such that $\tilde{y}_j \sim \mathcal{N}_d(0, \mathrm{Cov}(\tilde{x}_j))$ for $j = 1, \ldots, s_n$ and

$$\left| \sum_{j=1}^{s_n} \tilde{x}_j - \sum_{j=1}^{s_n} \tilde{y}_j \right| = \mathcal{O}_{a.s.}(d^\theta \log^*(d) C_\xi^{1/p} \tau_n \log n).$$

Furthermore, since a Brownian motion at integer times coincides with a sum of i.i.d. Gaussian random variables, there exists a standard Brownian motion B such that

$$\left| \sum_{j=1}^{s_n} \tilde{x}_j - s_n^{-1/2} \mathrm{Cov}\left(\sum_{j=1}^{s_n} \tilde{x}_j \right)^{1/2} B(s_n) \right| = \mathcal{O}_{a.s.}(d^\theta \log^*(d) C_\xi^{1/p} \tau_n \log n),$$

where we also used the fact that $(\tilde{x}_j)$ is an i.i.d. sequence. Note that by scale invariance of Brownian motion, $W := (\sqrt{m}\ B(n/m))$ is also a Brownian motion. Now let $\tilde{Y}_n$ and Y_n be defined as follows:

$$\tilde{Y}_n := \frac{1}{\sqrt{n}} \operatorname{Cov}\left(\sum_{j=1}^{s_n} \tilde{x}_j\right)^{1/2} W(n).$$

Using the same Brownian motion we also define

$$Y_n := \frac{1}{\sqrt{n}} \operatorname{Cov}\left(\sum_{k=1}^{n} \xi_k\right)^{1/2} W(n),$$

Thus we have that

$$\begin{aligned}
\left|\sum_{j=1}^{n} \xi_j - \tilde{Y}_n\right| &\le \left|\sum_{j=1}^{n} \xi_j - \sum_{j=1}^{s_n} \tilde{x}_j\right| + \left|\sum_{j=1}^{s_n} \tilde{x}_j - \tilde{Y}_n\right| && (4.84)\\
&= \mathcal{O}_{a.s.}(d^{\theta} C_{\xi}^{2/p} (\log d) \left(\frac{n}{m}\right)^{1/2} \log(n/m)) && (4.85)\\
&+ \mathcal{O}_{a.s.}(d^{1/2} C_{\xi}^{1/p} n m^{p/2-1} \tau_n^{1-p}) && (4.86)\\
&+ \mathcal{O}_{a.s.}(d^{\theta} \log d C_{\xi}^{1/p} \tau_n \log n). && (4.87)
\end{aligned}$$

Therefore we have that

$$\begin{aligned}
|Y_n - \tilde{Y}_n| &= \left|\left(\operatorname{Cov}\left(\sum_{k=1}^{n} \xi_k\right)^{1/2} - \operatorname{Cov}\left(\sum_{j=1}^{s_n} \tilde{x}_j\right)^{1/2}\right) \frac{1}{\sqrt{n}} W(n)\right| \\
&\le \left|\operatorname{Cov}\left(\sum_{k=1}^{n} \xi_k\right)^{1/2} - \operatorname{Cov}\left(\sum_{j=1}^{s_n} \tilde{x}_j\right)^{1/2}\right| \frac{1}{\sqrt{n}} |W(n)| \\
&\le \frac{1}{(\sigma_0 n)^{1/2}} \left|\operatorname{Cov}\left(\sum_{k=1}^{n} \xi_k\right) - \operatorname{Cov}\left(\sum_{j=1}^{s_n} \tilde{x}_j\right)\right| \frac{1}{\sqrt{n}} |W(n)|,
\end{aligned}$$

where the second inequality follows by Ando–van Hemmen's inequality, Lemma 4.5.17, using the assumption that the smallest eigenvalue of $\frac{1}{\sqrt{n}} \sum_{k=1}^{n} \xi_k$ is

bounded below by σ_0. Following the proof of Proposition 4.5.18, we see that

$$\begin{aligned}
&\left|\mathrm{Cov}\left(\sum_{k=1}^{n}\xi_k\right)-\mathrm{Cov}\left(\sum_{j=1}^{s_n}\tilde{x}_j\right)\right| \\
&\leq \frac{1}{(\sigma_0 n)^{1/2}}2\sqrt{\mathbb{E}\left[\left|\sum_{k=1}^{n}\xi_k\right|^2\right]}\sqrt{\mathbb{E}\left[\left|\sum_{k=1}^{n}\xi_k-\sum_{j=1}^{\lfloor n/m\rfloor}\tilde{x}_j\right|^2\right]} \\
&+\frac{1}{(\sigma_0 n)^{1/2}}\mathbb{E}\left[\left|\sum_{k=1}^{n}\xi_k-\sum_{j=1}^{\lfloor n/m\rfloor}\tilde{x}_j\right|^2\right].
\end{aligned}$$

Note that by a coordinate-wise application of the law of iterated logarithm we obtain

$$|W(n)|=\left(\sum_{i=1}^{d}W_i^2(n)\right)^{1/2}=\mathcal{O}_{a.s.}(\sqrt{dn}\log\log(n)). \tag{4.88}$$

By Lemma 4.5.11 and Jensen's inequality we obtain

$$\sqrt{\mathbb{E}\left[\left|\sum_{k=1}^{n}\xi_k\right|^2\right]}\lesssim d^{1/2}C_\xi^{1/p}n^{1/2}.$$

Furthermore, we have

$$\begin{aligned}
&\mathbb{E}\left[\left|\sum_{j=1}^{\lfloor n/m\rfloor}x_j-\sum_{j=1}^{\lfloor n/m\rfloor}\tilde{x}_j\right|^2\right] \\
&=\mathbb{E}\left[\left|\sum_{j=1}^{\lfloor n/m\rfloor}\left(x_j\mathbb{1}_{\{|x_j|>d^{1/2}C_\xi^{1/p}\tau_n)\}}-\mathbb{E}[x_j\mathbb{1}_{\{|x_j|>d^{1/2}C_\xi^{1/p}\tau_n\}}]\right)\right|^2\right] \\
&=\sum_{j=1}^{\lfloor n/m\rfloor}\mathbb{E}\left[\left|x_j\mathbb{1}_{\{|x_j|>d^{1/2}C_\xi^{1/p}\tau_n)\}}-\mathbb{E}[x_j\mathbb{1}_{\{|x_j|>d^{1/2}C_\xi^{1/p}\tau_n\}}]\right|^2\right] \\
&\leq\sum_{j=1}^{\lfloor n/m\rfloor}\mathbb{E}\left[|x_j|^2\,\mathbb{1}_{\{|x_j|>d^{1/2}C_\xi^{1/p}\tau_n)\}}\right].
\end{aligned}$$

Hence it follows that

$$\begin{aligned}\mathbb{E}\left[\left|\sum_{k=1}^{n}\xi_k-\sum_{j=1}^{\lfloor n/m\rfloor}\tilde{x}_j\right|^2\right] &\le 2\sum_{j=1}^{\lfloor n/m\rfloor}\mathbb{E}[|\xi_{jm}|^2]+2\sum_{j=1}^{\lfloor n/m\rfloor}\mathbb{E}[|x_j|^2\mathbb{1}_{\{|x_j|>d^{1/2}C_\xi^{1/p}\tau_n\}}]\\ &\le 2\sum_{j=1}^{\lfloor n/m\rfloor}\mathbb{E}[|\xi_{jm}|^2]+2C_\xi^{2/p}d\tau_n^2\sum_{j=1}^{\lfloor n/m\rfloor}\mathbb{E}\left[\frac{|x_j|^p}{d^{p/2}C_\xi\tau_n^p}\mathbb{1}_{\{|x_j|>d^{1/2}C_\xi^{1/p}\tau_n\}}\right]\\ &= \mathcal{O}(dC_\xi^{2/p}nm^{-1})+\mathcal{O}(dC_\xi^{2/p}nm^{p/2-1}\tau_n^{2-p}).\end{aligned}$$

This gives

$$|Y_n-\tilde{Y}_n|\le \tag{4.89}$$

$$\begin{aligned}&\frac{|W(n)|}{n(\sigma_0)^{1/2}}\left[2\sqrt{\mathbb{E}\left[\left|\sum_{k=1}^{n}\xi_k\right|^2\right]}\sqrt{\mathbb{E}\left[\left|\sum_{k=1}^{n}\xi_k-\sum_{j=1}^{\lfloor n/m\rfloor}\tilde{x}_j\right|^2\right]}+\mathbb{E}\left[\left|\sum_{k=1}^{n}\xi_k-\sum_{j=1}^{\lfloor n/m\rfloor}\tilde{x}_j\right|^2\right]\right]\\ &=\frac{\mathcal{O}_{a.s.}(\sqrt{d}\log(n))}{(\sigma_0 n)^{1/2}}\left[\mathcal{O}(d^{1/2}C_\xi^{1/p}n^{1/2})\sqrt{\mathcal{O}(dC_\xi^{2/p}nm^{-1})+\mathcal{O}(dC_\xi^{2/p}nm^{p/2-1}\tau_n^{2-p})}\right]\\ &+\frac{\mathcal{O}_{a.s.}(\sqrt{d}\log(n))}{(\sigma_0 n)^{1/2}}\left[\mathcal{O}(dC_\xi^{2/p}nm^{p/2-1}\tau_n^{2-p})+\mathcal{O}(dC_\xi^{2/p}nm^{-1})\right]\\ &=\mathcal{O}_{a.s.}(d^{3/2}C_\xi^{2/p}n^{1/2}m^{-1/2}\log(n))+\mathcal{O}_{a.s.}(d^{3/2}C_\xi^{2/p}n^{1/2}m^{p/4-1/2}\tau_n^{\frac{2-p}{2}}\log(n)) \qquad (4.90)\\ &+\mathcal{O}_{a.s.}(d^{3/2}C_\xi^{2/p}n^{1/2}m^{p/2-1}\tau_n^{2-p}\log(n))\\ &=\mathcal{O}_{a.s.}(d^{3/2}C_\xi^{2/p}n^{1/2}m^{-1/2}\log(n))+\mathcal{O}_{a.s.}(d^{3/2}C_\xi^{2/p}n^{1/2}m^{p/4-1/2}\tau_n^{\frac{2-p}{2}}\log(n)).\end{aligned}$$

Here the first inequality follows from (4.88) and the last equality follows since τ_n and m should be chosen such that $m^{p/2-1}\tau_n^{2-p}$ tends to zero and hence $m^{p/2-1}\tau_n^{2-p}\ll m^{p/4-1/2}\tau_n^{1-p/2}$.

Now choosing the block size m and the truncation level τ_n as follows

$$m=\lfloor n^\alpha\rfloor,\quad \alpha:=\frac{p-2}{2(p-1)},\quad \tau_n:=n^{\alpha/2-\alpha/p+1/p},$$

we have

$$\mathbb{E}\left[\left|\sum_{k=1}^{n}\xi_k-\sum_{j=1}^{\lfloor n/m\rfloor}\tilde{x}_j\right|^2\right]=\mathcal{O}(dC_\xi^{2/p}n^{\frac{p}{2(p-1)}})$$

and consequently

$$\begin{aligned}|Y_n-\tilde{Y}_n|&=\mathcal{O}_{a.s.}(d^{3/2}C_\xi^{2/p}n^{1/2}m^{-1/2}\log(n))+\mathcal{O}_{a.s.}(d^{3/2}C_\xi^{2/p}n^{1/2}m^{p/4-1/2}\tau_n^{\frac{2-p}{2}}\log(n))\\&=\mathcal{O}_{a.s.}\left(d^{3/2}C_\xi^{2/p}n^{\frac{1}{4}+\frac{1}{4(p-1)}}\log(n)\right)+\mathcal{O}_{a.s.}\left(d^{3/2}C_\xi^{2/p}n^{\frac{1}{2(p-1)}})\log(n)\right)\\&=\mathcal{O}_{a.s.}\left(d^{3/2}C_\xi^{2/p}\ n^{\frac{1}{4}+\frac{1}{4(p-1)}}\log(n)\right).\end{aligned}\tag{4.91}$$

Note that since $p>2$ also we have that for the specified choice of τ_n and m that (4.82) is satisfied. Furthermore, from (4.85)-(4.87) we see that

$$\left|\sum_{k=1}^{n}\xi_k-\tilde{Y}_n\right|=\mathcal{O}_{a.s.}(C_\xi^{1/p}d^{3/2}n^{\frac{1}{4}+\frac{1}{4(p-1)}}\log n)\tag{4.92}$$

Note that since $\{\xi_k\}$ is a one dependent identically distributed sequence, we have that $\mathrm{Cov}\left(\sum_{k=1}^{n}\xi_k\right)=n\,\mathrm{Var}(\xi_1)+(n-1)\left(\mathrm{Cov}(\xi_1,\xi_2)+\mathrm{Cov}(\xi_1,\xi_2)^T\right)$. Consequently,

$$Y_n=\left(\mathrm{Var}(\xi_1)+\frac{n-1}{n}\left(\mathrm{Cov}(\xi_1,\xi_2)+\mathrm{Cov}(\xi_1,\xi_2)^T\right)\right)^{1/2}W(n).$$

Furthermore, we have that the asymptotic covariance matrix is given by $\Sigma_\xi=\mathrm{Var}(\xi_1)+\mathrm{Cov}(\xi_1,\xi_2)+\mathrm{Cov}(\xi_1,\xi_2)^T$. This gives us

$$\begin{aligned}\left|Y_n-\Sigma_\xi^{1/2}W(n)\right|&\leq\frac{1}{n(\sigma_0)^{1/2}}\left|\mathrm{Cov}(\xi_1,\xi_2)+\mathrm{Cov}(\xi_1,\xi_2)^T\right||W(n)|\\&=\frac{\mathcal{O}_{a.s.}(\sqrt{d}\log(n))}{(\sigma_0 n)^{1/2}}|\mathrm{Cov}(\xi_1,\xi_2)|,\end{aligned}\tag{4.93}$$

where the first inequality follows by Ando–van Hemmen's inequality, Lemma 4.5.17. We note that since (4.93) is of smaller asymptotic magnitude that the other approximation terms.

Overall, from (4.91), (4.92), (4.93) we have

$$\left|\sum_{k=1}^{n}\xi_k-\Sigma_\xi^{1/2}W(n)\right|=\mathcal{O}_{a.s.}\left(\left(d^{25/4+\theta_0}C_\xi^{2/p}\right)\left(\frac{\sigma_d}{\sigma_0}\right)^{1/2}n^{\frac{1}{4}+\frac{1}{4(p-1)}}\log(n)\right).$$

This finishes the proof for the strong Gaussian approximation case. The weak Gaussian approximation case follows from the same argument with minor differences. The key point is that instead of applying Lemma 4.5.15 for approximating the truncated sequence $(\tilde{x}_k)_{k\in\mathbb{N}}$, we now make use of Lemma 4.5.13 instead. □

4.5.4. Proofs of Section 4.3

Proof of Theorem 4.3.2

Proof. Since the minorisation condition holds for some $m_0 > 1$, we have by Proposition 4.5.1 there exists a sequence of randomized stopping times $\{R_k\}$ such that we can define

$$\xi_k := \sum_{t=R_{k-1}}^{R_k-1} (f(X_t) - \pi(f)), \quad k \geq 1,$$

such that $(\xi_k)_{k\in\mathbb{N}}$ is a stationary one-dependent sequence under $\mathbb{P}_\nu$. Furthermore, by Asmussen [5; Theorem 3.2], we have that

$$\mathbb{E}_\nu \xi_1 = \mathbb{E}_\nu \sum_{t=0}^{R_1} \{f(X_t) - \pi(f)\} \ = \mu_\varrho \cdot \pi(f - \pi(f)) = 0.$$

Moreover, under Assumption A4.1, we have that

$$\sup_{i\in\{1,\ldots,d\}} \mathbb{E}_\nu \left[\left(\sum_{t=0}^{R_1} |f_i(X_t)| \right)^{p_0} \right] \lesssim \alpha^{-1} \left(\mathbb{E}_\nu [h(R_1)]^{\varepsilon/p_0} \right., \tag{4.94}$$

where $h(x) = x^q$ with $q = \eta/(1-\eta)$ and p_0 is defined in (4.58) in the case of a polynomial drift condition and $h(x) = e^{tx}$ with $|t| < -\ln(\lambda) m_0$ and $p_0 = p$ in the case of an exponential drift condition. By Theorem 4.5.19, we can redefine $(\xi_k)_k$ on a new probability space on which we can also construct a standard d-dimensional Brownian motion W such that

$$\left| \sum_{k=1}^{n} \xi_k - n\mathbb{E}_\nu \xi_1 - W(n) \right| = \begin{cases} \mathcal{O}_P \left(d^{3/2} \tilde{C}_N^{2/p} \left(\dfrac{\sigma_d}{\sigma_0} \right) T^{\frac{1}{4} + \frac{1}{4(p_0-1)}} \right), \\ \mathcal{O}_{a.s.} \left(d^{25/4+\theta_0} \log^*(d) \tilde{C}_N^{2/p} \left(\dfrac{\sigma_d}{\sigma_0} \right)^{1/2} T^{\frac{1}{4} + \frac{1}{4(p_0-1)}} \right), \end{cases} \tag{4.95}$$

where

$$C_N = \alpha^{-1} \left(\mathbb{E}_\nu[h(R_1])^{\varepsilon/p_0} .$$

In the remainder of the proof, we will consider the strong Gaussian approximation case, since the weak approximation case follows from similar and slightly easier arguments. Note, that since we assume that $X_0 \sim \pi$, we need to discard the first cycle. Similarly, in order to apply (4.95), we will need to discard the last cycle. Let $\eta(T)$ denote the number of regenerations of the m_0–skeleton chain up to time T, namely

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

It immediately follows that

$$\left| \sum_{t=0}^{T} (f(X_t) - \pi(f)) \ - \xi_0 - \sum_{k=1}^{\eta(T)} \xi_k \right| = |\xi_0| + \left| \sum_{t=R_{\eta(T)}}^{T} (f(X_t) - \pi(f)) \right|, \quad (4.96)$$

where

$$\xi_0 := \sum_{t=0}^{R_0} (f(X_t) - \pi(f)).$$

By a Borel–Cantelli argument, we will show that

$$\left| \sum_{t=R_{\eta(T)}}^{T} (f(X_t) - \pi(f)) \right| = \mathcal{O}_{a.s.} \left(\alpha^{-1/p} d^{1/2} n^{1/p_0} h(R_1)^{\varepsilon/p_0^2} \right) \quad (4.97)$$

and

$$|\xi_0| = \mathcal{O}_{a.s.} \left(d^{1/2} n^{1/p_0} h(R_1)^{\varepsilon/p_0^2} \right) . \quad (4.98)$$

In order to show (4.97), note that

$$\left| \sum_{t=R_{\eta(T)}}^{T} (f(X_t) - \pi(f)) \right| \leq \left| \sum_{t=R_{\eta(T)}}^{R_{\eta(T)+1}} g(X_t) \right|, \quad (4.99)$$

where $g(x) = (|f_1(x) - \pi(f_1)|, \ldots, |f_d(x) - \pi(f_d)|)^T$. Now let $\varepsilon > 0$ be given and introduce the event

$$A_n = \left\{ \left| \sum_{t=R_n}^{R_{n+1}} g(X_t) \right| > \alpha^{-1/p} d^{1/2} n^{1/p_0} h(R_1)^{\varepsilon/p_0^2} \right\} .$$

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

$$\begin{aligned}\sum_{n=1}^{\infty}\mathbb{P}_\nu\left(A_n\right) &\leq \sum_{n=1}^{\infty}\mathbb{P}_\nu\left(\left|\sum_{t=R_n}^{R_{n+1}} g(X_t)\right| > \alpha^{-1/p} d^{1/2} n^{1/p_0} h(R_1)^{\varepsilon/p_0^2}\right)\\ &= \sum_{n=1}^{\infty}\mathbb{P}_\nu\left(\left|\sum_{t=0}^{R_1} g(X_t)\right|^{p_0} > \alpha^{-1} d^{p_0/2} n h(R_1)^{\varepsilon/p_0}\right)\\ &\leq \mathbb{E}_\nu\left[\left|\sum_{t=0}^{R_{n1}} g(X_t)\right|^{p}\right](\alpha^{-1} d^{p_0/2} n h(R_1)^{\varepsilon/p_0})^{-1} < C,\end{aligned}$$

where C is some universal constant and the one for the last inequality follows by (4.54) of Lemma 4.5.8 and 4.5.9 in the case of a geometric and polynomial drift condition respectively. By the Borel–Cantelli lemma it follows that $\mathbb{P}_\nu(\limsup A_n) = 0$. Consequently, we have that $\mathbb{P}_\nu(\liminf A_n^c) = 1$. Hence it follows that

$$\left|\sum_{t=R_n}^{R_{n+1}} g(X_t)\right| = \mathcal{O}_{a.s.}(\alpha^{-1/p} d^{1/2} n^{1/p_0} h(R_1)^{\varepsilon/p_0^2}). \tag{4.100}$$

Moreover, since $\eta(T)$ is almost surely increasing and $\eta(T) = \mathcal{O}_{a.s.}(T)$, it follows that

$$\left|\sum_{t=R_n}^{R_{n+1}} g(X_t)\right| = \mathcal{O}_{a.s.}(\alpha^{-1/p} d^{1/2} \eta(T)^{1/p_0} h(R_1)^{\varepsilon/p_0^2}) = \mathcal{O}_{a.s.}(\alpha^{-1/p} d^{1/2} n^{1/p_0} h(R_1)^{\varepsilon/p_0^2}).$$

Hence the claim formulated in (4.97) directly follows. The claim (4.98) follows completely analogously to (4.100). Since we have that $\mathbb{E}_\nu R_1^{p_0} < \infty$, by Csörgö and Horváth [38; Theorem 2.4] we can construct a Brownian motion $\tilde{W}$ such that

$$\left|\eta(T) - \frac{T}{\mu_\varrho} - \frac{\sigma_\varrho}{\mu_\varrho^{3/2}}\tilde{W}_T\right| = o_{a.s.}(T^{1/p_0}), \tag{4.101}$$

By the law of iterated logarithm for Brownian motion we obtain

$$\eta(T) = \frac{T}{\varrho} + \mathcal{O}_{a.s.}(\sqrt{T\log\log T}) \quad \text{a.s.} \tag{4.102}$$

Furthermore, by (4.95), 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

$$\frac{1}{\left(\left(d^{25/4+\theta_0}\log^*(d)\tilde{C}_N^{2/p}\left(\frac{\sigma_d}{\sigma_0}\right)^{1/2}\right)n^{\frac{1}{4}+\frac{1}{4(p_0-1)}}\log n\right)}\left|\sum_{k=1}^{n}\xi_k(\omega)-W(n)\right| < C(\omega) \tag{4.103}$$

Since $\eta(T)$ 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 $\eta(T)(\omega) \geq N_0$ for all $T \geq T_0$. Hence we obtain from (4.103) that

$$\limsup_{T\to\infty}\frac{\left|\sum_{k=1}^{\eta(T)}\xi_k-\Sigma_\xi^{1/2}W(\eta(T))\right|}{\left(d^{25/4+\theta_0}\log^*(d)\tilde{C}_N^{2/p}\left(\frac{\sigma_d}{\sigma_0}\right)^{1/2}\right)\eta(T)^{\frac{1}{4}+\frac{1}{4(p_0-1)}}\log\eta(T)} < C \quad \text{a.s. .} \tag{4.104}$$

We see that (4.104) can be reformulated as

$$\left|\sum_{k=1}^{\eta(T)}\xi_k-(\Sigma_\xi)^{1/2}W(\eta(T))\right| \tag{4.105}$$

$$= \mathcal{O}_{a.s.}\left(d^{25/4+\theta_0}\log^*(d)\tilde{C}_N^{2/p}\left(\frac{\sigma_d}{\sigma_0}\right)^{1/2}\eta(T)^{\frac{1}{4}+\frac{1}{4(p_0-1)}}\log\eta(T)\right)$$

$$= \mathcal{O}_{a.s.}\left(d^{25/4+\theta_0}\log^*(d)\tilde{C}_N^{2/p}\left(\frac{\sigma_d}{\sigma_0}\right)^{1/2}T^{\frac{1}{4}+\frac{1}{4(p_0-1)}}\log T\right) \tag{4.106}$$

Here the second equality follows by (4.102). By a coordinate-wise application of Csörgö and Révész [40; Theorem 1.2.1]) it follows that

$$\begin{aligned}\left|W(\eta(T))-W(T/\mu_\varrho)\right| &\leq \left(\sum_{i=1}^{d}\left|W_i(\eta(T))-W_i(T/\mu_\varrho)\right|^2\right)^{1/2}\\ &= \mathcal{O}_{a.s.}(\sqrt{d}T^{1/4}\log T).\end{aligned} \tag{4.107}$$

From (4.96) and combining results (4.97), (4.98), (4.105), and (4.107) the asserted theorem follows. □

Proof of Theorem 4.3.1

Proof. We will first assume that $X_0 \sim \nu$ and consider the case where the geometric drift condition holds. Since the minorisation condition holds with $m_0 = 1$, we have by Proposition 4.5.1 there exists a sequence of randomized stopping times $\{R_k\}$ such that we can define

$$\xi_k := \sum_{t=R_{k-1}}^{R_k-1} (f(X_t) - \pi(f)), \quad k \geq 1,$$

such that $(\xi_k)_{k\in\mathbb{N}}$ is a mean-zero independent and identically distributed sequence under $\mathbb{P}_\nu$.

By Lemma 4.5.8 we have that for any t with $|t| \leq \ln(1/\lambda)/m_0$ we have that

$$\mathbb{E}_\nu[e^{tR_1}] \lesssim \frac{b}{\alpha\lambda(1-\lambda)} \tag{4.108}$$

and

$$\sup_{i\in\{1,\ldots,d\}} \mathbb{E}_\nu\left[\left(\sum_{t=0}^{R_1} |f_i(X_t)|\right)^{p_0}\right] \lesssim \alpha^{-1}\left(\mathbb{E}_\nu[h(R_1])^{\varepsilon/p_0}\right., \tag{4.109}$$

Note that from (4.108) we also have that

$$\mathbb{E}_\nu[R_1^p] \leq \left(\frac{p}{\ln(1/\lambda)e}\right)^p \left(\frac{b}{\alpha\lambda(1-\lambda)}\right).$$

Define $(\varrho_k)_{k\in\mathbb{N}}$ as $\varrho_k = R_k - R_{k-1}$. Note that this is an i.i.d sequence and let μ_ϱ and σ_ϱ^2 denote the respective mean and variance. The sequence of $(d+1)$-random vectors (ξ_k, ϱ_k) are independent dependent and identically distributed. Introduce the sequence $(\tilde{\xi}_k)_{k\in\mathbb{N}}$ as $\tilde{\xi}_k = \xi_k - \beta(\varrho_k - \mu_\varrho)$. Note that the sequence of random vectors $(\tilde{\xi}_k, \varrho_k)$ are also independent and identically distributed. If we choose $\beta = \mathrm{Cov}_\nu(\xi_1, \varrho_1)/\sigma_\varrho^2$, then it immediately follows that ρ_k and every component of $\tilde{\xi}_k$ are uncorrelated. Let $\tilde{\Sigma}$ denote the limiting covariance of $\frac{1}{n}\sum_k(\tilde{\xi}_k, \varrho_k)$. We see that $\tilde{\Sigma}$ is given by

$$\mathrm{Var}_\nu\begin{pmatrix}\tilde{\xi}_1\\ \varrho_1\end{pmatrix} = \mathrm{Var}_\nu\begin{pmatrix}\xi_1 - \beta(\varrho_1 - \mu_\varrho)\\ \varrho_1\end{pmatrix} = \begin{pmatrix}V_\xi & \underline{0}\\ \underline{0}^T & \sigma_\varrho^2\end{pmatrix}, \tag{4.110}$$

where by definition of β we see that the off-diagonal entries of block matrix (4.110) are $\underline{0}$, which denotes a d-dimensional vector of zeros, and V_ξ is given by

$$\mathrm{Var}_\nu(\xi_1 - \beta(\varrho_1 - \mu_\varrho)) = \mathrm{Var}_\nu(\xi_1) + \beta\beta^T\sigma_\varrho^2 - 2\beta\beta^T\sigma_\varrho^2 = \mathrm{Var}_\nu(\xi_1) - \beta\beta^T\sigma_\varrho^2. \tag{4.111}$$

By Sigman [140; Theorem 9], we have that

$$\frac{\mathrm{Var}_\nu(\xi_1)}{\mu_\varrho} = \Sigma_f. \tag{4.112}$$

Hence, we see that the conditions of Theorem 4.5.18 are satisfied with

$$\begin{aligned}\sup_{i\in\{1,\dots,d\}} \mathbb{E}_\nu \tilde{\xi}_{1i}^p &\le \left(1 + \left(\frac{p}{\ln(1/\lambda)e}\right)^p \left(\frac{b}{\alpha\lambda(1-\lambda)}\right)\right) \alpha^{-1} \left(\frac{b}{\alpha(1-\lambda)}\right)^{\varepsilon/p} \sup_{i\in\{1,\dots,d\}} \pi(|f_i|^{p+\varepsilon}) \\ &\le 2\alpha^{-1} \left(\frac{b}{\alpha(1-\lambda)}\right)^{1+\varepsilon/p} \left(\frac{p}{\ln(1/\lambda)e}\right)^p \sup_{i\in\{1,\dots,d\}} \pi(|f_i|^{p+\varepsilon}) =: C_\xi\end{aligned}$$

Applying the multivariate Gaussian approximation given in Theorem 4.5.18, we have that

$$\left|\sum_{k=1}^n \begin{pmatrix} \tilde{\xi}_k \\ \varrho_k \end{pmatrix} - \sum_{k=1}^n \begin{pmatrix} Y_k^1 \\ Y_k^2 \end{pmatrix}\right| = \mathcal{O}_{a.s.}\left(C_\xi \Psi_n\right),$$

where

$$\Psi_n = n^{1/p}\log n,$$

and

$$\tilde{C}_{\xi,d} := \left(d^{25/4+\theta_0}\log^*(d) C_\xi^{1/p} \vee d^{3/2} C_\xi^{2/p}\right)\left(\frac{\sigma_d}{\sigma_0}\right)^{1/2},$$

and $\sum_{k=1}^n \binom{Y_k^1}{Y_k^2}$ has a Gaussian distribution with the same mean and covariance matrix as $\sum_{k=1}^n \binom{\tilde{\xi}_k}{\varrho_k}$. Given the block structure of $\tilde{\Sigma}$ given in (4.110), we see that Y_2^k is independent of all components of Y_k^1. By the Skorohod embedding theorem, two independent Brownian motions B_1 and B_2 can be constructed, where B_1 is a d-dimensional Brownian motion and B_2 is one-dimensional, such that they coincide with the Gaussian sequences at all integer time points. Therefore we have that

$$\left|\sum_{k=1}^{n}\xi_k - \beta(\sum_{k=1}^{n}\varrho_k - \mu_\varrho) - V_\xi^{\frac{1}{2}}B_1\right| = \mathcal{O}_{a.s.}\left(\tilde{C}_{\xi,d}\Psi_n\right) \tag{4.113}$$

and

$$|R_n - n\mu_\varrho - \sigma_\varrho B_2(n)| = \mathcal{O}_{a.s.}\left(\tilde{C}_{\xi,d}\Psi_n\right). \tag{4.114}$$

By Komlos [90; Theorem 1(ii)], a Poisson process L with intensity $\lambda = \mu_\varrho^2/\sigma_\varrho^2$ can be constructed from the one-dimensional Brownian motion B_2 such that

$$\left|L(n) - \frac{\mu_\varrho}{\gamma}n - \frac{\sigma_\rho}{\gamma}B_2(n)\right| = \mathcal{O}_{a.s.}(\log n), \tag{4.115}$$

where $\gamma = \sigma_\varrho^2/\mu_\varrho$ and L is constructed increment-wise from B_2 in a deterministic way and is therefore also independent of B_1. From (4.114) and (4.115) it follows that

$$|R_n - \gamma L(n)| = \mathcal{O}_{a.s.}(\tilde{C}_{\xi,d}\Psi_n). \tag{4.116}$$

We claim that it therefore follows that

$$\left|\sum_{k=1}^{n}\xi_k - \sum_{k=1}^{\gamma L(n)}\xi_n\right| = \left|\sum_{k=0}^{R_n}(f(X_k)-\pi(f)) - \sum_{k=0}^{\gamma L(n)}(f(X_k)-\pi(f))\right| = \mathcal{O}_{a.s.}(\tilde{C}_{\xi,d}\Psi_n) \tag{4.117}$$

We see that

$$\left|\sum_{k=0}^{R_n}(f(X_k)-\pi(f)) - \sum_{k=0}^{\gamma L(n)}(f(X_k)-\pi(f))\right| = \left|\sum_{b_n}^{c_n}(f(X_k)-\pi(f))\right| \tag{4.118}$$

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

$$\kappa_n := c_n - b_n = |R_n - \gamma L(n)|.$$

From (4.116) it follows that $\kappa_n = \mathcal{O}_{a.s.}(C_{\xi,N,d}\Psi_n)$, hence it follows that there exists some universal almost surely finite C such that for almost every ω it holds that there exists an $N_1 := N_1(\omega)$ such that for all $n \geq N_1$ we have that $\kappa_n < C(\omega)C_{\xi,N,d}\Psi_n$ and hence $c_n = b_n + \kappa_n \leq b_n + \bar{\kappa}_n$, with

$\bar{\kappa}_n = \lceil C(\omega) C_{\xi,N,d} \Psi_n \rceil$. 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 amount of regenerative cycles up to time T and let $(Y_k)_{k\in\mathbb{N}}$ be defined as

$$Y_k = \left| \sum_{t=R_k}^{R_{k+1}-1} g(X_t) \right|, \tag{4.119}$$

where $g(x) = (|f_1(x) - \pi(f_1)|, \ldots, |f_d(x) - \pi(f_d)|)^T$. Then we see that for $n > N_1(\omega)$ we have that

$$\begin{aligned}
\left| \Psi_n^{-1} \sum_{b_n}^{c_n} (f(X_k) - \pi(f)) \right| &= \Psi_n^{-1} \left| \sum_{k=0}^{c_n - b_n} (f(X_{b_n+u}) - \pi(f)) \right| \\
&\leq \Psi_n^{-1} \sum_{k=0}^{\kappa_n} g(X_{b_n+u}) \\
&\leq \Psi_n^{-1} \sum_{k=0}^{\bar{\kappa}_n} g(X_{b_n+k}) \\
&= \Psi_n^{-1} \sum_{j=\eta(b_n)}^{\eta(b_n+\bar{\kappa}_n)} Y_j + \Psi_n^{-1} \sum_{t=R_{\eta(b_n+\bar{\kappa}_n)}}^{b_n+\bar{\kappa}_n} |f(X_t) - \pi(f)|
\end{aligned} \tag{4.120}$$

From (4.102) we see that $\eta(T)$ tends to infinity as $T \to \infty$ and $\lim_{T\to\infty} \eta(T)/T = 1/\mu_\varrho$ almost surely. Also for every positive sequence m_T that tends to infinity as $T \to \infty$, we have that $\lim_{T\to\infty} \eta(m_T)/m_T = 1/\mu_\varrho$ almost surely. By an application of the law of iterated logarithm to (4.114) and (4.115), we see that

$$R_n = n/\mu_\varrho + \mathcal{O}_{a.s.}(\sqrt{n \log\log n})$$

and

$$L_n = n/\lambda + \mathcal{O}_{a.s.}(\sqrt{n \log\log n}).$$

Consequently, we have that have that $b_n = \mathcal{O}_{a.s.}(n)$ and $\eta(b_n) = \mathcal{O}_{a.s.}(n)$. Note that $\eta(b_n + \bar{\kappa}_n)$, the amount of regenerations until time $b_n + \bar{\kappa}_n$ is equal to the amount of generation until time b_n and the amount of regenerations in the time interval $(b_n, b_n + \bar{\kappa}_n)$, i.e., $\eta(b_n + \bar{\kappa}_n) = \eta(b_n) + \eta(b_n + \bar{\kappa}_n) - \eta(b_n)$. Since $\eta(T)$ is a renewal process it is clear, that the amount of events should

be proportional to the time interval and the intensity, i.e., that we should have $\eta(b_n + \bar{\kappa}_n) - \eta(b_n) = O(\bar{\kappa}_n/\mu_\varrho)$ almost surely. We will now prove this claim. Since we have that $\mathbb{E}_\nu R_1^p < \infty$, by Csörgö and Horváth [38; Theorem 2.4] we can construct a Brownian motion $\tilde{B}_2$ such that

$$\left|\eta(T) - \frac{T}{\mu_\eta} - \sigma_\eta \tilde{B}_2(T)\right| = o_{a.s.}(T^{1/p}), \tag{4.121}$$

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 $T \geq T_1(\omega)$ we have that

$$\frac{1}{T^{1/p}}\left|\eta(T) - \frac{T}{\mu_\eta} - \sigma_\eta \tilde{B}_2(T)\right| < \varepsilon. \tag{4.122}$$

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 $\eta(b_n)(\omega) \geq T_1(\omega)$ for all $n \geq N_2(\omega)$ and hence

$$\frac{1}{b_n^{1/p}}\left|\eta(b_n) - \frac{b_n}{\mu_\eta} - \sigma_\eta \tilde{B}_2(b_n)\right| < \varepsilon. \tag{4.123}$$

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

$$\left|\eta(b_n) - \frac{b_n}{\mu_\eta} - \sigma_\eta \tilde{B}_2(b_n)\right| = o_{a.s.}(b_n^{1/p}) = o_{a.s.}(n^{1/p}). \tag{4.124}$$

Then by the triangle inequality, we obtain

$$\eta(b_n + \bar{\kappa}_n) - \eta(b_n) \leq \left|\eta(b_n + \bar{\kappa}_n) - (b_n + \bar{\kappa}_n)/\mu_\eta - \sigma_\eta \tilde{B}_2(\eta(b_n) + \bar{\kappa}_n)\right| \tag{4.125}$$

$$+ \bar{\kappa}_n/\mu_\eta + \left|-\eta(b_n) + b_n/\mu_\eta + \sigma_\eta \tilde{B}_2(b_n)\right| \tag{4.126}$$

$$+ \sigma_\eta\left|\tilde{B}_2(b_n + \bar{\kappa}_n) - \tilde{B}_2(b_n)\right| \tag{4.127}$$

$$\leq \bar{\kappa}_n/\mu_\eta + o_{a.s.}(n^{1/p}). \tag{4.128}$$

The last inequality follows, since by (4.124) the term in (4.125) and the second term in (4.126) are $o(n^{1/p})$. Furthermore, by Csörgö and Révész [39; Theorem 2] we have that for any $a_n \ll n$ that

$$\limsup_{n\to\infty} \sup_{0\leq s\leq a_n} \frac{\left|\tilde{B}_2(n+s) - \tilde{B}_2(n)\right|}{[a_n(\log(n/a_n) + \log\log n)]^{1/2}} = 1 \quad \text{a.s.} \tag{4.129}$$

Since we have $\bar{\kappa}_n = \mathcal{O}(n^{1/p})$, it follows that

$$\sup_{0\leq s\leq \bar{\kappa}_n} \left|\tilde{B}_2(n+s) - \tilde{B}_2(n)\right| = \mathcal{O}_{a.s.}\left(n^{1/2p}\log(n)\right). \tag{4.130}$$

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

$$\sup_{0\leq s\leq \bar{\kappa}_n} \left|\tilde{B}_2(\eta(b_n)+s) - \tilde{B}_2(\eta(b_n))\right| = o_{a.s.}\left(\eta(b_n)^{1/2p}\right) = o_{a.s.}\left(n^{1/2p}\right). \tag{4.131}$$

Hence, the inequality in (4.128) follows and we have shown that $\eta(b_n+\bar{\kappa}_n) - \eta(b_n) \leq \bar{\kappa}_n\mu_\eta + o\left(n^{1/p}\right)$ 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 $\eta(b_n+\bar{\kappa}_n) - \eta(b_n) < Kn^{1/p}$ almost surely. For notational convenience let $\tilde{a}_n$ be defined as $Kn^{1/p}$. Since $(Y_k)_{k\geq 0}$ form an i.i.d sequence we have by Lemma 4.5.8 and Theorem 4.5.18 that there exists a Brownian motion B_3 such that

$$\left|\sum_{k=0}^{n} Y_k - n\mu_Y - \Sigma_Y^{1/2}B_3(n)\right| = \mathcal{O}_{a.s.}(\tilde{C}_{\xi,d}\Psi_n). \tag{4.132}$$

where μ_Y and $\Sigma_Y^{1/2}$ denote the mean and square root of the covariance matrix of Y_1 respectively. It immediately follows that we also have

$$\left|\sum_{k=0}^{\eta(b_n)} Y_k - \eta(b_n)\mu_Y - \Sigma_Y^{1/2}B_3(\eta(b_n))\right| = \mathcal{O}_{a.s.}(\tilde{C}_{\xi,d}\Psi_{\eta(b_n)}) = \mathcal{O}_{a.s.}(\tilde{C}_{\xi,d}\Psi_n). \tag{4.133}$$

By the triangle inequality, we obtain

$$\left|\sum_{k=\eta(b_n)}^{\eta(b_n)+\tilde{a}_n} Y_k\right| \leq \left|\sum_{k=0}^{\eta(b_n)+\tilde{a}_n} Y_k - (\eta(b_n)+\tilde{a}_n)\mu_Y - \Sigma_Y^{1/2}B_3(\eta(b_n)+\tilde{a}_n))\right| \tag{4.134}$$

$$+ \tilde{a}_n\mu_Y + \left|-\sum_{k=0}^{\eta(b_n)} Y_k + \eta(b_n)\mu_Y + \Sigma_Y^{1/2}B_3(\eta(b_n))\right| \tag{4.135}$$

$$+ \left|\Sigma_Y^{1/2}\right||B_3(\eta(b_n)+\tilde{a}_n) - B_3(\eta(b_n))| \tag{4.136}$$

$$\leq \tilde{a}_n\mu_Y + \mathcal{O}_{a.s.}(\tilde{C}_{\xi,d}\Psi_n). \tag{4.137}$$

The last inequality follows, since by (4.133) both the term in (4.134) and the second term in (4.136) are $o(C_{\xi,N,d}n^{1/p})$ almost surely. By again applying Csörgö and Révész [39; Theorem 2] to every coordinate of B_3 we see that

$$\left|B_3(\eta(b_n)+\tilde{a}_n)-B_3(\eta(b_n))\right| = \left(\sum_{i=1}^{d}\left(B_{3i}(\eta(b_n)+\tilde{a}_n)-B_{3i}(\eta(b_n))\right)^2\right)^{1/2} \tag{4.138}$$

$$= \mathcal{O}_{a.s.}(d^{1/2}n^{1/2p}\log n). \tag{4.139}$$

Furthermore, by (4.131) the last inequality also follows. Hence it follows that

$$\mathbb{P}_\nu\left(\limsup_{n\to\infty}\frac{1}{C_{\xi,N,d}n^{1/p}}\left|\sum_{k=\eta(b_n)}^{\eta(b_n+a_n)}Y_k\right|\leq K\mu_Y\right)=1. \tag{4.140}$$

Hence the first term in the upper bound (4.120) is $O(1)$ almost surely. For the second term, we see that by a Borel–Cantelli argument that is the same as the one given to obtain (4.100) that

$$Y_n=\sum_{R_n}^{R_{n+1}}|f(X_k)-\pi(f)| = \mathcal{O}_{a.s.}(\alpha^{-1/p}d^{1/2}C_\xi^{1/p}n^{1/p}). \tag{4.141}$$

Therefore

$$\sum_{R_{\eta(b_n+\bar{\kappa}_n)}}^{b_n+\bar{\kappa}_n}|f(X_{R_n+s})-\pi(f)|ds \leq \sum_{R_{\eta(b_n+\bar{\kappa}_n)}}^{R_{\eta(b_n+\bar{\kappa}_n)+1}}|f(X_{R_n+s})-\pi(f)|ds \tag{4.142}$$

$$= Y_{\eta(b_n+\bar{\kappa}_n)} \tag{4.143}$$

$$= \mathcal{O}_{a.s.}\left(\alpha^{-1/p}d^{1/2}C_\xi^{1/p}(\eta(b_n+\bar{\kappa}_n))^{1/p}\right) \tag{4.144}$$

$$= \mathcal{O}_{a.s.}\left(d^{1/2}C_\xi^{1/p}(n+C(\omega)C_{\xi,N,d}\Psi_n)^{1/p}\right) \tag{4.145}$$

$$= \mathcal{O}_{a.s.}\left(d^{1/2}C_\xi^{1/p}n^{1/p}\right) \tag{4.146}$$

Hence our claim (4.120) follows, and consequently we have also shown (4.117). Combining (4.113), (4.116), and (4.117) it follows that

$$\left|\sum_{k=1}^{\gamma L(n)} \xi_k - \beta\gamma L(n) + \beta\varrho n - V_\xi^{\frac{1}{2}} B_1(n)\right| = \mathcal{O}_{a.s.}\left(\tilde{C}_{\xi,d} n^{1/p}\right) \tag{4.147}$$

Let $(\Gamma_s)_{s\geq 0}$ be defined as $\Gamma_0 := 0$ and $\Gamma_s := L^{-1}(s)$, the generalised inverse of the Poisson process. Taking $n' = \Gamma_n$ in (4.147) and subsequently making the substitution $n = n'/\gamma$, it follows that

$$\left|\sum_{k=1}^{n} \xi_k - \beta n + \beta\varrho\Gamma_{n/\gamma} - V_\xi^{\frac{1}{2}} B_1(\Gamma_{n/\gamma})\right| = \mathcal{O}_{a.s.}\left(\tilde{C}_{\xi,d}{\Gamma_n}^{1/p}\right) = \mathcal{O}_{a.s.}\left(\tilde{C}_{\xi,d} n^{1/p}\right) \tag{4.148}$$

Since Γ_n has a Gamma distribution, it follows from the Komlós–Major–Tusnády approximation [90; Theorem 1] that there exists a Brownian motion B_4 such that

$$\left|\Gamma_n - \frac{n}{\lambda} - \frac{1}{\lambda} B_4(n)\right| = \mathcal{O}_{a.s.}(\log n). \tag{4.149}$$

Since the Poisson process N is constructed deterministically from B_2 we have that N and its corresponding event time process Γ are independent of B_1. Moreover, the components of a standard d-dimensional Brownian motion are all independent. Therefore by a componentwise application of Lemma 4.5.16 it follows that there exists a standard d-dimensional Brownian motion B_5 independent of N and Γ such that

$$\left|B_1(n) - \frac{1}{\sqrt{\lambda}} B_5(L(n))\right| = \left(\sum_{i=1}^{d} |B_{1i}(n) - \frac{1}{\sqrt{\lambda}} B_{5i}(L(n))|^2\right)^{1/2} = \mathcal{O}_{a.s.}(\sqrt{d}\log n). \tag{4.150}$$

Furthermore, by (4.150), there exists an almost surely finite random variable C such that for almost all ω we have that for all $n \geq N_0 \equiv N_0(\omega)$ we have that

$$\limsup_{n\to\infty} \frac{1}{d\log n}\left|B_1(n) - \frac{1}{\sqrt{\lambda}} B_5(L(n))\right| < C(\omega) \tag{4.151}$$

Since Γ_n is an increasing process and tends to infinity, we have that for almost every ω there exists a $N_0' \equiv N_0'(\omega)$ such that $\Gamma_n(\omega) \geq N_0$ for all $n \geq N_0'$. Hence we obtain from (4.151) that

$$\limsup_{n\to\infty} \frac{1}{d\log \Gamma_n}\left|B_1(\Gamma_n) - \frac{1}{\sqrt{\lambda}} B_4(n)\right| < C(\omega) \quad \text{a.s. ,} \tag{4.152}$$

where we used that $L(\Gamma(n)) = n$. Therefore we see that

$$\left| B_1(\Gamma_n) - \frac{1}{\sqrt{\lambda}} B_5(n) \right| = \mathcal{O}_{a.s.}(d \log \Gamma_n) = \mathcal{O}_{a.s.}(d \log n). \tag{4.153}$$

Here the last equality follows since we can apply the law of iterated logarithm for Brownian motion to (4.149) which gives

$$\Gamma_n = \frac{n}{\lambda} + \mathcal{O}_{a.s.}(\sqrt{n \log \log n}).$$

Applying the obtained approximations given in (4.149) and (4.153) to (4.148) we see that

$$\left| \sum_{k=1}^{n} \xi_k - \left(\frac{V_\xi^{\frac{1}{2}}}{\sqrt{\lambda\gamma}} B_5(t) - \frac{\beta\varrho}{\lambda\sqrt{\gamma}} B_4(t) \right) \right| = \mathcal{O}_{a.s.}\left(\tilde{C}_{\xi,d} \Psi_n \right). \tag{4.154}$$

Note that since B_4 is independent of all components of B_5 we have that

$$W_t = \Sigma_f^{-1} \left(\frac{V_\xi^{\frac{1}{2}}}{\sqrt{\lambda\gamma}} B_5(t) - \frac{\beta\varrho}{\lambda\sqrt{\gamma}} B_4(t) \right) \tag{4.155}$$

is a standard d-dimensional Brownian motion since

$$\frac{\tilde{V}_\xi}{\gamma\lambda} + \frac{\beta\beta^T\varrho^2}{\gamma\lambda^2} = \frac{\mathrm{Var}_\nu(\xi_1)}{\mu_\varrho} = \Sigma_f, \tag{4.156}$$

where the last equality follows from Sigman [140; Theorem 9]. Note that by the same argument as given in Theorem 4.3.2, if we assume that the initial distribution is π, it can shown that the initial cycle is asymptotically negligible. Furthermore, in the case where the polynomial drift condition holds, Lemma 4.5.9 gives us the required moment conditions and the proof of the Theorem follows analogously.

□

Remark 4.5.20. Note that we can extend Theorem 4.3.1 and Theorem 4.3.2 to arbitrary initial distributions, provided that the first cycle, until the first draw from the small measure ν, is asymptotically negligible. For $g : \mathbb{E} \to \mathbb{R}$ we can define the norm $|g|_V := \sup_{x \in E} \frac{g(x)}{V(x)}$. Then if we assume that a geometric drift condition holds, we have for any f such that $\sup_{i=1,\cdots,d} |f_i|_V < \infty$ by an

application of the Comparison theorem; Meyn and Tweedie [107; Theorem 14.2.2] that

$$\left|\sum_{t=0}^{R_1} f(X_k)\right| \leq d \sup_{i=1,\cdots,d} \sum_{t=0}^{R_1} |f_i(X_k)| \tag{4.157}$$

$$\leq d \frac{\sup_{i=1,\cdots,d} |f_i|_V}{1-\lambda} \left(V(x) + b\mathbb{E}_x R_1\right), \tag{4.158}$$

where we used the fact that if $|g|_V < \infty$ then $g(x) \leq V(x)|g|_V$. By Lemma 4.5.2 and an application of Jensen's inequality, we have that

$$\left|\sum_{t=0}^{R_1} f(X_k)\right| \leq d \, \frac{\sup_{i=1,\cdots,d} |f_i|_V}{1-\lambda} \left(V(x) + b \log_r \left(\frac{\alpha G(r,x)}{1-(1-\alpha)r^a}\right)\right), \tag{4.159}$$

where

$$G(r,x) = V(x)\mathbb{1}_C(x) + r(\lambda \upsilon_C + b)\mathbb{1}_{C^c}(x)$$

and

$$a = 1 + \left(\ln \frac{\lambda \upsilon_V + b - \alpha}{1-\alpha}\right) / (\ln(\lambda^{-1})).$$

Similarly, if we assume that a polynomial drift condition holds, from Lemma 4.5.3 we see that,

$$\mathbb{E}_x[\tau_C] \leq \frac{1}{(1-\eta)c} \left(V^{1-\eta}(x) + (b^\eta + b_0)\mathbb{1}_C(x)\right).$$

A similar argument with the comparison theorem gives us

$$\left|\sum_{t=0}^{R_1} f(X_k)\right| \leq d \, \frac{\sup_{i=1,\cdots,d} |f_i|_{V^\eta}}{c} \left(V(x) + \frac{1}{(1-\eta)c} \left(V^{1-\eta}(x) + (b^\eta + b_0)\right)\right), \tag{4.160}$$

provided that $\sup_{i=1,\cdots,d} |f_i|_{V^\eta} < \infty$. △

4.5.5. Proofs of Section 4.4

Proof of Theorem 4.4.1

Theorem 4.5.21. *Suppose that $f : \mathbb{R}^N \to \mathbb{R}^d$, with $\sup_{i\in\{1,\cdots,d\}} \pi(|f_i|^{p+\varepsilon}) < \infty$ for some $p \geq 4$ and let X satisfy a strong Gaussian approximation with approximation error $\bar{\psi}_N \psi_d \Psi_T \log(T)$ with $\psi_d = d^a$ for some $a > 0$. Assume that Assumption 6 holds, and that*

$$\frac{\bar{\psi}_N d\psi_d \Psi_T \log(T)}{\ell_{T,d}^{1/2}} = o(1) \text{ and } \frac{\bar{\psi}_N^2 d\psi_d^2 \Psi_T^2 \log(T)}{T} = o(1), \tag{4.161}$$

then we have that $\hat{\Sigma}_T^{BM} \to \Sigma_f$ with probability 1 as $T \to \infty$. Moreover, if we assume that

$$\Psi_T = \begin{cases} \Psi_T^{(1)} := T^{1/p_0} \log(T), \\ \Psi_T^{(2)} := T^{1/4+1/4(p_0-1)} \log(T), \end{cases} \tag{4.162}$$

for some $p_0 > 2$ and we choose the simulation time

$$T = \begin{cases} \left(\bar{\psi}_N d\psi_d\right)^{\frac{2p_0}{(p_0-2)}(1+\bar{\delta})} & \text{under rate } \Psi_T^{(1)} \\ \left(\psi_N d^{1/4}\psi_d\right)^{\frac{p_0-1}{p_0-2}4(1+\bar{\delta})} & \text{under rate } \Psi_T^{(2)}, \end{cases} \tag{4.163}$$

for any $\bar{\delta} > 1/(1+a)$ then the choice of batch size $\ell_T = d^{-(p_0-2)/(2p_0(1+\bar{\delta}))}\lfloor T^{\alpha} \rfloor$ with

$$\alpha = \begin{cases} \frac{1}{2} + \frac{p_0-2}{2p_0(1+\bar{\delta})} + \frac{1}{p_0}, & \text{under rate } \Psi_T^{(1)} \\ \frac{3}{4} + \frac{1}{4(p_0-1)} + \frac{(p_0-2)}{4(p_0-1)(1+\bar{\delta})}, & \text{under rate } \Psi_T^{(2)}, \end{cases} \tag{4.164}$$

optimises the given convergence rate for $T \to \infty$.

Proof. In Vats et al. [152; Theorem 2] it is shown that for every i, j we have

$$\left|\widehat{\Sigma}_{T_{ij}}^{BM} - \Sigma_f\right| = \mathcal{O}_{a.s.}\left(\left(\frac{\ell_{T,d}}{T}\right)^{1/2}\right) + \mathcal{O}_{a.s.}\left(\bar{\psi}_N \psi_d \Psi_T \log(T) \ell_{T,d}^{-1/2}\right) + \mathcal{O}_{a.s.}\left(\frac{\bar{\psi}_N^2 \psi_d^2 \Psi_T^2 \log(T)}{T}\right).$$

Since we have that

$$\begin{aligned}\left|\widehat{\Sigma}_T^{BM} - \Sigma_f\right| &= \sqrt{\sum_{i=1}^{d}\sum_{j=1}^{d}\left|\widehat{\Sigma}_{T_{ij}}^{BM} - \Sigma_{f_{ij}}\right|^2} \\ &= \mathcal{O}_{a.s.}\left(d\left(\frac{\ell_{T,d}}{T}\right)^{1/2}\right) + \mathcal{O}_{a.s.}\left(\bar{\psi}_N d\psi_d \Psi_T \log(T)\ell_{T,d}^{-1/2}\right) \\ &+ \mathcal{O}_{a.s.}\left(\frac{\bar{\psi}_N^2 d\psi_d^2 \Psi_T^2 \log(T)}{T}\right) \\ &= o_{a.s.}(1).\end{aligned}$$

Firstly, note that if

$$T = \left(\bar{\psi}_N d\psi_d\right)^{\frac{2p_0}{(p_0-2)}(1+\bar{\delta})},$$

we have that for $T \to \infty$ that

$$\mathcal{O}_{a.s.}\left(\frac{\bar{\psi}_N^2 d\psi_d^2 \Psi_T^2 \log^2(T)}{T}\right) = \mathcal{O}_{a.s.}\left(d^{-1}\left(\bar{\psi}_N d\psi_d\right)^{-2\bar{\delta}} \log^3(\bar{\psi}_N d\psi_d)\right) = o_{a.s.}(1).$$

Note that in order to find the optimal batch size $\ell_T = \lfloor T^\alpha \rfloor$ such that $\left|\widehat{\Sigma}_T^{BM} - \Sigma_f\right|$ tends to zero at the fastest rate, we equate the error terms $\mathcal{O}_{a.s.}\left(d\left(\frac{\ell_{T,d}}{T}\right)^{1/2}\right)$ and $\mathcal{O}_{a.s.}\left(\bar{\psi}_N d\psi_d \Psi_T \log(T)\ell_{T,d}^{-1/2}\right)$. This gives us that the optimal batch size, up to a logarithmic factor, should be of an asymptotic magnitude

$$\begin{aligned}\ell_T &\asymp \bar{\psi}_N \psi_d \Psi_T T^{1/2} \\ &\asymp d^{-(p_0-2)/(2p_0(1+\bar{\delta}))} T^{1/2+1/p_0+(p_0-2)/(2p_0(1+\bar{\delta}))} \\ &\asymp d^{-(p_0-2)/(2p_0(1+\bar{\delta}))}\left(\bar{\psi}_N d\psi_d\right)^{1+\frac{p_0+2}{p_0-2}(1+\bar{\delta})},\end{aligned}$$

which gives us

$$\left|\widehat{\Sigma}_T^{BM} - \Sigma_f\right| = \mathcal{O}_{a.s.}\left(\sqrt{\frac{\bar{\psi}_N d^2 \psi_d}{T^{1/2-1/p_0}}}\right).$$

With our choice for the simulation time $T = \left(\bar{\psi}_N d\psi_d\right)^{\frac{2p_0}{(p_0-2)}(1+\bar{\delta})}$ we see that

$$\left|\widehat{\Sigma}_T^{BM} - \Sigma_f\right| = \mathcal{O}_{a.s.}\left(\sqrt{d}T^{\frac{-\bar{\delta}(p-2)}{4p_0(1+\delta)}}\right) = \mathcal{O}_{a.s.}\left(\sqrt{d}\left(\bar{\psi}_N d\psi_d\right)^{-\bar{\delta}/2}\right)$$
$$= \mathcal{O}_{a.s.}\left(\bar{\psi}_N^{-\bar{\delta}/2} d^{1/2-(a+1)\bar{\delta}/2}\right) = o_{a.s.}(1),$$

where the last equality follows since $\bar{\delta} > 1/(1+a)$. □

Remark 4.5.22. In the case that $\Psi_T = T^{1/4+1/4(p_0-1)}$ and $\psi_d = d^a$ for some given $a > 0$ we require

$$T = \left(\psi_N d^{1/4}\psi_d\right)^{\frac{p_0-1}{p_0-2}4(1+\bar{\delta})},$$

with $\delta > 1/(1+a)$ and consequently

$$\ell_T \asymp \bar{\psi}_N \psi_d \Psi_T T^{1/2}$$
$$\asymp T^{\frac{3}{4}+\frac{1}{4(p_0-1)}+\frac{(p_0-2)}{4(p_0-1)(1+\delta)}}$$

△

Proof of Theorem 4.4.4

Theorem 4.5.23. *Suppose that X satisfies the following strong Gaussian approximation*

$$\left|\sum_{t=1}^{T} f(X_t) - T\pi(f) - \Sigma_f^{1/2} W_T\right| = \mathcal{O}_{a.s.}\left(\bar{\psi}_N \psi_d \Psi_T\right), \tag{4.165}$$

with approximation error $\Psi_T = T^{1/p_0}\log(T)$ for some $p_0 > 4$ and $\psi_d = d^a$ for some $a > 0$. Let $T_1(\varepsilon)$ be given by

$$T_1(\varepsilon) = \inf\{t > 0 : \mathrm{Vol}(C(t))^{1/d} + \varepsilon\Lambda(t) < \varepsilon\}, \tag{4.166}$$

with $C(t)$ the confidence ellipsoid given in (4.23) and $\Lambda(t) = \mathbb{1}_{\{t<T^*(\varepsilon,d,N)\}} + t^{-1}$, with

$$T^*(\varepsilon, d, N) = \left(\bar{\psi}_N \left(\frac{\mathrm{tr}(\Sigma_f)}{\sigma_0}\right)^2 d^3\psi_d\right)^{\frac{2p_0}{(p_0-2)}(1+\bar{\delta}_1)} \left(\frac{1}{\varepsilon}\right)^{\frac{4p_0}{(p_0-2)}(1+\bar{\delta}_2)} \vee e^{\frac{10p_0}{p_0-2}} \tag{4.167}$$

for any $\bar{\delta}_1 > 3/(3+a)$ *and* $\bar{\delta}_2 > 0$. *Let* $\widehat{\Sigma}_T$ *in* (4.23) *denote the batch means estimator defined in* (4.16), *with batch size* ℓ_T *set as*

$$\ell_T = \bar{\psi}_N \psi_d T^{1/2+1/p_0}. \tag{4.168}$$

Suppose that Assumptions 4 and 5 hold. Then we have as $\varepsilon \downarrow 0$ *the following:*

1. *The asymptotic behaviour of the termination time* $T_1(\varepsilon)$ *is characterised by*

$$\frac{\varepsilon^2 T_1(\varepsilon)}{c_{\alpha,d}^{2/d} \det(\Sigma_f)^{1/d}} = 1 + o_{a.s.}\left(\log^2(\bar{\psi}_N d^{5/2}\psi_d)\bar{\psi}_N^{-\bar{\delta}_1/2} d^{-1/2}\varepsilon\right), \tag{4.169}$$

   where $c_{\alpha,d}$ *denotes the product of* $q_\alpha^{d/2}$ *and the volume of a standard d-dimensional hyper-sphere.*

2. *Asymptotic validity of the resulting confidence set*

$$\mathbb{P}_\pi\left(C(T_1(\varepsilon)) \ni \pi(f)\right) \to 1-\alpha. \tag{4.170}$$

Proof. To prove the first claim, we first show that the difference between termination rules based on the volumes of the ellipsoids based on the estimated and asymptotic covariance matrix tends to zero at an appropriate rate. Note that the volume of the confidence ellipsoid is given by

$$\mathrm{Vol}(C_T(\alpha)) = T^{-d/2} q_{\alpha,d}^{d/2} \frac{2\pi^{d/2}}{d\Gamma(d/2)} \det\left(\widehat{\Sigma}_T^{1/2}\right). \tag{4.171}$$

Furthermore under Assumption 5, we have that $\sigma_0^{1/2d} I_d \preccurlyeq \Sigma_f^{1/2d} \preccurlyeq \sigma_d^{1/2d} I_d$, where the matrix inequalities hold in the positive semi-definite sense. Furthermore, by Jacobi's formula, we have that

$$\frac{\partial \det(\Sigma)}{\partial \Sigma} = \det(\Sigma)\Sigma^{-1}. \tag{4.172}$$

Note that the choice $\Sigma = \sigma_d^{1/2d} I_d$ maximises the function $\det(\Sigma)|\Sigma^{-1}|$ subject to the constraint $\sigma_0^{1/2d} I_d \preccurlyeq \Sigma \preccurlyeq \sigma_d^{1/2d} I_d$. Hence we have the following Lipschitz property for the determinant on this domain

$$\left|\det\left(\widehat{\Sigma}_T^{1/2d}\right) - \det\left(\Sigma_f^{1/2d}\right)\right| \leq \sigma_d^{1/2-1/2d}\sqrt{d}\left|\widehat{\Sigma}_T^{1/2d} - \Sigma_f^{1/2d}\right|. \tag{4.173}$$

By Ando–van Hemmen's inequality given in Lemma 4.5.17, we have under Assumption 5, that

$$\left|\widehat{\Sigma}_T^{1/2d} - \Sigma_f^{1/2d}\right| \leq \left(\frac{1}{\sigma_0}\right)^{1-1/2d} \left|\widehat{\Sigma}_T - \Sigma_f\right|. \tag{4.174}$$

By Theorem 4.4.1 it follows that $\left|\widehat{\Sigma}_T - \Sigma_f\right| = R_T$ with

$$R_T = \mathcal{O}_{a.s.}\left(d\left(\frac{\ell_{T,d}}{T}\right)^{1/2}\right) + \mathcal{O}_{a.s.}\left(\bar{\psi}_N d\psi_d \Psi_T \log(T)\ell_{T,d}^{-1/2}\right) + \mathcal{O}_{a.s.}\left(\frac{\bar{\psi}_N^2 d\psi_d^2 \Psi_T^2 \log(T)}{T}\right). \tag{4.175}$$

Combining (4.173) and (4.174), and since $\det\left(\Sigma_f^{1/2d}\right) \geq \sigma_0^{1/2}$ we see that

$$\left|\frac{\det\left(\widehat{\Sigma}_T^{1/2d}\right)}{\det\left(\Sigma_f^{1/2d}\right)} - 1\right| \leq \left(\frac{\sigma_d}{\sigma_0}\right)^{1/2-1/2d} \frac{\sqrt{d}}{\sigma_0} R_T \leq \left(\frac{\sigma_d}{\sigma_0}\right)^{1/2} \frac{\sqrt{d}}{\sigma_0} R_T. \tag{4.176}$$

Under Assumption 5, we have by the Gershgorin circle theorem that σ_d is bounded by

$$\sigma_d \leq \sup_{i\in\{1,\ldots,d\}} \Sigma_{f_{ii}} + \sup_{i\in\{1,\ldots,d\}} \sum_{j\neq i} \left|\Sigma_{f_{ij}}\right| \lesssim d^2. \tag{4.177}$$

Let $c_{\alpha,d} := q_{\alpha,d}^{d/2} \frac{2\pi^{d/2}}{d\Gamma(d/2)}$, then from (4.171), (4.176), and (4.177) we see that

$$\begin{aligned}
\sqrt{T}\frac{\mathrm{Vol}(C_T(\alpha))^{1/d}}{c_{\alpha,d}^{1/d}\det(\Sigma_f)^{1/2d}} &= \frac{\det\left(\widehat{\Sigma}_T^{1/2d}\right)}{\det\left(\Sigma_f^{1/2d}\right)} \\
&= 1 + \left(\frac{\sigma_d}{\sigma_0}\right)^{1/2} \frac{\sqrt{d}}{\sigma_0} \mathcal{O}_{a.s.}(R_T) \\
&= 1 + \mathcal{O}_{a.s.}(d^{3/2} R_T).
\end{aligned}$$

Let

$$\tilde{T}^*(\varepsilon, d, N) := \left(\bar{\psi}_N d^3 \psi_d\right)^{\frac{2p_0}{(p_0-2)}(1+\bar{\delta}_1)} \left(\frac{1}{\varepsilon}\right)^{\frac{4p_0}{(p_0-2)}(1+\bar{\delta}_2)} \wedge e^{\frac{8p_0}{p_0-2}}$$

,

then for $T \geq \tilde{T}^*(\varepsilon, d, N)$ we have that

$$\begin{aligned} d^{3/2} R_T &= \mathcal{O}_{a.s.}\left(\bar{\psi}_N^{-\bar{\delta}_1/2} \psi_d^{-\bar{\delta}_1/2} d^{1-3/2\bar{\delta}_1} \varepsilon^{1+\bar{\delta}_2} \log^2(\tilde{T}^*(\varepsilon, d, N))\right) \\ &= \mathcal{O}_{a.s.}\left(\bar{\psi}_N^{-\bar{\delta}_1/2} d^{-1/2} \varepsilon^{1+\bar{\delta}_2} \log^2(\tilde{T}^*(\varepsilon, d, N))\right) \\ &= \mathcal{O}_{a.s.}\left(\bar{\psi}_N^{-\bar{\delta}_1/2} d^{-1/2} \varepsilon^{1+\bar{\delta}_2} \left(\log^2(\bar{\psi}_N d^3 \psi_d) + \log^2(1/\varepsilon)\right)\right) \\ &= o_{a.s.}\left(\log^2(\bar{\psi}_N d^3 \psi_d) \bar{\psi}_N^{-\bar{\delta}_1/2} d^{-1/2} \varepsilon\right), \end{aligned} \tag{4.178}$$

where the second equality follows since $\bar{\delta}_1 > 3/(3+a)$ and subsequent inequalities follow by definition of $\tilde{T}^*(\varepsilon, d, N)$, the fact that $\log^a(x)/x^b$ is decreasing on $x \geq e^{a/b}$ for any $a, b > 0$, and some basic computations. Therefore we see that

$$\sqrt{T} \frac{\mathrm{Vol}(C_T(\alpha))^{1/d}}{c_{\alpha,d}^{1/d} \det(\Sigma_f)^{1/2d}} = 1 + o_{a.s.}\left(\log^2(\bar{\psi}_N d^3 \psi_d) \bar{\psi}_N^{-\bar{\delta}_1/2} d^{-1/2} \varepsilon\right) \tag{4.179}$$

and consequently

$$\sqrt{T} \frac{\mathrm{Vol}(C_T(\alpha))^{1/d}}{c_{\alpha,d}^{1/d} \det(\Sigma_f)^{1/2d}} \xrightarrow{a.s.} 1 \text{ as } \varepsilon \downarrow 0. \tag{4.180}$$

The remainder of the first part of the proof now follows by the argument of Glynn and Whitt [73; Theorem 1]. Let $V(T) = \mathrm{Vol}(C_T(\alpha))^{1/d} + a(T)$, then by definition of $T_1(\varepsilon)$ we have that $V(T_1(\varepsilon) - 1) > \varepsilon$ and that there exists a random variable $Z(\varepsilon) \in [0, 1]$ such that $V(T_1(\varepsilon) + Z(\varepsilon)) \leq \varepsilon$. This gives us

$$\limsup_{\varepsilon \downarrow 0} \varepsilon T_1^{1/2}(\varepsilon) \leq \limsup_{\varepsilon \downarrow 0} V(T_1(\varepsilon) - 1) T_1^{1/2}(\varepsilon).$$

Since $T_1(\varepsilon) \to \infty$ almost surely as ε tends to zero it follows from (4.179) that

$$\begin{aligned} \limsup_{\varepsilon \downarrow 0} \frac{\varepsilon T_1^{1/2}(\varepsilon)}{\left(c_{\alpha,d} \det\left(\Sigma_f^{1/2}\right)\right)^{1/d}} &\leq \limsup_{\varepsilon \downarrow 0} \frac{V(T_1(\varepsilon) - 1) T_1^{1/2}(\varepsilon)}{\left(c_{\alpha,d} \det\left(\Sigma_f^{1/2}\right)\right)^{1/d}} \\ &= 1 + \ o_{a.s.}\left(\log^2(\bar{\psi}_N d^3 \psi_d) \bar{\psi}_N^{-\bar{\delta}_1/2} d^{-1/2} \varepsilon\right). \end{aligned}$$

By a similar argument, we also have that

$$\begin{aligned}\liminf_{\varepsilon\downarrow 0} \frac{\varepsilon T_1^{1/2}(\varepsilon)}{(c_{\alpha,d}^2 \det(\Sigma_f))^{1/2d}} &\geq \liminf_{\varepsilon\downarrow 0} \frac{V(T_1(\varepsilon)+Z(\varepsilon))T_1^{1/2}(\varepsilon)}{(c_{\alpha,d}^2 \det(\Sigma_f))^{1/2d}} \\ &= 1 + \; o_{a.s.}\left(\log^2(\bar{\psi}_N d^3 \psi_d)\bar{\psi}_N^{-\bar{\delta}_1/2} d^{-1/2}\varepsilon\right).\end{aligned}$$

This proves the first part of the Theorem. To prove the second claim, we first bound the difference between the confidence ellipsoids based on the estimated and asymptotic covariance matrix. Recall that $\hat{\pi}_T(f) = T^{-1}\sum_{t=1}^T f(X_t)$ and let

$$E_T := \left|T(\hat{\pi}_T(f)-\pi(f))^T\Sigma_f^{-1}(\hat{\pi}_T(f)-\pi(f)) - T(\hat{\pi}_T(f)-\pi(f))^T\widehat{\Sigma}_T^{-1}(\hat{\pi}_T(f)-\pi(f))\right|$$

Then by Cauchy–Schwarz's inequality we have that

$$E_T \leq T\left|\widehat{\Sigma}_T^{-1}-\Sigma_f^{-1}\right||\hat{\pi}_T(f)-\pi(f)|^2. \tag{4.181}$$

Since $\Sigma_f^{-1}(\widehat{\Sigma}_T-\Sigma_f)\widehat{\Sigma}_T^{-1} = \Sigma_f^{-1}-\widehat{\Sigma}_T^{-1}$, we have by sub-multiplicativity of the Frobenius norm that

$$T\left|\widehat{\Sigma}_T^{-1}-\Sigma_f^{-1}\right||\hat{\pi}_T(f)-\pi(f)|^2 \leq T\left|\widehat{\Sigma}_T-\Sigma_f\right|\left|\widehat{\Sigma}_T^{-1}\right|\left|\Sigma_f^{-1}\right||\hat{\pi}_T(f)-\pi(f)|^2.$$

From our assumed Gaussian approximation result and the law of iterated logarithm, we see that

$$T|\hat{\pi}_T(f)-\pi(f)|^2 \leq 2T\left|\hat{\pi}_T(f)-\pi(f)-T^{-1}\Sigma_f^{1/2}W_T\right|^2 + 2T^{-1}\left|\Sigma_f^{1/2}W_T\right|^2.$$

By a coordinate-wise application of the law of iterated logarithm, we have that

$$\left|\Sigma_f^{1/2}W_T\right|^2 \leq \operatorname{tr}(\Sigma_f)|W_T|^2 = \mathcal{O}_{a.s}(\operatorname{tr}(\Sigma_f)dT\log\log T). \tag{4.182}$$

This gives us

$$T|\hat{\pi}_T(f)-\pi(f)|^2 = \mathcal{O}_{a.s.}\left(\frac{\bar{\psi}_N^2\psi_d^2\Psi_T^2}{T}\right) + \mathcal{O}_{a.s.}\left(d\operatorname{tr}(\Sigma_f)\log\log T\right)$$

Let $\hat{\sigma}_1$ and σ_1 denote the smallest eigenvalues of $\widehat{\Sigma}_T$ and Σ_f respectively. Then by the equivalence of the Frobenius and spectral norm that

$$\left|\hat{\Sigma}_T^{-1}\right| \leq \sqrt{d}\left|\hat{\Sigma}_T^{-1}\right|_* \leq \frac{\sqrt{d}}{\hat{\sigma}_1}.$$

Therefore we have that

$$\begin{aligned}\left|\hat{\Sigma}_T^{-1}\right| &\leq \frac{\sqrt{d}}{\sigma_1} + \sqrt{d}\left|\frac{1}{\hat{\sigma}_1} - \frac{1}{\sigma_1}\right| \\ &\leq \frac{\sqrt{d}}{\sigma_1} + \frac{\sqrt{d}}{\sigma_0^2}|\hat{\sigma}_1 - \sigma_1|,\end{aligned}$$

where the last inequality follows since by Assumption 5 we have that $\sigma_1, \hat{\sigma}_1 \geq \sigma_0 > 0$. By Weyl's Perturbation Theorem, see Bhatia [17; Corollary III.2.6], we have that for all Hermitian matrices Σ_1, Σ_2 that

$$\max_i |\sigma_i(\Sigma_1) - \sigma_i(\Sigma_2)| \leq |\Sigma_1 - \Sigma_2|_*,$$

where $\sigma_i(\Sigma)$ denotes the i-th eigenvalue of Σ for $i = 1, \cdots, d$. Hence the eigenvalues of $\hat{\Sigma}_T$ converge at the same rate as the matrix itself, see also Vats et al. [151; Theorem 3]. This gives us

$$|\hat{\sigma}_1 - \sigma_1| \leq \left|\widehat{\Sigma}_T - \Sigma_f\right|_* \leq R_T.$$

Furthermore, we also have that $\left|\Sigma_f^{-1}\right| \leq \sqrt{d}/\sigma_1$. Therefore it follows that

$$\begin{aligned}\left|\widehat{\Sigma}_T^{-1} - \Sigma_f^{-1}\right||\hat{\pi}_T(f) - \pi(f)|^2 &\leq \left|\widehat{\Sigma}_T - \Sigma_f\right|\mathcal{O}_{a.s.}\left(d^2\frac{\mathrm{tr}(\Sigma_f)}{\sigma_0^2}\log\log T\right) \\ &\leq R_T\mathcal{O}_{a.s.}\left(d^2\frac{\mathrm{tr}(\Sigma_f)}{\sigma_0^2}\log^{1/2} T\right). \qquad (4.183)\end{aligned}$$

Given the specification of R_T given in (4.175) we have that

$$E_T \leq E_{1T} + E_{2T} + E_{3T},$$

where

$$E_{1T} = \mathcal{O}_{a.s.}\left(d^3 \frac{\operatorname{tr}(\Sigma_f)}{\sigma_0^2}\left(\frac{\ell_{T,d}}{T}\right)^{1/2} \log^{1/2}(T)\right) \tag{4.184}$$

$$E_{2T} = \mathcal{O}_{a.s.}\left(\frac{\bar{\psi}_N d^3 \operatorname{tr}(\Sigma_f)\psi_d \Psi_T \log^{3/2}(T)}{\sigma_0^2 \ell_{T,d}^{1/2}}\right) \tag{4.185}$$

$$E_{3T} = \mathcal{O}_{a.s.}\left(\frac{\bar{\psi}_N^2 d^3 \operatorname{tr}(\Sigma_f)\psi_d^2 \Psi_T^2 \log^{3/2}(T)}{\sigma_0^2 T}\right) \tag{4.186}$$

Note that for $T \geq T^*(\varepsilon, d, N) = \left(\bar{\psi}_N \frac{\operatorname{tr}(\Sigma_f)}{\sigma_0^2} d^3 \psi_d\right)^{\frac{2p_0}{(p_0-2)}(1+\bar{\delta}_1)} \left(\frac{1}{\varepsilon}\right)^{\frac{4p_0}{(p_0-2)}(1+\bar{\delta}_2)} \vee e^{\frac{10p_0}{p_0-2}}$ we have that $E_{3T} = o_{a.s.}(\bar{\psi}_N^{-2\bar{\delta}_1} \psi_d^{-2\bar{\delta}_1} d^{-3-6\bar{\delta}_1} \varepsilon^{4(1+\bar{\delta}_2)} \log^{3/2}(T^*(\varepsilon, d, N)))$. Note that by the choice of ℓ_T it follows that E_{1T} and E_{2T} are almost surely of the same asymptotic magnitude up to a log factor. By the same argument as given in (4.178) it follows that both E_{1T} and E_{2T} are of order $o_{a.s.}(\log^{5/2}(\bar{\psi}_N d^3 \psi_d)\bar{\psi}_N^{-\bar{\delta}_1/2} d^{3/2-(3+a)\bar{\delta}_1/2}\varepsilon)$ and consequently, it follows that

$$E_T = o_{a.s.}(\log^{5/2}(\bar{\psi}_N d^3 \psi_d)\bar{\psi}_N^{-\bar{\delta}_1/2} d^{3/2-(3+a)\bar{\delta}_1/2}\varepsilon). \tag{4.187}$$

Now we show that the confidence ellipsoid based on Σ_f has the asymptotically correct coverage. Note that

$$\begin{aligned}
&\left|T(\hat{\pi}_T(f) - \pi(f))^T \Sigma_f^{-1}(\hat{\pi}_T(f) - \pi(f)) - T^{-1}\langle W_T, W_T\rangle\right| \\
&= \left|T^{-1/2}\left(\sum_{t=1}^T f(X_t) - T\pi(f) - \Sigma_f^{1/2} W_T\right)^T \Sigma_f^{-1} T^{-1/2}\left(\sum_{t=1}^T f(X_t) - T\pi(f) + \Sigma_f^{1/2} W_T\right)\right| \\
&\leq \frac{\sqrt{d}}{\sigma_0 T}\left|\sum_{t=1}^T f(X_t) - T\pi(f) - \Sigma_f^{1/2} W_T\right|^2 + \frac{2\left|\Sigma_f^{-1/2}\right|}{T}\left|\sum_{t=1}^T f(X_t) - T\pi(f) - \Sigma_f^{1/2} W_T\right| |W_T|,
\end{aligned}$$

where the last inequality follows from Cauchy–Schwarz and since we have by the equivalence of the Frobenius and spectral norm that $\left|\Sigma_f^{-1}\right| \leq \sqrt{d}/\sigma_0$. Moreover, by the assumed weak Gaussian approximation and (4.182) we

obtain

$$\left|T(\hat{\pi}_T(f)-\pi(f))^T\Sigma_f^{-1}(\hat{\pi}_T(f)-\pi(f))-\frac{\langle W_T,W_T\rangle}{T}\right| \tag{4.188}$$

$$= \mathcal{O}_{a.s.}\left(\frac{\bar{\psi_N}^2\sqrt{d}\psi_d^2\Psi_T^2}{T}\right)+\mathcal{O}_p\left(\frac{\bar{\psi}_N d^{3/2}\psi_d\Psi_T'}{\sqrt{T}}\right)$$

$$= \mathcal{O}_{a.s.}\left(\frac{\bar{\psi}_N d^{3/2}\psi_d\Psi_T\log^{1/4}(T)}{\sqrt{T}}\right)$$

$$= o_{a.s.}\left(\frac{\log\left(\bar{\psi}_N d^{3/2}\psi_d\right)}{(\bar{\psi}_N d^{3/2}\psi_d)^{\bar{\delta}_3}}\varepsilon\right), \tag{4.189}$$

with $\Psi_T' = \Psi_T(\log\log T)^{1/2}$ and for all $T \geq \left(\frac{\bar{\psi}_N d^{3/2}\psi_d}{\varepsilon}\right)^{\frac{2p_0}{(p_0-2)}(1+\bar{\delta}_3)} \vee e^{\frac{5p_0}{2(p_0-2)}}$ for any $\bar{\delta}_3 > 0$. Finally, given (4.187) and (4.188), we can use the argument of Glynn and Whitt [73; Theorem 1] to show that at termination time $T_1(\varepsilon)$ the empirical confidence interval also has the correct coverage as $\varepsilon \downarrow 0$. □

Proof of Theorem 4.4.5

Proof. The proof follows completely analogous to the proof of Theorem 4.4.4. Note that we now obtain

$$\left|T(\hat{\pi}_T(f)-\pi(f))^T\hat{\Sigma}_f^{-1}(\hat{\pi}_T(f)-\pi(f))-T^{-1}\langle W_T,W_T\rangle\right|$$

$$= \mathcal{O}_{a.s.}\left(\frac{\bar{\psi}_N d^3\,\mathrm{tr}(\Sigma_f)\psi_d\log^2(T)}{\sigma_1 T^{\frac{p_0-2}{8(p_0-1)}}}\right).$$

□

Chapter 5

Conclusion

The main contribution of this thesis are the obtained novel Gaussian approximation results and their application to uncertainty quantification and the analysis of termination criteria. Firstly, we obtain strong invariance principles for a broad class of ergodic Markov processes. For the Zig-Zag sampler, which is a specific PDMP, we show that the optimal Gaussian approximation rate can be obtained. This is the first MCMC sampler, within the considered class of problems, for which this optimal rate has been obtained. We demonstrate how these results can be used to analyse the batch means method for simulation output of PDMC samplers. Consequently, many results regarding uncertainty quantification now carry over to PDMC samplers. Since previous work on estimation of the MCMC standard error is based on strong invariance principles with limited accuracy, our results improve the currently available regularity conditions guaranteeing strong convergence of the batch means estimator in any MCMC setting. Furthermore, we also discuss multiple convergence diagnostics suitable for PDMP based sampling algorithms.

The largest contribution of this thesis are the novel quantitative bounds on (functional) Markov central limit theorems in both multivariate and high-dimensional settings, given in Chapter 4. Our results provide the first Gaussian approximation result for multivariate MCMC samplers, for which the optimal rate is attained. Furthermore, our results cover a larger class of polynomially ergodic Markov chains than previous works and are the first to quantify the influence of polynomial ergodicity on the Gaussian approximation rate.

Another key contribution of our paper is obtaining the dependence of the approximation errors on the dimension of both the target distribution and the feature space. For applications in Bayesian statistics, our results provide a direct link between the statistical model complexity and the computational complexity of the MCMC algorithm. Our results provide us with explicit simulation requirements such that a Markov CLT holds. The results can also be used to adapt the tuning parameters of widely used variance estimation methods for MCMC simulation output. Moreover, in high-dimensional settings they imply simulation requirements that guarantee the validity of these variance estimation methods. Therefore, we are able to give conditions for valid uncertainty quantification for high-dimensional MCMC algorithms.

Another key contribution of our work is the application of these results to MCMC output analysis. Our Gaussian approximations enable us to extend results on the estimation of variance and effective sample size to a broader class of multivariate and high-dimensional settings. Furthermore, we provide quantitative convergence bounds for termination criteria and show that the termination time of a wide class of MCMC algorithms scales polynomially with dimension while ensuring a desired level of precision. Our results offer guidance to practitioners in obtaining appropriate standard errors, as well as determining the minimum simulation effort and termination time for MCMC algorithms in high-dimensional applications.

Bibliography

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

[2] R. Altmeyer. Polynomial time guarantees for sampling based posterior inference in high-dimensional generalised linear models. *arXiv preprint arXiv:2208.13296*, 2022.

[3] C. Andrieu, G. Fort, and M. Vihola. Quantitative convergence rates for subgeometric Markov chains. *Journal of Applied Probability*, 52(2): 391–404, 2015.

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

[5] S. Asmussen. *Applied probability and queues*, volume 2. Springer, 2003.

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

[7] A. S. Bandeira, A. Maillard, R. Nickl, and S. Wang. On free energy barriers in gaussian priors and failure of cold start mcmc for high-dimensional unimodal distributions. *Philosophical Transactions of the Royal Society A*, 381(2247):20220150, 2023.

[8] A. Banerjee and D. Vats. Multivariate strong invariance principles in markov chain monte carlo. *Electronic Journal of Statistics*, 18(1): 2450–2476, 2024.

[9] P. H. Baxendale. Renewal theory and computable convergence rates for geometrically ergodic Markov chains. *Annals of Applied Probability*, pages 700–738, 2005.

[10] W. Bednorz and K. Łatuszyński. A few remarks on "fixed-width output analysis for Markov chain Monte Carlo" by Jones et al. *Journal of the American Statistical Association*, 102(480):1485–1486, 2007.

[11] A. Belloni and V. Chernozhukov. On the computational complexity of MCMC-based estimators in large samples. *The Annals of Statistics*, 37 (4), 2009.

[12] I. Berkes, S. Hörmann, J. Schauer, et al. Split invariance principles for stationary processes. *The Annals of Probability*, 39(6):2441–2473, 2011.

[13] I. Berkes, W. Liu, and W. B. Wu. Komlós–Major–Tusnády approximation under dependence. *The Annals of Probability*, 42(2):794–817, 2014.

[14] P. Bertail and G. Ciołek. New Bernstein and Hoeffding type inequalities for regenerative Markov chains. 2018.

[15] A. Bertazzi, J. Bierkens, and P. Dobson. Approximations of piecewise deterministic markov processes and their convergence properties. *Stochastic Processes and their Applications*, 154:91–153, 2022.

[16] A. Bertazzi, P. Dobson, and P. Monmarché. Piecewise deterministic sampling with splitting schemes. *arXiv preprint arXiv:2301.02537*, 2023.

[17] R. Bhatia. *Matrix analysis*, volume 169. Springer Science & Business Media, 2013.

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

[19] J. Bierkens and A. Duncan. Limit theorems for the zig-zag process. *Advances in Applied Probability*, 49(3):791–825, 2017.

[20] J. Bierkens and G. Roberts. A piecewise deterministic scaling limit of lifted Metropolis–Hastings in the Curie–Weiss model. *The Annals of Applied Probability*, 27(2):846–882, 2017.

[21] J. Bierkens, P. Fearnhead, G. Roberts, et al. The zig-zag process and super-efficient sampling for Bayesian analysis of big data. *The Annals of Statistics*, 47(3):1288–1320, 2019.

[22] J. Bierkens, G. O. Roberts, and P.-A. Zitt. Ergodicity of the zigzag process. *Ann. Appl. Probab.*, 29(4):2266–2301, 08 2019. doi: 10.1214/18-AAP1453. URL https://doi.org/10.1214/18-AAP1453.

[23] J. Bierkens, S. Grazzi, K. Kamatani, and G. Roberts. The boomerang sampler. In *International conference on machine learning*, pages 908–918. PMLR, 2020.

[24] T. Bonis. Stein's method for normal approximation in Wasserstein distances with application to the multivariate central limit theorem. *Probability Theory and Related Fields*, 178(3):827–860, 2020.

[25] N. Bou-Rabee and A. Eberle. Mixing time guarantees for unadjusted Hamiltonian Monte Carlo. *Bernoulli*, 29(1):75–104, 2023.

[26] A. Bouchard-Côté, S. J. Vollmer, and A. Doucet. The bouncy particle sampler: A nonreversible rejection-free Markov chain Monte Carlo method. *Journal of the American Statistical Association*, 113(522): 855–867, 2018.

[27] A. Bovier, M. Eckhoff, V. Gayrard, and M. Klein. Metastability in reversible diffusion processes. i. sharp asymptotics for capacities and exit times. *J. Eur. Math. Soc.(JEMS)*, 6(4):399–424, 2004.

[28] R. C. Bradley. Basic properties of strong mixing conditions. a survey and some open questions. *Probability Surveys*, 2:107–144, 2005.

[29] A. E. Brockwell and J. B. Kadane. Identification of regeneration times in mcmc simulation, with application to adaptive schemes. *Journal of Computational and Graphical Statistics*, 14(2):436–458, 2005.

[30] S. Brooks, A. Gelman, G. Jones, and X.-L. Meng. *Handbook of Markov chain Monte Carlo*. CRC press, 2011.

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

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

[33] J. Chang, X. Chen, and M. Wu. Central limit theorems for high dimensional dependent data. *Bernoulli*, 30(1):712–742, 2024.

[34] C. Chien, D. Goldsman, and B. Melamed. Large-sample results for batch means. *Management Science*, 43(9):1288–1295, 1997.

[35] C.-H. Chien. Small-sample theory for steady state confidence intervals. In *Proceedings of the 20th conference on Winter simulation*, pages 408–413, 1988.

[36] M. K. Cowles and B. P. Carlin. Markov chain Monte Carlo convergence diagnostics: a comparative review. *Journal of the American Statistical Association*, 91(434):883–904, 1996.

[37] E. Csáki and M. Csörgő. On additive functionals of Markov chains. *Journal of Theoretical Probability*, 8(4):905–919, 1995.

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

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

[40] M. Csörgö and P. Révész. *Strong approximations in probability and statistics*. Academic press, 2014.

[41] S. Csörgö and P. Hall. The Komlós–Major–Tusnády approximations and their applications. *Australian Journal of Statistics*, 26(2):189–218, 1984.

[42] C. Cuny, J. Dedecker, and F. Merlevède. On the komlós, major and tusnády strong approximation for some classes of random iterates. *Stochastic Processes and their Applications*, 128(4):1347–1385, 2018.

[43] A. S. Dalalyan. Theoretical guarantees for approximate sampling from smooth and log-concave densities. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 79(3):651–676, 2017.

[44] H. Damerdji. Strong consistency and other properties of the spectral variance estimator. *Management Science*, 37(11):1424–1440, 1991.

[45] H. Damerdji. Strong consistency of the variance estimator in steady-state simulation output analysis. *Mathematics of Operations Research*, 19(2):494–512, 1994.

[46] H. Damerdji. Mean-square consistency of the variance estimator in steady-state simulation output analysis. *Operations Research*, 43(2): 282–291, 1995.

[47] Y. A. Davydov. Convergence of distributions generated by stationary stochastic processes. *Theory of Probability & Its Applications*, 13(4): 691–696, 1968.

[48] V. De la Pena and E. Giné. *Decoupling: from dependence to independence*. Springer Science & Business Media, 2012.

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

[50] G. Deligiannidis, A. Bouchard-Côté, A. Doucet, et al. Exponential ergodicity of the bouncy particle sampler. *The Annals of Statistics*, 47 (3):1268–1287, 2019.

[51] R. Douc, G. Fort, E. Moulines, and P. Soulier. Practical drift conditions for subgeometric rates of convergence. 2004.

[52] R. Douc, G. Fort, and A. Guillin. Subgeometric rates of convergence of f-ergodic strong Markov processes. *Stochastic processes and their applications*, 119(3):897–923, 2009.

[53] R. Douc, E. Moulines, P. Priouret, and P. Soulier. *Markov chains*. Springer, 2018.

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

[55] S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth. Hybrid monte carlo. *Physics letters B*, 195(2):216–222, 1987.

[56] A. B. Duncan, T. Lelievre, and G. Pavliotis. Variance reduction using nonreversible Langevin samplers. *Journal of statistical physics*, 163(3): 457–491, 2016.

[57] A. Durmus and É. Moulines. Nonasymptotic convergence analysis for the unadjusted Langevin algorithm. *Annals of Applied Probability*, 27 (3):1551–1587, 2017.

[58] A. Durmus and E. Moulines. High-dimensional Bayesian inference via the unadjusted Langevin algorithm. *Bernoulli*, 25(4A):2854–2882, 2019.

[59] A. Durmus, A. Guillin, P. Monmarché, et al. Geometric ergodicity of the bouncy particle sampler. *Annals of Applied Probability*, 30(5): 2069–2098, 2020.

[60] U. Einmahl. Extensions of results of komlós, major, and tusnády to the multivariate case. *Journal of multivariate analysis*, 28(1):20–68, 1989.

[61] U. Einmahl and D. M. Mason. Gaussian approximation of local empirical processes indexed by functions. *Probability Theory and Related Fields*, 107:283–311, 1997.

[62] C. El-Nouty. On the large increments of fractional brownian motion. *Statistics & probability letters*, 41(2):169–178, 1999.

[63] R. Eldan, D. Mikulincer, and A. Zhai. The CLT in high dimensions. *The Annals of Probability*, 48(5):2494–2524, 2020.

[64] M. Fathi. Stein kernels and moment maps. *The Annals of Probability*, 47(4):2172–2185, 2019.

[65] P. Fearnhead, J. Bierkens, M. Pollock, G. O. Roberts, et al. Piecewise deterministic Markov processes for continuous-time monte carlo. *Statistical Science*, 33(3):386–412, 2018.

[66] J. M. Flegal and G. L. Jones. Batch means and spectral variance estimators in Markov chain Monte Carlo. *The Annals of Statistics*, 38 (2):1034–1070, 2010.

[67] G. Fort and G. O. Roberts. Subgeometric ergodicity of strong Markov processes. *The Annals of Applied Probability*, 15(2):1565–1589, 2005.

[68] J. Gawlikowski, C. R. N. Tassi, M. Ali, J. Lee, M. Humt, J. Feng, A. Kruspe, R. Triebel, P. Jung, R. Roscher, et al. A survey of uncertainty in deep neural networks. *Artificial Intelligence Review*, 56(Suppl 1):1513–1589, 2023.

[69] A. E. Gelfand and A. F. Smith. Sampling-based approaches to calculating marginal densities. *Journal of the American statistical association*, 85(410):398–409, 1990.

[70] W. R. Gilks, G. O. Roberts, and S. K. Sahu. Adaptive markov chain monte carlo through regeneration. *Journal of the American statistical association*, 93(443):1045–1054, 1998.

[71] P. W. Glynn and W. Whitt. Ordinary clt and wlln versions of l= λ w. *Mathematics of operations research*, 13(4):674–692, 1988.

[72] P. W. Glynn and W. Whitt. Estimating the asymptotic variance with batch means. *Operations Research Letters*, 10(8):431–435, 1991.

[73] P. W. Glynn and W. Whitt. The asymptotic validity of sequential stopping rules for stochastic simulations. *The Annals of Applied Probability*, 2(1):180–198, 1992.

[74] D. Goldsman, M. Meketon, and L. Schruben. Properties of standardized time series weighted area variance estimators. *Management Science*, 36(5):602–612, 1990.

[75] L. Gong and J. M. Flegal. A practical sequential stopping rule for high-dimensional Markov chain Monte Carlo. *Journal of Computational and Graphical Statistics*, 25(3):684–700, 2016.

[76] S. Gouëzel. Almost sure invariance principle for dynamical systems by spectral methods. *The Annals of Probability*, 38(4):1639–1671, 2010.

[77] M. Hairer, A. M. Stuart, and S. J. Vollmer. Spectral gaps for a Metropolis–Hastings algorithm in infinite dimensions. *The Annals of Applied Probability*, 24(6), 2014.

[78] W. K. Hastings. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 1970.

[79] A. J. Heunis. Strong invariance principle for singular diffusions. *Stochastic processes and their applications*, 104(1):57–80, 2003.

[80] A. Hinrichs, E. Novak, M. Ullrich, and H. Woźniakowski. The curse of dimensionality for numerical integration of smooth functions. *Mathematics of Computation*, 83(290):2853–2863, 2014.

[81] J. P. Hobert, G. L. Jones, B. Presnell, and J. S. Rosenthal. On the applicability of regenerative simulation in Markov chain Monte Carlo. *Biometrika*, 89(4):731–743, 2002.

[82] R. Höpfner and E. Löcherbach. *Limit theorems for null recurrent Markov processes*. Number 768. American Mathematical Soc., 2003.

[83] C.-R. Hwang, S.-Y. Hwang-Ma, and S.-J. Sheu. Accelerating Gaussian diffusions. *The Annals of Applied Probability*, pages 897–913, 1993.

[84] S. F. Jarner and G. O. Roberts. Polynomial convergence rates of Markov chains. *The Annals of Applied Probability*, 12(1):224–247, 2002.

[85] G. L. Jones. On the Markov chain central limit theorem. *Probability Surveys*, 1:299–320, 2004.

[86] G. L. Jones and J. P. Hobert. Honest exploration of intractable probability distributions via Markov chain Monte Carlo. *Statistical Science*, pages 312–334, 2001.

[87] G. L. Jones, M. Haran, B. S. Caffo, and R. Neath. Fixed-width output analysis for Markov chain Monte Carlo. *Journal of the American Statistical Association*, 101(476):1537–1547, 2006.

[88] O. Kallenberg. *Foundations of modern probability*, volume 2. Springer, 1997.

[89] J. Komlós, P. Major, and G. Tusnády. An approximation of partial sums of independent RV's, and the sample DF. II. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 34(1):33–58, 1976.

[90] M. Komlos. Tusnady (1975) An approximation of partial sums of RV's, and the sample DF, I, Z. *Wahrscheinlichkeitstheorie und Verw. Gebiete*, 32:111–131.

[91] J. Kuelbs and W. Philipp. Almost sure invariance principles for partial sums of mixing b-valued random variables. *The Annals of Probability*, pages 1003–1036, 1980.

[92] P. Lazić and N. Sandrić. On sub-geometric ergodicity of diffusion processes. *Bernoulli*, 2021.

[93] T. Lelievre, F. Nier, and G. A. Pavliotis. Optimal non-reversible linear drift for the convergence to equilibrium of a diffusion. *Journal of Statistical Physics*, 152(2):237–274, 2013.

[94] P. W. Lewis and G. S. Shedler. Simulation of nonhomogeneous poisson processes by thinning. *Naval research logistics quarterly*, 26(3):403–413, 1979.

[95] H. Li and Q. Qin. Multivariate strong invariance principle and uncertainty assessment for time in-homogeneous cyclic MCMC samplers. *arXiv preprint arXiv:2405.10194*, 2024.

[96] W. V. Li. Limit theorems for the square integral of brownian motion and its increments. *Stochastic processes and their applications*, 41(2): 223–239, 1992.

[97] Y. Liu, D. Vats, and J. M. Flegal. Batch size selection for variance estimators in MCMC. *Methodology and Computing in Applied Probability*, pages 1–29, 2021.

[98] E. Löcherbach and D. Loukianova. 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(8):1301–1321, 2008.

[99] E. Löcherbach and D. Loukianova. The law of iterated logarithm for additive functionals and martingale additive functionals of Harris recurrent Markov processes. *Stochastic processes and their applications*, 119(7):2312–2335, 2009.

[100] J. Lu, W. B. Wu, Z. Xiao, and L. Xu. Almost sure invariance principle of beta mixing time series in Hilbert space. *arXiv preprint arXiv:2209.12535*, 2022.

[101] P. Major. An improvement of Strassen's invariance principle. *The Annals of Probability*, 7(1):55–61, 1979.

[102] K. L. Mengersen, C. P. Robert, and C. Guihenneuc-Jouyaux. MCMC convergence diagnostics: a reviewww. *Bayesian statistics*, 6:415–440, 1999.

[103] F. Merlevède, E. Rio, et al. Strong approximation for additive functionals of geometrically ergodic Markov chains. *Electronic Journal of Probability*, 20, 2015.

[104] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21(6):1087–1092, 1953.

[105] S. P. Meyn and R. L. Tweedie. Stability of Markovian processes ii: Continuous-time processes and sampled chains. *Advances in Applied Probability*, pages 487–517, 1993.

[106] S. P. Meyn and R. L. Tweedie. Computable bounds for geometric convergence rates of Markov chains. *The Annals of Applied Probability*, pages 981–1011, 1994.

[107] S. P. Meyn and R. L. Tweedie. *Markov chains and stochastic stability*. Springer Science & Business Media, 2012.

[108] F. Mies and A. Steland. Sequential Gaussian approximation for nonstationary time series in high dimensions. *Bernoulli*, 29(4):3114–3140, 2023.

[109] S. Mihalache. Strong approximations and sequential change-point analysis for diffusion processes. *Statistics & Probability Letters*, 82(3): 464–472, 2012.

[110] P. Monmarché. Piecewise deterministic simulated annealing. *arXiv preprint arXiv:1410.1656*, 2014.

[111] I. Murray, R. Adams, and D. MacKay. Elliptical slice sampling. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pages 541–548. JMLR Workshop and Conference Proceedings, 2010.

[112] P. Mykland, L. Tierney, and B. Yu. Regeneration in Markov chain samplers. *Journal of the American Statistical Association*, 90(429): 233–241, 1995.

[113] R. M. Neal. Slice sampling. *The annals of statistics*, 31(3):705–767, 2003.

[114] R. Nickl and S. Wang. On polynomial-time computation of high-dimensional posterior measures by Langevin-type algorithms. *Journal of the European Mathematical Society*, 2022.

[115] E. Nummelin. A splitting technique for Harris recurrent Markov chains. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 43 (4):309–318, 1978.

[116] E. Nummelin. *General irreducible Markov chains and non-negative operators*. Number 83. Cambridge University Press, 2004.

[117] E. Nummelin and P. Tuominen. Geometric ergodicity of harris recurrent Markov chains with applications to renewal theory. *Stochastic Processes and Their Applications*, 12(2):187–202, 1982.

[118] E. Nummelin and P. Tuominen. The rate of convergence in orey's theorem for harris recurrent Markov chains with applications to renewal theory. *Stochastic Processes and Their Applications*, 15(3):295–311, 1983.

[119] S. Orey. *Limit theorems for Markov chain Monte Carlo transition probabilities*. van Nostrand London, 1971.

[120] J. Ortega. On the size of the increments of nonstationary Gaussian processes. *Stochastic processes and their applications*, 18(1):47–56, 1984.

[121] A. Pengel and J. Bierkens. Strong invariance principles for ergodic Markov processes. *Electronic Journal of Statistics*, 18(1):191–246, 2024.

[122] A. Pengel, J. Yang, and Z. Zhou. Gaussian approximation and output analysis for high-dimensional mcmc. *arXiv preprint arXiv:2407.05492*, 2024.

[123] W. Philipp and W. Stout. *Almost sure invariance principles for partial sums of weakly dependent random variables*, volume 161. American Mathematical Soc., 1975.

[124] V. I. Piterbarg and I. V. Rodionov. High excursions of bessel and related random processes. *Stochastic Processes and their Applications*, 130(8):4859–4872, 2020.

[125] Q. Qin and J. P. Hobert. Convergence complexity analysis of Albert and Chib's algorithm for Bayesian probit regression. *The Annals of Statistics*, 47(4):2320–2347, 2019.

[126] Q. Qin and J. P. Hobert. On the limitations of single-step drift and minorization in Markov chain convergence analysis. *The Annals of Applied Probability*, 31(4):1633–1659, 2021.

[127] Q. Qin and J. P. Hobert. Wasserstein-based methods for convergence complexity analysis of MCMC with applications. *The Annals of Applied Probability*, 32(1):124–166, 2022.

[128] B. Rajaratnam and D. Sparks. MCMC-based inference in the era of big data: A fundamental analysis of the convergence complexity of high-dimensional chains. *arXiv preprint arXiv:1508.00947*, 2015.

[129] P. Révész. On the increments of Wiener and related processes. *The Annals of Probability*, pages 613–622, 1982.

[130] D. Revuz. *Markov chains*. Elsevier, 2008.

[131] L. Rey-Bellet and K. Spiliopoulos. Irreversible Langevin samplers and variance reduction: a large deviations approach. *Nonlinearity*, 28(7): 2081, 2015.

[132] E. Rio. Covariance inequalities for strongly mixing processes. In *Annales de l'IHP Probabilités et statistiques*, volume 29, pages 587–597, 1993.

[133] G. O. Roberts and J. S. Rosenthal. Markov chains and de-initializing processes. *Scandinavian Journal of Statistics*, 28(3):489–504, 2001.

[134] G. O. Roberts and R. L. Tweedie. Exponential convergence of langevin distributions and their discrete approximations. *Bernoulli*, 1996.

[135] L. C. G. Rogers and D. Williams. *Diffusions, Markov processes and martingales: Volume 2, Itô calculus*, volume 2. Cambridge university press, 2000.

[136] J. S. Rosenthal. Minorization conditions and convergence rates for Markov chain Monte Carlo. *Journal of the American Statistical Association*, 90(430):558–566, 1995.

[137] V. Roy. Convergence diagnostics for Markov chain Monte Carlo. *Annual Review of Statistics and Its Application*, 7:387–412, 2020.

[138] M. Sherman and D. Goldsman. Large-sample normality of the batch-means variance estimator. *Operations Research Letters*, 30(5):319–326, 2002.

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

[140] K. Sigman. One-dependent regenerative processes and queues in continuous time. *Mathematics of Operations Research*, 15(1):175–189, 1990.

[141] K. Sigman and R. W. Wolff. A review of regenerative processes. *SIAM review*, 35(2):269–288, 1993.

[142] W. T. Song and B. W. Schmeiser. Optimal mean-squared-error batch sizes. *Management Science*, 41(1):110–123, 1995.

[143] J. T. Springenberg, A. Klein, S. Falkner, and F. Hutter. Bayesian optimization with robust Bayesian neural networks. *Advances in neural information processing systems*, 29, 2016.

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

[145] O. Stramer and R. Tweedie. Langevin-type models i: Diffusions with given stationary distributions and their discretizations. *Methodology and Computing in Applied Probability*, 1(3):283–306, 1999.

[146] D. W. Stroock and S. S. Varadhan. *Multidimensional diffusion processes*, volume 233. Springer Science & Business Media, 1997.

[147] R. Tang and Y. Yang. On the computational complexity of Metropolis-adjusted Langevin algorithms for Bayesian posterior sampling. *arXiv preprint arXiv:2206.06491*, 2022.

[148] S. Tweedie. Generalized resolvents and harris recurrence. *Doeblin and modern probability*, 149, 1993.

[149] A. W. Van der Vaart. *Asymptotic statistics*, volume 3. Cambridge university press, 2000.

[150] J. L. van Hemmen and T. Ando. An inequality for trace ideals. *Communications in Mathematical Physics*, 76:143–148, 1980.

[151] D. Vats, J. M. Flegal, and G. L. Jones. Strong consistency of multivariate spectral variance estimators in Markov chain Monte Carlo. *Bernoulli*, 24(3):1860–1909, 2018.

[152] D. Vats, J. M. Flegal, and G. L. Jones. Multivariate output analysis for Markov chain Monte Carlo. *Biometrika*, 106(2):321–337, 2019.

[153] C. Villani et al. *Optimal transport: old and new*, volume 338. Springer, 2009.

[154] M. Welling and Y. W. Teh. Bayesian learning via stochastic gradient Langevin dynamics. In *Proceedings of the 28th international conference on machine learning (ICML-11)*, pages 681–688, 2011.

[155] D. Williams. *Stochastic integrals: proceedings of the LMS Durham Symposium, July 7-17, 1980*, volume 851. Springer, 2006.

[156] J. Yang and J. S. Rosenthal. Complexity results for MCMC derived from quantitative bounds. *The Annals of Applied Probability*, 33(2): 1459–1500, 2023.

[157] A. Y. Zaitsev. Estimates of the Lévy–Prokhorov distance in the multivariate central limit theorem for random variables with finite exponential moments. *Theory of Probability & Its Applications*, 31(2):203–220, 1987.

[158] A. Y. Zaitsev. On the Gaussian approximation of convolutions under multidimensional analogues of SN Bernstein's inequality conditions. *Probability theory and related fields*, 74(4):535–566, 1987.

[159] A. Y. Zaitsev. Multidimensional version of the results of komlós, major and tusnády for vectors with finite exponential moments. *ESAIM: Probability and Statistics*, 2:41–108, 1998.

[160] A. Y. Zaitsev. The accuracy of strong Gaussian approximation for sums of independent random vectors. *Russian Mathematical Surveys*, 68(4):721, 2013.

[161] Q. Zhou, J. Yang, D. Vats, G. O. Roberts, and J. S. Rosenthal. Dimension-free mixing for high-dimensional Bayesian variable selection. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 84(5):1751–1784, 2022.

Summary

Markov Chain Monte Carlo (MCMC) methods are generally acknowledged to be the most versatile algorithms for simulating from a probability distribution of interest. MCMC methods are extensively applied in an array of fields, ranging from statistics and machine learning to physics. The main idea is to construct a relatively simple process such that, in its equilibrium state, the process can be used to approximate the distribution of interest. Any MCMC algorithm can be conceptualised as a set of rules dictating the movement of a particle through space. The implied motion of the particle ensures that, over time, the particle explores the space in accordance with the probability distribution of interest. In the long run, the particle thus spends time in any region of the space that is proportional to the probability mass assigned to that region. The simulated path of our particle can subsequently be used to estimate quantities related to the probability distribution of interest. Most quantities of interest can naturally be expressed as high-dimensional integrals with respect to the distribution of interest, and are estimated using the corresponding time-averages of the particle trajectory.

For every application of MCMC it must be assessed when it is justified to terminate the simulation. Prematurely terminating our simulation algorithm, i.e., the exploring particle, could lead to inaccurate sample statistics or even to output that does not resemble the equilibrium distribution. This problem is closely related to the uncertainty quantification of our algorithm. Since MCMC methods are only able to sample approximately from the target distribution, quantifying the uncertainty of the simulation output is of central importance. Termination rules are the gold standard for determining the appropriate running time of an MCMC algorithm. These rules allow termination of the simulation when the uncertainty of our estimation

process is below some predetermined tolerance level.

In order to quantify the uncertainty of our sampling algorithms, it is required that a so-called Gaussian approximation holds. These quantitative invariance principles tell us how the trajectory of the time-average of the sampling process will fluctuate around the value it is designed to approximate. Moreover, they justify the construction of confidence ellipsoids which are required for the implementation of various termination criteria. A confidence ellipsoid provides a range of values such that, if the sampling process were repeated infinitely often, the proportion of ellipsoids containing the quantity of interest would equal the pre-determined confidence level.

In Chapter 3, novel Gaussian approximation results for a wide range of continuous-time Markov processes are presented, particularly achieving the optimal approximation rate for the Zig-Zag sampler, a Piecewise Deterministic Markov Process (PDMP). These processes enjoy favourable scaling properties in modern big-data settings. Our obtained results improve rularity conditions for standard error estimation of PDMP-based algorithms.

In Chapter 4, we obtain novel dimension-dependent Gaussian approximation results which allow us to adapt and justify methods for uncertainty quantification and termination criteria in high-dimensional settings. Moreover, for a large class of processes, our Gaussian approximation results attain the optimal approximation rate.

Samenvatting

Markov Chain Monte Carlo (MCMC)-methoden worden erkend als de meest toepasbare algoritmen voor het simuleren van kansverdelingen. MCMC-methoden worden op grote schaal toegepast in verschillende vakgebieden, variërend van statistiek en machine learning tot natuurkunde. Het achterliggende idee is om een relatief eenvoudig proces te construeren, zodat dit proces in zijn evenwichtstoestand kan worden gebruikt om de gewenste kansverdeling te benaderen. Elk MCMC-algoritme kan worden geconceptualiseerd als een reeks instructies die de beweging van een deeltje (partikel) door de ruimte beschrijven. De motie van het deeltje is geconstrueerd zodat het deeltje na verloop van tijd de ruimte verkent volgens de gewenste waarschijnlijkheidsverdeling. Op de lange termijn brengt het deeltje tijd door in elk gebied van de ruimte die evenredig is met de kansmassa die aan dat gebied is toegewezen. Het gesimuleerde pad van het deeltje kan vervolgens worden gebruikt om grootheden te schatten die gerelateerd zijn aan de gewenste waarschijnlijkheidsverdeling. De meeste grootheden van belang kunnen op natuurlijke wijze worden uitgedrukt als hoog-dimensionale integralen met betrekking tot de gewenste kansverdeling, en kunnen worden geschat met de bijbehorende tijdsgemiddelden van het deeltjestraject.

Voor elke toepassing van MCMC moet worden beoordeeld wanneer het gerechtvaardigd is om de simulatie te beëindigen. Het voortijdig beëindigen van ons simulatie-algoritme, het verkennende deeltje, kan ons onnauwkeurige steekproefstatistieken of zelfs gesimuleerde waarden die niet overeenkomen met de evenwichtsverdeling opleveren. Dit probleem hangt nauw samen met de onzekerheidskwantificatie van ons algoritme. Aangezien MCMC-methoden per constructie alleen in staat zijn om de gewenste kansverdeling te benaderen, is het kwantificeren van de onzekerheid van de simulatie-uitkomst van essentieel belang. Terminatiecriteria zijn de gouden standaard voor het

bepalen van de juiste looptijd van een MCMC-algoritme. Deze regels maken het mogelijk om de simulatie te beëindigen wanneer de onzekerheid van ons schattingsproces onder een vooraf bepaald tolerantieniveau ligt.

Om de onzekerheid van onze simulatie-algoritmen te kwantificeren, is het vereist dat een zogenoemde Gaussische benadering geldt. Deze kwantitatieve invariantieprincipes vertellen ons hoe het traject van het tijdsgemiddelde van het simulatieproces zal fluctueren rond de waarde die het moet benaderen. Bovendien rechtvaardigen Gaussische benadering de constructie van betrouwbaarheidsellipsoïden die nodig zijn voor de implementatie van verschillende terminatiecriteria. Een betrouwbaarheidsellipsoïde bescrhijft een verzameling van waarden zodat, als het simulatieproces oneindig vaak herhaald zou worden, het aandeel ellipsoïden dat de gezochte grootheid bevat gelijk zal zijn aan het vooraf bepaalde betrouwbaarheidsniveau.

In Hoofdstuk 3 worden nieuwe Gaussische benaderingsresultaten gepresenteerd voor een breed scala aan tijdscontinue Markovprocessen, waarbij met name de optimale benaderingsfout voor de Zig-Zag-simulatie-algoritme, een *Piecewise Deterministic Markov Process* (PDMP), wordt aangetoond. Deze processen hebben gunstige schaalbaarheidseigenschappen in moderne big data-applicaties. Onze verkregen resultaten verbeteren de regulariteitsvoorwaarden voor het schatten van standaardfouten voor PDMP-gebaseerde algoritmen.

In Hoofdstuk 4 verkrijgen we nieuwe dimensie-afhankelijke Gaussische benaderingsresultaten, waarmee we methoden voor onzekerheidskwantificering en terminatiecriteria in hoog-dimensionale toepassingen kunnen rechtvaardigen. Bovendien bezitten onze aangetoonde Gaussische benaderingsresultaten voor een grote klasse van processen de optimale benaderingsfout.

Acknowledgements

Pursuing a PhD in the field of Stochastics has been a long-held aspiration of mine, and I am deeply grateful to all those who have been there with me along the way. Firstly, I would like to express my gratitude to Joris Bierkens for giving me the opportunity to pursue this PhD in Mathematical & Computational Statistics. Thank you for always believing in my ideas and giving me the freedom to explore them with your support and guidance throughout this process. Furthermore, I am deeply thankful to the entire Statistics and Applied Probability group for fostering a vibrant and supportive working environment. I would especially like to acknowledge Geurt Jongbloed, whose daily enthusiasm and kindness always lifted my spirits.

I would also like to thank the external members of the committee, Randolf Altmeyer, Krzysztof Łatuszyński, Frank van der Meulen, Frank Redig, Aad van der Vaart (reserve member), and Jun Yang (external collaborator) for their valuable feedback on my thesis and for participating in my PhD defence and ceremony. I am honoured to have you in my committee.

I am deeply grateful to my collaborators Jun Yang and Zhou Zhou. Working with both of you was a joy, and all the Zoom calls and hours spent in front of the blackboard truly paid off. The research visits to the University of Oxford and the University of Copenhagen, where we had the chance to work together in person, were highlights of my PhD.

I would like to express my heartfelt gratitude to the many inspiring professors I had the privilege of learning from throughout my education. In particular, I am deeply thankful to Rein Nobel, who challenged me like never before and ensured that I became the best student I could possibly be. He redefined my understanding of the word 'understanding' and fostered

my love for Stochastics, in particular Probability Theory, which ultimately inspired me to pursue a degree in Mathematics. I am also very grateful to Kees-Jan van Garderen and Harry van Zanten for encouraging me to pursue a PhD. Additionally, I want to thank Hans van Ophem for the course in Microeconometrics, Harry van Zanten for the course in Statistical Theory for High and Infinite-dimensional models, Jakob Söhl for his course Statistics for Stochastic Processes, and Aad van der Vaart and Botond Szabo for their course in Nonparametric Bayesian Statistics, all of whom played a pivotal role in turning me into a proper statistician.

I am thankful for my loving and supportive family. Julian and Shanta, my incredible parents who always believe in me. My brother and best friend, Zubin, who inspires me to strive for my best. Maya and Hans, thank you for making me feel the warmth of Suriname here in the Netherlands.

For my amazing colleagues, many of whom have also become friends throughout the years. We made sure that Delft had the most fun math departement out there! Lasse, thank you for all the inspiring conversations we've had about research. Your enthusiasm for research is contagious. I'm truly grateful to have you as a friend. Nevena, thank you for listening to all my stories. Cherishing the good times with Sebastiano, Serena, and Matteo. Thank you for all the pizzas, laughs, and tennis matches. Sharing the best office with Marc, Wieger, Sebastiano, Shixiang, and Mate. It's always a good day with Francesco around! The best coffee breaks with Andrea, working from exotic places with Thorben, the best boardgame nights with Jan-Tino, Serena, Marc, and Hidde. Solving complicated brainteasers with Bart G., relaxed talks about food with Bart E., nice walks with Paul, good intellectual discussions with Jasper and Jonathan. Much-needed breaks with Chris and Andreas, lunch at the best food-truck with Vicente, and always a reading group to attend with Stefan. Furthermore, I would also like to thank Michal, Thomas, Jakob, Hanne, Alexis, Nestor, Martina, Fabian, Elena, Chenguang, Richard, and Dorota for the enjoyable lunches at the department.

Needless to say, I am also very grateful to all my friends with whom I could take my mind off work and enjoy our time together. I would especially like to thank you, Alyssa, for being the lifelong friend I can always count on and for always bringing out the best in me.

Publications

Published

A. Pengel and J. Bierkens. Strong invariance principles for ergodic Markov processes. *Electronic Journal of Statistics*, 18(1):191–246, 2024.

Submitted

A. Pengel, J. Yang, and Z. Zhou. Gaussian approximation and output analysis for high-dimensional MCMC.
arXiv preprint, arXiv:2407.05492, 2024.

In preparation

A. Pengel, J. Yang, and Z. Zhou. High-dimensional Gaussian approximation for continuous-time processes. *In preparation.*

Conferences & Invited Talks

Conferences

Bernoulli-IMS World Congress
Talk: Gaussian Approximation and Output Analysis for High-dimensional MCMC — *Bochum, Germany, August 2024*

International Society for Bayesian Analysis (ISBA)
Poster presentation — *Venice, Italy, July 2024*

BayesComp
Talk: Strong approximation results for PDMC — *Levi, Finland, March 2023*

Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing
Talk: Strong invariance principles for ergodic Markov processes — *Linz, Austria, July 2022*

Invited Talks

Algorithms and Computationally Intensive Inference Seminar
University of Warwick — *Warwick, UK*
Talk: Output Analysis for High-dimensional MCMC — *April 2024*

UCPH Statistics Seminar
University of Copenhagen — *Copenhagen, Denmark*
Talk: Gaussian approximation and variance estimation — *October 2023*

for high-dimensional MCMC

Oxford Computational Statistics and Machine Learning Seminar
University of Oxford *Oxford, UK*
Talk: Gaussian Approximation and Output Analysis *June 2023*
for high-dimensional MCMC