Invariance Principles in Rough Path Topology for Independent Random Variables and Markov Chains

By
Manuela Ott

Master Thesis

Supervisor: Prof. Dr. Erwin Bolthausen

Institute of Mathematics
Faculty of Science
University of Zurich

University of Zurich[UZH]

October 2012

Abstract

In this thesis, we first introduce some basics of the theory of rough paths and then we present a Donsker-type invariance principle in rough path topology due to E. Breuillard, P. Friz and M. Huesmann [3]. We extend their invariance principle to Markov chains. The limit process is shown to be a diffusion process with drift. The strategy of our proof is to establish tightness and to characterize the law of the limit process as the unique solution to a martingale problem.

Acknowledgments

Firstly, I would like to thank Professor Erwin Bolthausen for the guidance and for numerous enlightening discussions.
I would also like to thank Philipp Thomann for helpful hints.
Thanks also go to Wito Traub for tips concerning paper [3].

Contents

1 Introduction

1.1 Motivation and overview

In this thesis, we study weak convergence results of the following form: We start with a sequence $(X_i)_{i\geq 1}$ of (possibly dependent) identically distributed random variables. From these random variables we then construct by interpolation a sequence of stochastic processes $(Y^n)_{n\geq 1}$ such that Y^n converges weakly to a limit process Y as $n \to \infty$ – in the sense that the measures induced by these stochastic processes converge weakly in an appropriate topological function space. If the law of the limit process does not depend on the exact distribution of the X_i, but only on some moments of the X_i and the dependence between X_i and X_j $(i \neq j)$, such results are called invariance principles. A basic invariance principle is Donsker's theorem, which says that a sequence of suitably interpolated (real-valued) random walks[1] converges weakly to Brownian motion.

A generalization of Donsker's theorem to random variables taking values in a certain Lie group – denoted G_d (depending on the dimension $d \in \mathbb{N}$) in this section – was established in the paper "From Random Walks to Rough Paths" by E. Breuillard, P. Friz and M. Huesmann [3]. That paper is the motivation for this thesis, which aims at establishing a similar invariance principle for a stochastic process constructed from a Markov chain (also taking values G_d) in place of the sequence of independent random variables. The Lie group G_d is central to the theory of rough paths. To explain the interest in this state space, we briefly give some basic ideas of this theory, which was developed by T. Lyons and his co-authors in the 1990s.

The theory of rough paths is a deterministic, non-linear extension of the classical theory of integration and differential equations to more irregular paths. This extension was motivated by stochastic modeling. To introduce the terminology, let us first consider a classical differential equation of the following form

$$dy_t = f(y_t)dx_t,$$

where y is the solution, x the driving signal or controlling term and f a vector field (a function in the one-dimensional case). y_t models the state of a dynamical system (for example the speed of a parachutist) whose evolution is modified by the changes of an external parameter x_t (for example the time t). If we want to model a system that is driven by an irregular random source x which is not of finite variation (for example Brownian motion), we cannot use classical differential equations. Rough path theory allows us to deal with more irregular driving signals, which are of finite

[1] i.e. the sequence $(S_n)_{n\geq 1}$, where $S_n = \sum_{i=1}^n X_i$ and $(X_i)_{i\geq 1}$ is a sequence of independent, identically distributed random variables.

p-variation for some $p \in (1, \infty)$. A continuous map $x : [0,1] \to \mathbb{R}^d$ is said to be of finite p-variation if

$$\sup_{\substack{0=t_0<t_1<\cdots<t_k=1 \\ \text{partition of } [0,1],\ k\in\mathbb{N}}} \sum_{i=0}^{k-1} \left\| x_{t_{i+1}} - x_{t_i} \right\|^p < \infty.$$

Since there are a lot of interesting stochastic processes which are not of finite variation, but of finite p-variation for some $p > 2$, rough path theory has various probabilistic applications.
There is a fundamental problem with the case $p \geq 2$: There exist real-valued paths x, which are of finite p-variation for some $p \geq 2$, such that the Riemann sums $\sum_i x_{t_i}(x_{t_{i+1}} - x_{t_i})$ do not converge as the mesh of the partition tends to zero. Thus, the integral $\int_0^t x_u \ dx_u$ cannot be defined in the usual way as a limit of Riemann sums.
The approach taken in rough path theory for the case $p \geq 2$ is to replace the original path x by a path $\mathbf{x}$ in a higher-dimensional vector space. The new path $\mathbf{x}$ – called the lift of x – is constructed to encode the information needed to solve differential equations driven by x in an efficient way. Differential equations can then be defined and solved with respect to these lifted paths. Before the development of rough path theory, iterated integrals of paths were used by different authors to obtain pathwise Taylor series for the solutions of controlled differential equations. It was also shown by K.T. Chen that a smooth path is uniquely determined by a Taylor series involving its iterated integrals. Thus, for a smooth path, it makes sense to define its lift as the sequence of its (low order) iterated integrals. For paths of finite p-variation for some $p \geq 2$, the lifted paths are usually taken as elements of the closure of lifted smooth paths with respect to an $1/p$-Hölder (or p-variation) distance. Elements of this closure are called geometric rough paths. The appropriate dimension of the vector space for the lift of a path x depends on the irregularity of x: The more irregular the path, the higher the dimension of the vector space.
The main result of rough path theory is that the solution to a differential equation driven by a geometric rough path $\mathbf{x}$ depends continuously on $\mathbf{x}$ with respect to a suitable $1/p$-Hölder (or p-variation) distance. In this thesis, we call the topology induced by the $1/p$-Hölder distance on geometric rough paths the "rough path topology". The above result has a lot of applications to stochastic analysis. In particular, we have the following application to invariance principles: Assume $(\mathbf{Z^n})_{n\geq 1}$ is a sequence of stochastic processes the paths of which are geometric rough paths and $\mathbf{Z^n}$ converges weakly to a limit process $\mathbf{Z}$ in rough path topology. Then, as weak convergence is preserved under continuous maps, the sequence of solutions to the differential equations driven by $\mathbf{Z^n}$ will converge weakly to the solution of the differential equation driven by $\mathbf{Z}$ in rough path topology. Results of this type are particularly useful if the limit process is enhanced Brownian motion[2] i.e. the rough path lift of Brownian motion. This is due to the fact that several objects studied

[2]This term is used by P. Friz and his co-authors in [3] and [5].

in the theory of stochastic integration and stochastic differential equations are continuous functions of enhanced Brownian motion in rough path topology, which is stronger than the uniform topology on geometric rough paths. Following Breuillard-Friz-Huesmann, we will therefore focus on weak convergence results in rough path topology.

Some parts of the proof given by Breuillard et al. did not easily allow an extension to the case of dependent random variables. We therefore relied on a different concept, called the martingale problem formulation. To see how the martingale problem formulation can be used to prove invariance principles, we first studied two simpler cases: A stochastic process constructed from a real-valued Markov chain and a special case of the setting in Breuillard et al.'s paper. The arguments developed in these two cases could then be extended to Markov chains taking values in the Lie group G_2 [3]. To our knowledge, this is a new result.

The structure of this thesis is as follows:
In the next section of the introduction, we give the precise statement of Donsker's theorem and a related invariance principle in a stronger topology due to Lamperti. At the end of the introduction, we fix some notation.
Chapters 2 and 3 contain basics which will be used to state and prove the invariance principles in the subsequent chapters. Firstly, we introduce some relevant objects and results from the theory of rough paths. Secondly, we give some results related to weak convergence and explain the basic proof strategy used by Breuillard-Friz-Huesmann (i.e. proving tightness and convergence of the finite-dimensional distributions to establish weak convergence in a function space).
Chapter 4 is devoted to the Breuillard-Friz-Huesmann invariance principle. We present their proof with more intermediate steps and explanations than in the original paper.
In Chapter 5, we introduce two important notions for the proofs of our invariance principles: The martingale problem formulation and a measure for dependence between random variables called the strong mixing coefficient.
In the last three chapters, we finally state and prove our three invariance principles. A central ingredient for our proofs involving Markov chains $(X_i)_{i\geq 1}$ are covariance bounds. In the two cases we study, $\mathrm{Cov}(X_i, X_j)$ decreases exponentially fast as a function of the distance $|i-j|$.

[3] We restrict our argument to the case $d = 2$ to concentrate on the conceptual aspects and to keep the notation at a manageable level.

1.2 Basic invariance principles and Hölder topology

Donsker's theorem involves the space of continuous functions $[0,1] \to \mathbb{R}$, denoted $C([0,1],\mathbb{R})$, endowed with the uniform topology.

Theorem 1.1 (Donsker's theorem, [1])**.** *Let $(X_i)_{i\geq 1}$ be a sequence of iid*[4]*, real-valued random variables with mean 0 and finite variance $\sigma^2 > 0$. Define $S_0 = 0$ and $S_n = \sum_{i=1}^n X_i$ for $n \in \mathbb{N}$. Consider the rescaled random walk $Y^{(n)} = \{Y_t^{(n)}\}_{t\in[0,1]}$ defined by*

$$Y_t^{(n)} = \frac{1}{\sigma\sqrt{n}} S_{nt}$$

for $t \in \left\{0, \frac{1}{n}, \frac{2}{n}, \ldots, 1\right\}$ and piecewise linearly interpolated in between, i.e.

$$Y_t^{(n)} = \frac{1}{\sigma\sqrt{n}} \left(S_{\lfloor nt\rfloor} + (nt - \lfloor nt\rfloor)X_{\lfloor nt\rfloor+1}\right)$$

for $t \in [0,1]$.
Then, as $n \to \infty$, $Y^{(n)}$ converges weakly to a standard Brownian motion B in $C([0,1],\mathbb{R})$, i.e. for any bounded, continuous function $f : C([0,1],\mathbb{R}) \to \mathbb{R}$, we have

$$\lim_{n\to\infty} E[f(Y^{(n)})] = E[f(B)]. \tag{1.1}$$

Note that Donsker's theorem can be generalized to $\mathbb{R}^d$-valued random variables. One then obtains weak convergence to d-dimensional Brownian motion.
Donsker's theorem has the following physical interpretation: Consider a particle which is displaced according to X_i at time point $i\tau$, where τ is a small time interval and $X_1, X_2, \ldots$ are independent. Viewed from a large distance, this particle seems to perform a Brownian motion.

Let us explain some arguments related to statement (1.1), since results of this form will be central to this thesis: First note that for each $\omega \in \Omega$, the map $Y_\cdot^{(n)}(\omega) : [0,1] \to \mathbb{R}$ is continuous i.e. $Y_\cdot^{(n)}(\omega) \in C([0,1],\mathbb{R})$. Let us consider the Borel σ-algebra $\mathcal{C}$ on $C([0,1],\mathbb{R})$ endowed with the uniform topology. One can show that the map

$$Y^{(n)} : (\Omega, \mathcal{F}) \to (C([0,1],\mathbb{R}), \mathcal{C})$$

is measurable i.e. $Y^{(n)}$ can be viewed as a random variable with values in $C([0,1],\mathbb{R})$. Hence, $Y^{(n)}$ induces the image measure $P\left(Y^{(n)}\right)^{-1}$ on $C([0,1],\mathbb{R})$ and (1.1) is equivalent to saying that $P\left(Y^{(n)}\right)^{-1}$ converges weakly to the Wiener measure.

In fact, $Y^{(n)}$ and Brownian motion are not only continuous, but α-Hölder continuous for any $\alpha \in (0, 1/2)$. Hence, we can restrict our attention to the space of α-Hölder continuous functions, an approach due to Lamperti (see [11]). He strengthened

[4] independent and identically distributed.

Donsker's invariance principle in the following way: Let $C^{\alpha-H\ddot{o}l}([0,1],\mathbb{R})$ denote the space of functions $x:[0,1]\to\mathbb{R}$ such that $x(0)=0$ and

$$\|x\|_\alpha := \sup_{0\leq s<t\leq 1}\frac{|x(t)-x(s)|}{|t-s|^\alpha} + \sup_{t\in[0,1]}|x(t)| < \infty,$$

endowed with the topology induced by the norm $\|.\|_\alpha$. If the X_i have a finite moment of order $2p$ for some $p\in\mathbb{N}\setminus\{1\}$, then the function space $C([0,1],\mathbb{R})$ in Donsker's theorem can be replaced by $C^{\alpha-H\ddot{o}l}([0,1],\mathbb{R})$ for any $\alpha\in(0,(p-1)/(2p))$.
Since the above topology on $C^{\alpha-H\ddot{o}l}([0,1],\mathbb{R})$ is stronger than the uniform one, there are more continuous functions on $C^{\alpha-H\ddot{o}l}([0,1],\mathbb{R})$, which is advantageous for statistical applications such as estimation and testing hypotheses.

1.3 Notation and conventions

$\mathbb{N}$	natural numbers excluding 0
$\lfloor x \rfloor$	the largest integer smaller or equal to $x \in \mathbb{R}$
$\lceil x \rceil$	the smallest integer larger or equal to $x \in \mathbb{R}$
$\lvert x \rvert$	Euclidean norm of $x \in \mathbb{R}^d$
$C(I, E)$	set of continuous maps $I \to E$ for $I \subset \mathbb{R}$ interval, (E, d) metric space
$X \stackrel{D}{=} Y$	The random variables X and Y have the same distribution.
$X_n \stackrel{D}{\to} X$	convergence in distribution
$X_n \stackrel{P}{\to} X$	convergence in probability

Short notation for paths

If f is a map having preimage $I \subset \mathbb{R}$, we often write f_t instead of $f(t)$ for $t \in I$.

The Landau symbols $\boldsymbol{O}$ and $\boldsymbol{o}$

Consider two functions $f, g : \mathbb{N} \to \mathbb{R}$.

$f(n) = O(g(n))$ (for $n \to \infty$) $:\Leftrightarrow$
There exist constants $K > 0$ and $N \in \mathbb{N}$ such that $|f(n)| \leq K\,|g(n)|$ for all $n \geq N$.

$f(n) = o(g(n))$ (for $n \to \infty$) $:\Leftrightarrow$
For every $\epsilon > 0$ there exists a constant $N \in \mathbb{N}$ such that $|f(n)| \leq \epsilon\,|g(n)|$ for all $n \geq N$.

Probability spaces

If we do not mention anything else, the probability space on which random variables are defined will always be arbitrary and denoted by $(\Omega, \mathcal{F}, P)$.

Random variables

Random variables taking values in a Lie group will be written in bold face. Random variables in normal type take their values in $\mathbb{R}^d$ or in a Lie algebra.

2 Some basics of the theory of rough paths

In this chapter, we introduce some objects and results from the theory of rough paths. We also included the proofs of some results to illustrate how one can work with these objects. However, we will not introduce integrals with respect to rough paths and rough differential equations. For this, we refer the interested reader to [12] or [14].
An interesting, but also challenging feature of rough path theory is that it combines analytical and algebraic aspects and it also has connections to differential geometry. Rough path theory can be studied from several different conceptual viewpoints. We will mainly follow the approach taken by P. Friz and his co-authors in [3] and [5]. The latter[1] is the main reference for this chapter.
We will only develop the theory for handling driving signals of finite p-variation for $p \in (2,3]$ – a case that is often used in probabilistic applications. We will first introduce the collection of iterated integrals up to level 2, called the (step-2) signature. Next, we will study the underlying tensor space. Using some of the definitions and results given for this tensor space, we will then investigate some properties of the signature. In the next section, we will see that the set of all signatures forms a Lie group. We will then define a metric on this Lie group. This metric will be used to define geometric rough paths. Finally, we will treat the rough path lift of Brownian motion.

2.1 The signature of a path

Definition 2.1. For $s,t \in \mathbb{R}_{\geq 0}$, $s < t$, let $D([s,t])$ denote the set of partitions of the interval $[s,t]$. An element D of $D([s,t])$ is of the form $s = t_0 < t_1 < \cdots < t_k = t$ for some $k \in \mathbb{N}$ and will be denoted $D = (t_i)_{i=0,\dots,k}$ or simply $D = (t_i)$ for short. We define the mesh $|D|$ of D by $|D| := \max_{i=0,\dots k-1}(t_{i+1} - t_i)$.
A path $x : [s,t] \to \mathbb{R}^d$ is said to be of **finite 1-variation** if

$$|x|_{1-var;[s,t]} := \sup_{(t_i) \in D([s,t])} \sum_i \left| x_{t_{i+1}} - x_{t_i} \right| < \infty.$$

$C^{1-var}\left([s,t], \mathbb{R}^d\right) := \left\{ x : [s,t] \to \mathbb{R}^d \mid x \text{ is continuous and of finite 1-variation} \right\}.$

Remark 2.2. Note that $|.|_{1-var;[s,t]}$ is only a semi-norm on $C^{1-var}\left([s,t], \mathbb{R}^d\right)$. But if we identify two paths x, $y : [s,t] \to \mathbb{R}^d$ satisfying $x_u - y_u = \text{const.}$ for all $u \in [s,t]$, then $|.|_{1-var;[s,t]}$ is a norm on these equivalence classes of paths.

Let $(e_i)_{i=1,\dots,d}$ be the canonical basis of $\mathbb{R}^d$. Then, $(e_i \otimes e_j)_{i,j \in \{1,\dots,d\}}$ is the canonical basis of the vector space $\left(\mathbb{R}^d\right)^{\otimes 2} := \mathbb{R}^d \otimes \mathbb{R}^d$.

[1] mainly Chapters 7-9.

Definition 2.3. Let $x \in C^{1-var}\left([s,t], \mathbb{R}^d\right)$. We define

$$\left(\int_{s<u<t} dx_u^i\right)_{i=1,\dots,d} := \sum_{i=1}^{d} \left(\int_s^t 1\ dx_u^i\right) e_i \in \mathbb{R}^d; \tag{2.1}$$

$$\left(\int_{s<u<v<t} dx_u^i dx_v^j\right)_{i,j=1,\dots,d} := \sum_{i,j=1}^{d} \left(\int_s^t \int_s^v 1\ dx_u^i dx_v^j\right) (e_i \otimes e_j) \in \left(\mathbb{R}^d\right)^{\otimes 2}. \tag{2.2}$$

For (2.1), we will use the short notation $\int_{s<u<t} dx_u$ and for (2.2), we will write $\int_{s<u<v<t} dx_u \otimes dx_v$.

Remark 2.4. The integrals on the right-hand side of (2.1) and (2.2) are Riemann-Stieltjes integrals. We clearly have $\int_{s<u<t} dx_u = x_t - x_s$.
If the path $x : [s,t] \to \mathbb{R}^d$ is absolutely continuous, then $\int_{s<u<t} dx_u = \int_s^t \dot{x}_u du$.

Definition 2.5. For $x \in C^{1-var}\left([s,t], \mathbb{R}^d\right)$, the **step-2 signature** is defined by

$$S_2(x)_{s,t} := \left(1, \int_{s<u<t} dx_u, \int_{s<u<v<t} dx_u \otimes dx_v\right) \in \mathbb{R} \oplus \mathbb{R}^d \oplus \left(\mathbb{R}^d\right)^{\otimes 2}.$$

The path $u \mapsto S_2(x)_{s,u}$ is called the **(step-2) lift** of x.

2.2 Tensor spaces

We now study the vector space $\mathbb{R} \oplus \mathbb{R}^d \oplus \left(\mathbb{R}^d\right)^{\otimes 2}$ underlying the step-2 signature and some useful subsets of this space.
First, we recall the definition of the tensor product on $\mathbb{R}^d$.

Definition 2.6. For

$$a = \sum_{i=1}^{d} a^i e_i \in \mathbb{R}^d, \quad b = \sum_{i=1}^{d} b^j e_j \in \mathbb{R}^d,$$

the tensor product, which is a bilinear map $\mathbb{R}^d \times \mathbb{R}^d \to \left(\mathbb{R}^d\right)^{\otimes 2}$, is defined as

$$a \otimes b := \sum_{i,j=1}^{d} a^i b^j (e_i \otimes e_j) \in \left(\mathbb{R}^d\right)^{\otimes 2}.$$

Definition 2.7. Let $d \in \mathbb{N}$. We define the vector space $T^2\left(\mathbb{R}^d\right) := \mathbb{R} \oplus \mathbb{R}^d \oplus (\mathbb{R}^d)^{\otimes 2}$.
For $g \in T^2\left(\mathbb{R}^d\right)$, we write $g = (g_0, g_1, g_2) := g_0 + g_1 + g_2$.
We set $\left(\mathbb{R}^d\right)^{\otimes 0} := \mathbb{R}$.
For $k \in \{0,1,2\}$, let $\pi_k : T^2\left(\mathbb{R}^d\right) \to \left(\mathbb{R}^d\right)^{\otimes k}$ be the projection to the k-th tensor level.

Remark 2.8. We can identify $\left(\mathbb{R}^d\right)^{\otimes 2}$ with the set of real-valued $d \times d$-matrices by writing the factor in front of $e_i \otimes e_j$ in the i-th row, j-th column of the matrix. So we can think of an element of $T^2\left(\mathbb{R}^d\right)$ as a collection of a real number, an element of $\mathbb{R}^d$ and a $d \times d$ matrix with real entries.

Definition 2.9. We define the product of $(g_0, g_1, g_2), (h_0, h_1, h_2) \in T^2\left(\mathbb{R}^d\right)$ by

$$(g_0, g_1, g_2) \otimes (h_0, h_1, h_2) := (g_0 h_0, g_0 h_1 + h_0 g_1, g_0 h_2 + h_0 g_2 + g_1 \otimes h_1).$$

Note that the product $\otimes$ is not commutative.
It is easy to see that this product satisfies the laws of associativity and distributivity and that it is compatible with scalar multiplication (in the sense that for $r, s \in \mathbb{R}$ and $g, h \in T^2\left(\mathbb{R}^d\right)$, we have $(rg) \otimes (sh) = (rs)(g \otimes h)$). Thus, we have

Lemma 2.10. *The space* $(T^2\left(\mathbb{R}^d\right), \otimes)$ *is an associative algebra over* $\mathbb{R}$ *with neutral element* $\mathbf{1} := (1, 0, 0)$.

We now define a norm on $T^2\left(\mathbb{R}^d\right)$.

Definition 2.11. First, we introduce a norm on $\left(\mathbb{R}^d\right)^{\otimes 2}$: For

$$a = \sum_{i,j=1}^{d} a^{ij} e_i \otimes e_j \in \left(\mathbb{R}^d\right)^{\otimes 2}, \quad |a|_{(\mathbb{R}^d)^{\otimes 2}} := \sqrt{\sum_{i,j=1}^{d} |a^{ij}|^2}.$$

Then, for $g = (g_0, g_1, g_2) \in T^2\left(\mathbb{R}^d\right)$,

$$|g|_{T^2\left(\mathbb{R}^d\right)} := \max\left\{ |g_0|, |g_1|, |g_2|_{(\mathbb{R}^d)^{\otimes 2}} \right\}.$$

In the sequel, we will just write $|.|$ in place of $|.|_{(\mathbb{R}^d)^{\otimes 2}}$. The meaning will be clear from the context.
Remark 2.12. $(T^2\left(\mathbb{R}^d\right), |.|_{T^2\left(\mathbb{R}^d\right)})$ is a Banach space of dimension $1 + d + d^2$.

Next, we consider some useful subsets of $T^2\left(\mathbb{R}^d\right)$ having fixed zeroth components.

Definition 2.13. $t^2\left(\mathbb{R}^d\right) := \left\{ g \in T^2\left(\mathbb{R}^d\right) : \pi_0(g) = 0 \right\}$.
$1 + t^2\left(\mathbb{R}^d\right) := \left\{ g \in T^2\left(\mathbb{R}^d\right) : \pi_0(g) = 1 \right\}$.
Let us also define the map

$$\begin{aligned} [.,.] : t^2\left(\mathbb{R}^d\right) \times t^2\left(\mathbb{R}^d\right) &\to t^2\left(\mathbb{R}^d\right) \\ (g, h) &\mapsto [g, h] := g \otimes h - h \otimes g. \end{aligned}$$

Clearly, the bracket operation $[.,.]$ is anticommutative (i.e. $[g, h] = -[h, g]$). As the product $\otimes$ on $T^2\left(\mathbb{R}^d\right)$ is bilinear, $[.,.]$ is also bilinear. One can check by a simple computation that this bracket satisfies Jacobi's identity (i.e. [f,[g,h]] + [g,[h,f]] + [h,[f,g]]=0 for all $f, g, h \in t^2\left(\mathbb{R}^d\right)$). Consequently, this bracket operation is a Lie bracket and we have

Lemma 2.14. $(t^2(\mathbb{R}^d), [.,.])$ *is a Lie algebra over* $\mathbb{R}$.

Let us endow $1 + t^2(\mathbb{R}^d)$ with the topology induced by the metric

$$\rho(g,h) := |g-h|_{T^2(\mathbb{R}^d)} = \max_{i=1,2} |\pi_i(g-h)|. \tag{2.3}$$

Then, we have

Lemma 2.15. $1 + t^2(\mathbb{R}^d)$ *is a Lie group with respect to* $\otimes$.

Proof. $1 + t^2(\mathbb{R}^d)$ is a group with respect to $\otimes$:
$\mathbf{1}$ is the neutral element. For $g = (1, g_1, g_2) \in 1 + t^2(\mathbb{R}^d)$, the inverse is

$$g^{-1} = (1, -g_1, -g_2 + g_1 \otimes g_1).$$

$1 + t^2(\mathbb{R}^d)$ is a C^∞-manifold since it is diffeomorphic to the vector space $t^2(\mathbb{R}^d) \cong \mathbb{R}^{d+d^2}$, which is a C^∞-manifold. The group operations $\otimes$ and $(.)^{-1}$ are polynomial functions in the coordinates and hence they are smooth. $\square$

2.3 The exponential map and the Campbell-Baker-Hausdorff formula

The exponential and logarithm maps between the two subspaces $t^2(\mathbb{R}^d)$ and $1+ t^2(\mathbb{R}^d)$ of $T^2(\mathbb{R}^d)$ are just truncated versions of the usual exponential and logarithm series.

Definition 2.16.

$$\begin{aligned} exp : t^2\left(\mathbb{R}^d\right) &\to 1 + t^2\left(\mathbb{R}^d\right) \\ a &\mapsto \mathbf{1} + a + \frac{a \otimes a}{2} = \left(1, a_1, a_2 + \frac{1}{2} a_1 \otimes a_1\right); \\ log : 1 + t^2\left(\mathbb{R}^d\right) &\to t^2\left(\mathbb{R}^d\right) \\ \mathbf{1} + a &\mapsto a - \frac{a \otimes a}{2} = \left(0, a_1, a_2 - \frac{1}{2} a_1 \otimes a_1\right). \end{aligned}$$

Clearly, exp and log are inverse maps. Hence, they are bijective.
For the Lie algebra $t^2(\mathbb{R}^d)$, the Campbell-Baker-Hausdorff formula from the theory of Lie algebras and Lie groups has a very simple form and can be obtained by an easy computation (which is left to the reader).

Theorem 2.17 (Campbell-Baker-Hausdorff formula for $t^2(\mathbb{R}^d)$)**.** *For* $a, b \in t^2(\mathbb{R}^d)$, *we have*

$$\exp(a) \otimes \exp(b) = \exp\left(a + b + \frac{1}{2}[a,b]\right).$$

2.4 The signature revisited

We now consider an important example – the signature of a linear path. This result will be used several times in the following proofs.

Example 2.18. Let $a \in \mathbb{R}^d$. We compute the step-2 signature of the path $x : [0,1] \to \mathbb{R}^d; t \mapsto ta$:

$$\begin{aligned} S_2(x)_{0,1} &= \left(1, \int_{0<u<1} dx_u, \int_{0<u<v<1} dx_u \otimes dx_v\right) \\ &= \left(1,\ a\int_{0<u<1} du,\ a \otimes a \int_{0<u<v<1} du\ dv\right) \\ &= \left(1,\ a,\ \frac{a\otimes a}{2}\right) = \exp\left((0,a,0)\right). \end{aligned}$$

Remark 2.19. In the sequel, we will write $\exp(a)$ in place of $\exp\left((0,a,0)\right)$ for $a \in \mathbb{R}^d$.

Let us consider the second-order iterated integrals appearing in the signature more closely. They are elements of $\mathbb{R}^d \otimes \mathbb{R}^d$, which can be identified with the set of real-valued $d \times d$ matrices (Remark 2.8). We fix a path $x \in C^{1-var}\left([0,1], \mathbb{R}^d\right)$. We can decompose $M := \pi_2\left(S_2(x)_{0,1}\right)$ into its symmetric and antisymmetric parts

$$S = \frac{1}{2}\left(M + M^T\right) \text{ and } A = \frac{1}{2}\left(M - M^T\right).$$

We write $S = (s^{ij})_{i,j\in\{1,\dots,d\}}$ and $A = (a^{ij})_{i,j\in\{1,\dots,d\}}$. The symmetric part S can be expressed by the increments of the path x since

$$\begin{aligned} s^{ij} =& \frac{1}{2}\left(\int_{0<u<v<1} dx_u^i dx_v^j + \int_{0<u<v<1} dx_u^j dx_v^i\right) \\ =& \frac{1}{2}\left(\int_0^1 x_v^i\ dx_v^j + \int_0^1 x_v^j\ dx_v^i - \left(\int_0^1 x_0^i\ dx_v^j + \int_0^1 x_0^j\ dx_v^i\right)\right) \\ =& \frac{1}{2}\left(x_1^i x_1^j - x_0^i x_0^j - \left(x_0^i\left(x_1^j - x_0^j\right) + x_0^j\left(x_1^i - x_0^i\right)\right)\right) \qquad (2.4) \\ =& \frac{1}{2}\left(x_1^i - x_0^i\right)\left(x_1^j - x_0^j\right), \end{aligned}$$

where in (2.4), we have done integration by parts for the first term. Hence, the symmetric part S is determined by $\pi_1\left(S_2(x)_{0,1}\right)$. Thus, all additional information contained in the second-order iterated integrals must be encoded in the antisymmetric part A.

For $i \neq j$, the entry a^{ij} of the antisymmetric part has a nice geometric interpretation: Consider the path $u \in [0,1] \mapsto (x_u^i, x_u^j) \in \mathbb{R}^2$ and add the line segment joining (x_1^i, x_1^j) to (x_0^i, x_0^j) to obtain a closed directed curve. Then, an application of Green's theorem yields that a^{ij} is the area enclosed by this curve if orientation and multiplicity are taken into account (see Figure 2.1).

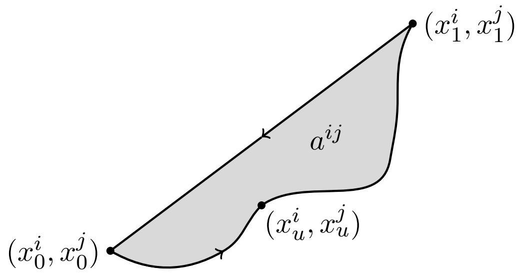

Figure 2.1: The entry in the i-th row, j-th column of the antisymmetric part of $\pi_2\left(S_2(x)_{0,1}\right)$ is the signed area enclosed by the above directed curve.

2.5 The Lie group $G^2\left(\mathbb{R}^d\right)$

In the following, the set of all step-2 signatures of continuous paths of finite 1-variation will be the state space in which the random variables take their values. We therefore study this set in some detail. First, we introduce some additional objects which will turn out to be closely related to this set of signatures.

Preliminaries

Definition 2.20. $G^2\left(\mathbb{R}^d\right) := \left\{S_2(x)_{0,1} : x \in C^{1-var}\left([0,1], \mathbb{R}^d\right)\right\}$.

Remark 2.21. In the last section, we have seen that an element of $G^2\left(\mathbb{R}^d\right)$ is determined by a path increment and an area associated to this increment.

Definition 2.22. For $v, w \in \mathbb{R}^d$, let $[v,w] := v \otimes w - w \otimes v \in \mathbb{R}^d \otimes \mathbb{R}^d$.
We define $[\mathbb{R}^d, \mathbb{R}^d] := \operatorname{span}\left([v,w] : v, w \in \mathbb{R}^d\right)$, where span denotes the linear span over $\mathbb{R}$.
Let $g^2\left(\mathbb{R}^d\right)$ be the smallest Lie subalgebra of $t^2\left(\mathbb{R}^d\right)$ which contains $\pi_1(t^2\left(\mathbb{R}^d\right)) \cong \mathbb{R}^d$. Thus, $g^2\left(\mathbb{R}^d\right) \cong \mathbb{R}^d \oplus [\mathbb{R}^d, \mathbb{R}^d]$.

Remark 2.23. Note that $[\mathbb{R}^d, \mathbb{R}^d]$ can be identified with the set of skew-symmetric matrices.

Definition 2.24. $\left\langle \exp\left(\mathbb{R}^d\right)\right\rangle := \left\{\otimes_{i=1}^m \exp(v_i) : m \in \mathbb{N}, v_1, \ldots, v_m \in \mathbb{R}^d\right\}$.

Remark 2.25.

a) $\left\langle \exp\left(\mathbb{R}^d\right)\right\rangle$ is a subgroup of $1 + t^2\left(\mathbb{R}^d\right)$:

   $\exp(0) = \mathbf{1}$ is the neutral element and for $v_1, \ldots, v_m \in \mathbb{R}^d$, the inverse of $\exp(v_1) \otimes \cdots \otimes \exp(v_m)$ is $\exp(-v_m) \otimes \cdots \otimes \exp(-v_1)$.

b) $\left\langle \exp\left(\mathbb{R}^d\right)\right\rangle$ is the smallest subgroup of $1 + t^2\left(\mathbb{R}^d\right)$ which contains the set $\exp(\mathbb{R}^d)$.

Different representations

Theorem 2.26 (Representation theorem for $G^2\left(\mathbb{R}^d\right)$)**.** *We have*

$$G^2\left(\mathbb{R}^d\right)=\exp\left(g^2\left(\mathbb{R}^d\right)\right)=\left\langle\exp\left(\mathbb{R}^d\right)\right\rangle,$$

and $G^2\left(\mathbb{R}^d\right)$ is a closed Lie subgroup of $\left(1+t^2\left(\mathbb{R}^d\right),\otimes\right)$.

Remark 2.27.

a) Due to the equality $G^2\left(\mathbb{R}^d\right)=\exp\left(g^2\left(\mathbb{R}^d\right)\right)$, each $g\in G^2\left(\mathbb{R}^d\right)$ can be written as $g=(1,v,A+\frac{1}{2}v\otimes v)$, where A is a skew-symmetric matrix and $v\in\mathbb{R}^d$.

b) Recall the definition of the metric ρ on $1+t^2\left(\mathbb{R}^d\right)$ given in (2.3). The metric $\rho|_{G^2\left(\mathbb{R}^d\right)}$ induces the (sub)manifold topology of $G^2\left(\mathbb{R}^d\right)$.

To prove the representation theorem for $G^2\left(\mathbb{R}^d\right)$, we will need several results related to paths and their signatures. These results will be stated in the next subsection. The proofs are given in [5].

Some results for paths and their signatures

Definition 2.28. Let $\gamma\in C^{1-var}\left([0,T],\mathbb{R}^d\right)$ and $\eta\in C^{1-var}\left([T,2T],\mathbb{R}^d\right)$. The concatenation $\gamma\sqcup\eta$ of these two paths is defined by

$$\gamma\sqcup\eta(t):=\begin{cases}\gamma(t) & \text{for } t\in[0,T]\\ \eta(t)-\eta(0)+\gamma(T) & \text{for } t\in[T,2T],\end{cases}$$

so that we have $\gamma\sqcup\eta\in C^{1-var}\left([0,2T],\mathbb{R}^d\right)$. If γ and η are both parametrized over the interval [0,T], we reparametrize the concatenation of the two paths, so that it is parametrized over $[0,T]$, too[2]:

$$\gamma\sqcup\eta(t):=\begin{cases}\gamma(2t) & \text{for } t\in[0,T/2]\\ \eta(2t-T)-\eta(0)+\gamma(T) & \text{for } t\in[T/2,T].\end{cases}$$

Theorem 2.29 (Chen, [5], Th. 7.11)**.** *For $\gamma\in C^{1-var}\left([0,T],\mathbb{R}^d\right)$ and $\eta\in C^{1-var}\left([T,2T],\mathbb{R}^d\right)$, we have*

$$S_2(\gamma\sqcup\eta)_{0,2T}=S_2(\gamma)_{0,T}\otimes S_2(\eta)_{T,2T}.$$

Equivalently, for $x\in C^{1-var}\left([0,T],\mathbb{R}^d\right)$ and $0\leq s<t<u\leq T$, we have

$$S_2(x)_{s,u}=S_2(x)_{s,t}\otimes S_2(x)_{t,u}. \tag{2.5}$$

[2]The signature of a path can be shown to be independent of the chosen parametrization.

The idea for the proof below comes from [14].

Proof. For $t_1, t_2 \in [0,\infty)$, $t_1 < t_2$ and a path $\gamma \in C^{1-var}\big([t_1,t_2],\mathbb{R}^d\big)$, we use the following notation for the signature: $S_2(\gamma)_{t_1,t_2} = (1, \gamma^1_{t_1,t_2}, \gamma^2_{t_1,t_2})$. Let $\gamma \in C^{1-var}\big([0,T],\mathbb{R}^d\big)$ and $\eta \in C^{1-var}\big([T,2T],\mathbb{R}^d\big)$. Then, for the path $\theta := \gamma \sqcup \eta \in C^{1-var}\big([0,2T],\mathbb{R}^d\big)$, we have

$$\begin{aligned}\theta^1_{0,2T} = \int_0^{2T} d\theta_u &= \int_0^T d\gamma_u + \int_T^{2T} d(\eta_u - \eta_0 - \gamma_T)\\ &= \int_0^T d\gamma_u + \int_T^{2T} d\eta_u = \gamma^1_{0,T} + \eta^1_{T,2T}\end{aligned}$$

and

$$\begin{aligned}\theta^2_{0,2T} &= \int_{0<u<v<2T} d\theta_u \otimes d\theta_v\\ &= \int_{0<u<v<T} d\gamma_u \otimes d\gamma_v + \int_{0<u<T<v<2T} d\gamma_u \otimes d\eta_v + \int_{T<u<v<2T} d\eta_u \otimes d\eta_v\\ &= \gamma^2_{0,T} + \left(\int_0^T d\gamma\right) \otimes \left(\int_T^{2T} d\eta\right) + \eta^2_{T,2T}\\ &= \gamma^2_{0,T} + \gamma^1_{0,T} \otimes \eta^1_{T,2T} + \eta^2_{T,2T}.\end{aligned}$$

Hence, we obtain $S_2(\theta)_{0,2T} = S_2(\gamma)_{0,T} \otimes S_2(\eta)_{T,2T}$. □

Chow's theorem states that each element of $\exp\big(g^2\left(\mathbb{R}^d\right)\big)$ can be represented as the step-2 signature of a piecewise linear path.

Theorem 2.30 (Chow, [5], Th. 7.28)**.** *For each $g \in \exp\big(g^2\left(\mathbb{R}^d\right)\big)$, there exist $v_1, \dots, v_m \in \mathbb{R}^d$ such that*

$$g = \exp(v_1) \otimes \cdots \otimes \exp(v_m).$$

In other words, there exists a piecewise linear path $x : [0,1] \to \mathbb{R}^d$ such that $S_2(x)_{0,1} = g$.

From Chow's theorem, one can deduce that

Corollary 2.31 ([5], Cor. 7.29)**.** *Let $(g_n)_{n\geq 1}$ be a sequence in $\exp\big(g^2\left(\mathbb{R}^d\right)\big)$ such that $g_n \xrightarrow[n\to\infty]{} \mathbf{1}$[3]. Then, there exists a sequence of piecewise linear paths $(x^n)_{n\geq 1}$ (parametrized over $[0,1]$) such that x^n has signature g_n and the length of the path x^n, that is $\int_0^1 |dx^n_u|$, converges to 0 as $n \to \infty$.*

Definition 2.32. Let $x : [0,T] \to \mathbb{R}^d$ and let $D = (t_i)_{i=0,\dots k}$ be a partition of $[0,T]$. Then the **piecewise linear approximation** of x is defined by

$$x^D_t := x_{t_i} + \frac{t - t_i}{t_{i+1} - t_i}(x_{t_{i+1}} - x_{t_i})$$

for $t \in [t_i, t_{i+1}]$, $i = 0, \dots, k-1$.

[3]in the metric ρ on $1 + t^2\left(\mathbb{R}^d\right)$.

We will use the fact that the 1-variation of any piecewise linear approximation of a path x does not exceed the 1-variation of x.

Proposition 2.33 ([5], Prop. 1.28)**.** *Let $x \in C^{1-var}\left([0,T],\mathbb{R}^d\right)$. Then, for any partition D of $[0,T]$ and any $s,t \in [0,T]$, $s<t$, we have*

$$\left|x^D\right|_{1-var;[s,t]} \leq |x|_{1-var;[s,t]}\,.$$

If $(D_n)_{n\geq 1}$ is a sequence of partitions of $[0,T]$ such that $\lim_{n\to\infty}|D_n|=0$, then the sequence of piecewise linear approximations $(x^{D_n})_{n\geq 1}$ converges uniformly to x as $n\to\infty$.

Proposition 2.34 ([5], Prop. 7.15)**.** *Let $(x_n)_{n\geq 1}$ be a sequence of paths in $C^{1-var}\left([0,1],\mathbb{R}^d\right)$ with $\sup_{n\geq 1}|x_n|_{1-var;[0,1]}<\infty$, uniformly convergent to some $x \in C^{1-var}\left([0,1],\mathbb{R}^d\right)$. Then, $S_2(x_n)_{0,\cdot}$ converges uniformly to $S_2(x)_{0,\cdot}$.*
In particular, we have

$$\lim_{n\to\infty} S_2(x_n)_{0,1} = S_2(x)_{0,1}.$$

Proof of the representation theorem for $G^2\left(\mathbb{R}^d\right)$ (Theorem 2.26)

Step 1: $\exp\left(g^2\left(\mathbb{R}^d\right)\right) = \left\langle \exp\left(\mathbb{R}^d\right)\right\rangle$

$\subseteq$: This is due to Chow's theorem (Theorem 2.30).

$\supseteq$: Clearly, $\exp\left(\mathbb{R}^d\right) \subset \exp\left(g^2\left(\mathbb{R}^d\right)\right)$. $\left\langle \exp\left(\mathbb{R}^d\right)\right\rangle \subseteq \exp\left(g^2\left(\mathbb{R}^d\right)\right)$ then follows from the Campbell-Baker-Hausdorff formula (Theorem 2.17).

Step 2: $\left\langle \exp\left(\mathbb{R}^d\right)\right\rangle$ is a closed subset of $1+t^2\left(\mathbb{R}^d\right)$.

By Step 1, it is enough to show that $\exp\left(g^2\left(\mathbb{R}^d\right)\right)$ is closed in $1+t^2\left(\mathbb{R}^d\right)$. As the Lie bracket on $t^2\left(\mathbb{R}^d\right)$ is continuous in both arguments, $g^2\left(\mathbb{R}^d\right)$ is a closed subset of $t^2\left(\mathbb{R}^d\right)$. In addition, we have $\exp\left(g^2\left(\mathbb{R}^d\right)\right) = \log^{-1}(g^2\left(\mathbb{R}^d\right))$ and log is a continuous map. Consequently, $\exp\left(g^2\left(\mathbb{R}^d\right)\right)$ is a closed subset of $1+t^2\left(\mathbb{R}^d\right)$.

Step 3: $G^2\left(\mathbb{R}^d\right) = \left\langle \exp\left(\mathbb{R}^d\right)\right\rangle$

$\supseteq$: This is a consequence of Example 2.18 (where we computed the signature of a linear path) combined with Chen's theorem (Theorem 2.29): For $v_1,\ldots,v_m \in \mathbb{R}^d$

$$\exp(v_1)\otimes\cdots\otimes\exp(v_m)$$

is the step-2 signature of the path $x:[0,1]\mapsto\mathbb{R}^d$ defined by $x_0=0$, $x_{i/m} = x_{(i-1)/m}+v_i$ for $i=1,\ldots,m$ and linearly connected in between.

$\subseteq$: By Step 2, we have $\overline{\langle \exp\left(\mathbb{R}^d\right)\rangle} = \left\langle \exp\left(\mathbb{R}^d\right)\right\rangle$. Hence, it suffices to prove

$$G^2\left(\mathbb{R}^d\right) \subseteq \overline{\langle \exp\left(\mathbb{R}^d\right)\rangle}.$$

Let $g \in G^2\left(\mathbb{R}^d\right)$. Then, there exists $x \in C^{1-var}\left([0,1], \mathbb{R}^d\right)$ such that $g = S_2(x)_{0,1}$. It follows from Proposition 2.33 that there exists a sequence $(x_n)_{n\geq 1}$ of piecewise linear approximations of x which satisfies

$$\sup_{n\geq 1} |x_n|_{1-var;[0,1]} \leq |x|_{1-var;[0,1]} < \infty$$

and converges uniformly to x on $[0,1]$. We now consider the signatures $g_n := S_2(x_n)_{0,1}$ of these piecewise linear paths. As seen before, we have

$$g_n \in \left\langle \exp\left(\mathbb{R}^d\right)\right\rangle$$

for all $n \in \mathbb{N}$. In addition,

$$\lim_{n\to\infty} g_n = g$$

holds by Proposition 2.34. Consequently, $g \in \overline{\langle \exp\left(\mathbb{R}^d\right)\rangle}$.

Step 4: $\left\langle \exp\left(\mathbb{R}^d\right)\right\rangle$ is a closed Lie subgroup of $(1 + t^2\left(\mathbb{R}^d\right), \otimes)$.

We have shown topological closedness in Step 2. The fact that $\left\langle \exp\left(\mathbb{R}^d\right)\right\rangle$ is an (abstract) subgroup of $1 + t^2\left(\mathbb{R}^d\right)$ was already discussed in Remark 2.25. The claim can now be obtained by applying a theorem from the theory of Lie groups which states that each closed abstract subgroup of a Lie group is also a Lie subgroup (see for example [10], Th. 2.9). □

2.6 The Carnot-Carathéodory norm and metric

The Carnot-Carathéodory norm and homogenous norms on $G^2\left(\mathbb{R}^d\right)$

The Carnot-Carathéodory norm of $g \in G^2\left(\mathbb{R}^d\right)$ is defined by using paths in $\mathbb{R}^d$ having signature g.

Definition 2.35. For $g \in G^2\left(\mathbb{R}^d\right)$, we define

$$\|g\| := \inf\left\{\int_0^1 |d\gamma| : \gamma \in C^{1-var}\left([0,1], \mathbb{R}^d\right) \text{ and } S_2(\gamma)_{0,1} = g\right\}.$$

Remark 2.36. $\int_0^1 |d\gamma|$ is the length of the path γ, measured in the Euclidean distance on $\mathbb{R}^d$.

Theorem 2.37 (existence of a minimizing path, [5], Th. 7.32)**.** *For every $g \in G^2\left(\mathbb{R}^d\right)$, $\|g\|$ is finite and achieved at some minimizing path γ^* i.e. there exists $\gamma^* \in C^{1-var}\left([0,1], \mathbb{R}^d\right)$ such that*

$$\|g\| = \int_0^1 |d\gamma^*| \text{ and } S_2\left(\gamma^*\right)_{0,1} = g.$$

The minimizing path γ^* can be parametrized to be Lipschitz continuous and of constant speed i.e. there exists a constant $c \geq 0$ such that $\left|\dot{\gamma}^*(t)\right| = c$ for almost every $t \in [0,1]$.

Remark 2.38. Observe that combining Chow's theorem (Theorem 2.30) with the fact $G^2\left(\mathbb{R}^d\right) = \exp\left(g^2\left(\mathbb{R}^d\right)\right)$ gives that for any $g \in G^2\left(\mathbb{R}^d\right)$, there exists a piecewise linear path $[0,1] \to \mathbb{R}^d$ with signature g. Thus, it is clear that $\|g\|$ is finite for each $g \in G^2\left(\mathbb{R}^d\right)$. For a proof of the remaining statements in Theorem 2.37, we refer the reader to [5].

We will now introduce the concept of a homogenous norm on $G^2\left(\mathbb{R}^d\right)$ and show that it applies to the map $\|.\| : G^2\left(\mathbb{R}^d\right) \to \mathbb{R}_{\geq 0}$ defined above. Note that the Lie group $G^2\left(\mathbb{R}^d\right)$ is not a vector space, but we can define the following "substitute of scalar multiplication" on the underlying tensor space $T^2\left(\mathbb{R}^d\right)$.

Definition 2.39. For $r \in \mathbb{R}$, we define the **dilation map**

$$\begin{aligned} \delta_r : T^2\left(\mathbb{R}^d\right) &\to T^2\left(\mathbb{R}^d\right) \\ (g_0, g_1, g_2) &\mapsto (g_0, rg_1, r^2 g_2). \end{aligned}$$

Definition 2.40. A **homogenous norm** on $G^2\left(\mathbb{R}^d\right)$ is a continuous[4] map $|||.||| : G^2\left(\mathbb{R}^d\right) \to \mathbb{R}_{\geq 0}$ having the following two properties:

(*i*) $|||g||| = 0 \Leftrightarrow g = \mathbf{1} \in G^2\left(\mathbb{R}^d\right)$;

(*ii*) Homogeneity with respect to the dilation map δ_r:
$|||\delta_r(g)||| = |r|\,|||g|||$ for all $r \in \mathbb{R}$.

Example 2.41. A simple example for a homogenous norm on $G^2\left(\mathbb{R}^d\right)$ is given by

$$|||g||| := \max\left\{|\pi_1(g)|, |\pi_2(g)|^{1/2}\right\}.$$

Proposition 2.42 (Properties of the map $\|.\|$)**.** *Let $g, h \in G^2\left(\mathbb{R}^d\right)$. Then, we have*

(*i*) $\|g\| = 0 \Leftrightarrow g = \mathbf{1}$*;*

(*ii*) *Homogeneity with respect to the dilation map δ_r:*
$\|\delta_r(g)\| = |r|\,\|g\|$ *for all $r \in \mathbb{R}$;*

(*iii*) *Symmetry:* $\|g\| = \left\|g^{-1}\right\|$*;*

(*iv*) *Sub-additivity:* $\|g \otimes h\| \leq \|g\| + \|h\|$*;*

(*v*) *The map $g \mapsto \|g\|$ is continuous.*

In particular, $\|.\|$ is a homogenous norm on $G^2\left(\mathbb{R}^d\right)$.

Definition 2.43. $\|.\|$ is called **Carnot-Carathéodory norm**.

[4]with respect to the manifold topology on $G^2\left(\mathbb{R}^d\right)$.

Proof of Proposition 2.42. For $g \in G^2\left(\mathbb{R}^d\right)$, we write γ_g^* for an arbitrary path of minimal length having signature g i.e. $\gamma_g^* \in C^{1-var}\left([0,1],\mathbb{R}^d\right)$ such that

$$\|g\| = \int_0^1 \left|d\gamma_g^*\right| \text{ and } S_2(\gamma_g^*)_{0,1} = g.$$

In the following, let $g, h \in G^2\left(\mathbb{R}^d\right)$.
(i): This is left to the reader.
(ii): Case $r = 0$: $\|\delta_0(g)\| = \|\mathbf{1}\| = 0 = 0\,\|g\|$.
Case $r \neq 0$: By definition of the signature, we have $S_2(r\gamma*_g)_{0,1} = \delta_r(g)$. Consequently,

$$\|\delta_r(g)\| \leq \int_0^1 \left|d(r\gamma_g^*)\right| = |r| \int_0^1 \left|d(\gamma_g^*)\right| = |r|\,\|g\|\,.$$

To obtain the opposite inequality, we replace r by $\frac{1}{r}$ and g by $\delta_r(g)$ and proceed as above.

(iii): For $\overleftarrow{\gamma_g^*} : [0,1] \to \mathbb{R}^d; t \mapsto \gamma_g^*(1-t)$, the reversal of the path γ_g^*, we have $S_2(\overleftarrow{\gamma*_g})_{0,1} = g^{-1}$ ([5], Prop. 7.12). This implies

$$\left\|g^{-1}\right\| \leq \int_0^1 \left|d(\overleftarrow{\gamma_g^*})\right| = \int_0^1 \left|d(\gamma_g^*)\right| = \|g\|\,.$$

For the proof of the opposite inequality, we just replace g by g^{-1}.

(iv): Let $\gamma_{g,h}^* := \gamma_g^* \sqcup \gamma_h^*$ be the concatenation of the paths γ_g^* and γ_h^* (reparametrized to the interval $[0,1]$). Chen's theorem implies that

$$S_2(\gamma_{g,h}^*)_{0,1} = S_2(\gamma_g^*)_{0,1} \otimes S_2(\gamma_h^*)_{0,1} = g \otimes h.$$

It follows that

$$\|g \otimes h\| \leq \int_0^1 \left|d(\gamma_{g,h}^*)\right| = \|g\| + \|h\|\,.$$

(v): We consider $G^2\left(\mathbb{R}^d\right)$ with the metric $\widetilde{\rho} := \rho|_{G^2\left(\mathbb{R}^d\right)}$, which induces the manifold topology on $G^2\left(\mathbb{R}^d\right)$. Since this metric space satisfies the first axiom of countability, it is enough to prove sequential continuity.
For this, let $g \in G^2\left(\mathbb{R}^d\right)$ and let $(g_n)_{n\geq 1}$ be a sequence in $G^2\left(\mathbb{R}^d\right)$ satisfying

$$\lim_{n\to\infty} \widetilde{\rho}\,(g_n, g) = 0. \tag{2.6}$$

The aim is to show that

$$\lim_{n\to\infty} |\|g_n\| - \|g\|| = 0. \tag{2.7}$$

As the group operations $\otimes$ and $(.)^{-1}$ (which are polynomial functions in the coordinates) are continuous in the metric $\widetilde{\rho}$, (2.6) implies

$$\lim_{n\to\infty} \widetilde{\rho}\left(g_n^{-1} \otimes g, \mathbf{1}\right) = 0.$$

Hence, it follows from Corollary 2.31 that

$$\lim_{n\to\infty} \left\| g_n^{-1} \otimes g \right\| = 0. \tag{2.8}$$

Now, using sub-additivity and symmetry of the Carnot-Carathéodory norm, we get

- $\|g\| - \|g_n\| \leq \left\| g_n^{-1} \otimes g \right\|$
- $\|g_n\| = \left\| g_n^{-1} \right\| \leq \left\| g_n^{-1} \otimes g \right\| + \left\| g^{-1} \right\| = \left\| g_n^{-1} \otimes g \right\| + \|g\|$
  $\Rightarrow \|g_n\| - \|g\| \leq \left\| g_n^{-1} \otimes g \right\|$.

Thus,

$$|\|g_n\| - \|g\|| \leq \left\| g_n^{-1} \otimes g \right\|. \tag{2.9}$$

Combining (2.9) and (2.8) gives (2.7). □

The next result will be used in the proof of the Breuillard-Friz-Huesmann invariance principle.

Theorem 2.44. *All homogenous norms on $G^2(\mathbb{R}^d)$ are equivalent i.e. if $\|.\|_1$ and $\|.\|_2$ are two homogenous norms on $G^2(\mathbb{R}^d)$, then there exists a constant $C \geq 1$ such that for all $g \in G^2(\mathbb{R}^d)$*

$$\frac{1}{C} \|g\|_1 \leq \|g\|_2 \leq C \|g\|_1 . \tag{2.10}$$

Proof. Let $\|.\|_1$ and $\|.\|_2$ be two homogenous norms on $G^2(\mathbb{R}^d)$. It follows easily from the definition of equivalent norms that it suffices to consider one specific homogenous norm $\|.\|_1$ on $G^2(\mathbb{R}^d)$. Let us take

$$\|g\|_1 = \max\left\{ |\pi_1(g)|, |\pi_2(g)|^{1/2} \right\}.$$

We consider the set $B := \left\{ g \in G^2(\mathbb{R}^d) : \|g\|_1 = 1 \right\}$. If we view B as a subset of the normed vector space $(T^2(\mathbb{R}^d), |.|_{T^2(\mathbb{R}^d)})$, then it is clear that B is compact as it is bounded and closed. Since the manifold topology of $G^2(\mathbb{R}^d)$ is induced by the metric $\rho|_{G^2(\mathbb{R}^d)}$, which is in turn induced by $|.|_{T^2(\mathbb{R}^d)}$, B is also compact in $G^2(\mathbb{R}^d)$. As $\|.\|_2$ is continuous, it attains a maximum and a minimum on B. This minimum is positive, because $\|.\|_2$ satisfies property (i) of the definition of a homogenous norm (Definition 2.40). Hence, there exist two positive constants m and M such that for all $g \in B$

$$m \leq \|g\|_2 \leq M. \tag{2.11}$$

For $g = \mathbf{1}$, it is clear that (2.10) holds (All terms are 0 in this case.). Thus, we can assume $g \neq \mathbf{1}$, so that $\|g\|_1 > 0$. We set $\epsilon = \frac{1}{\|g\|_1}$. Then, we have $\delta_\epsilon(g) \in B$. Thus, by inequality (2.11), we have that for all $g \in G^2(\mathbb{R}^d) \setminus \{\mathbf{1}\}$

$$m \leq \|\delta_\epsilon(g)\|_2 = \frac{\|g\|_2}{\|g\|_1} \leq M.$$

This implies the claim. □

The Carnot-Carathéodory metric and geodesics in $G^2\left(\mathbb{R}^d\right)$

We now use the Carnot-Carathéodory norm to define a metric on $G^2\left(\mathbb{R}^d\right)$.

Proposition 2.45. *The map*

$$\begin{aligned} d : G^2\left(\mathbb{R}^d\right) \times G^2\left(\mathbb{R}^d\right) &\to \mathbb{R}_{\geq 0} \\ (g,h) &\mapsto \left\|g^{-1} \otimes h\right\| \end{aligned}$$

is a metric on $G^2\left(\mathbb{R}^d\right)$. Moreover, d is left-invariant[5] *and continuous.*

Definition 2.46. d is called **Carnot-Carathéodory metric**.

We omit the proof of Proposition 2.45 as the claims follow easily from the definition of d and the properties of the Carnot-Carathéodory norm stated in Proposition 2.42. We will use that

Proposition 2.47. $(G^2\left(\mathbb{R}^d\right), d)$ *is a Polish space.*

For a proof of this fact, see [5] (p. 153).

For the next result, we will need the notion of geodesics in a metric space.

Definition 2.48. Let (E,d) be a metric space and $g,h \in E$. A **geodesic** joining g and h is a continuous path $Y^{g,h} : [0,1] \to E$ such that $Y^{g,h}(0) = g$, $Y^{g,h}(1) = h$ and

$$d\left(Y^{g,h}(s), Y^{g,h}(t)\right) = |t-s|\, d(g,h) \tag{2.12}$$

for all $s,t \in [0,1]$, $s < t$.
E is called a **geodesic space** if any two points in E can be joined by a (not necessarily unique) geodesic.

Thus, a geodesic is a path which is a shortest connection between any two of its points. In $\mathbb{R}^d$, the geodesics are straight lines.
The following fact will be fundamental for interpolating random variables taking values in $G^2\left(\mathbb{R}^d\right)$.

Theorem 2.49. $(G^2\left(\mathbb{R}^d\right), d)$ *is a geodesic space. For $g,h \in G^2\left(\mathbb{R}^d\right)$, a connecting geodesic is given by*

$$t \in [0,1] \mapsto Y_t := g \otimes S_2(\gamma*)_{0,t},$$

where $\gamma*$ is a minimizing path associated to $g^{-1} \otimes h$ (i.e. $\int_0^1 |d\gamma*| = \left\|g^{-1} \otimes h\right\|$ and $S_2(\gamma*)_{0,1} = g^{-1} \otimes h$).

[5] i.e. $d(g \otimes h, g \otimes k) = d(h,k)$ for all $g,h,k \in G^2\left(\mathbb{R}^d\right)$.

Proof. Y is continuous since the map $t \mapsto S_2(\gamma*)_{0,t}$ is continuous. It is clear that $Y_0 = g$ and $Y_1 = h$.
By Theorem 2.37, we can assume that $\gamma*$ is parametrized to be Lipschitz continuous and satisfies

$$|\dot{\gamma*}_u| = \text{const. for almost all } u \in [0,1]. \tag{2.13}$$

Then, for $0 \leq s < t \leq 1$, we have

$$\begin{aligned}
d(Y_s, Y_t) &= \left\| (g \otimes S_2(\gamma*)_{0,s})^{-1} \otimes (g \otimes S_2(\gamma*)_{0,t}) \right\| \\
&= \| S_2(\gamma*)_{s,t} \| && (2.14) \\
&\leq \int_s^t |d\gamma*| && (2.15) \\
&= \int_s^t |\dot{\gamma*}_u| \, du && (2.16) \\
&= (t-s) \int_0^1 |\dot{\gamma*}_u| \, du && (2.17) \\
&= (t-s) \int_0^1 |d\gamma*| && (2.18) \\
&= (t-s) \left\| g^{-1} \otimes h \right\| \\
&= (t-s) \; d(g,h).
\end{aligned}$$

(2.14) is a consequence of Chen's theorem. (2.15) follows from the definition of the Carnot-Carathéodory norm by reparametrization. In (2.16) and (2.18), we have used that $\gamma*$ is Lipschitz continuous and hence absolutely continuous. (2.17) is due to the fact that $\gamma*$ is almost everywhere of constant speed (assumption (2.13)).
If we assume the inequality in (2.15) to be strict, we will obtain
$d(Y_s, Y_t) < (t-s) d(g,h)$. This implies

$$\begin{aligned}
d(g,h) &= d(Y_0, Y_1) \\
&\leq d(Y_0, Y_s) + d(Y_s, Y_t) + d(Y_t, Y_1) \\
&< (s + (t-s) + (1-t)) \; d(g,h) \\
&= d(g,h),
\end{aligned}$$

which is a contradiction. Consequently, there is equality in (2.15) so that we have $d(Y_s, Y_t) = (t-s) \; d(g,h)$.
Hence, Y is a geodesic connecting g and h. □

2.7 Geometric α-Hölder rough paths

To introduce the notion of geometric α-Hölder rough paths, we need a Hölder distance for paths in $G^2\left(\mathbb{R}^d\right)$. Such a distance induces a topology which is stronger than the uniform one. First, we need some preparatory definitions.
In the whole section, we assume $\alpha \in (0,1]$ if nothing else is mentioned.

Definition 2.50. We define the **increments** of a path $x : [0,1] \to G^2(\mathbb{R}^d)$ by

$$x_{s,t} := x_s^{-1} \otimes x_t \in G^2\left(\mathbb{R}^d\right)$$

for $0 \leq s \leq t \leq 1$.

Definition 2.51. For a path $x : [0,1] \to G^2(\mathbb{R}^d)$, the **$\alpha$-Hölder norm** is defined as

$$\|x\|_{\alpha-H\ddot{o}l} := \sup_{0 \leq s < t \leq 1} \frac{\|x_{s,t}\|}{|t-s|^\alpha} = \sup_{0 \leq s < t \leq 1} \frac{d(x_s, x_t)}{|t-s|^\alpha}.$$

Let $C^{\alpha-H\ddot{o}l}([0,1], G^2(\mathbb{R}^d)) := \left\{x \in C([0,1], G^2(\mathbb{R}^d)) : \|x\|_{\alpha-H\ddot{o}l} < \infty\right\}$.

Remark 2.52. $C^{1-H\ddot{o}l}([0,1], G^2(\mathbb{R}^d))$ is the space of Lipschitz continuous maps $[0,1] \to G^2(\mathbb{R}^d)$. Note that for $0 < \alpha < \alpha' \leq 1$, we have

$$C^{\alpha'-H\ddot{o}l}\left([0,1], G^2\left(\mathbb{R}^d\right)\right) \subset C^{\alpha-H\ddot{o}l}\left([0,1], G^2\left(\mathbb{R}^d\right)\right).$$

Definition 2.53. Let x, y be two paths in $C([0,1], G^2(\mathbb{R}^d))$. We define the **$\alpha$-Hölder distance** of x and y by

$$d_{\alpha-H\ddot{o}l}(x,y) := \sup_{0 \leq s < t \leq 1} \frac{d(x_{s,t}, y_{s,t})}{|t-s|^\alpha}.$$

On the set $C^{\alpha-H\ddot{o}l}([0,1], G^2(\mathbb{R}^d))$, $d_{\alpha-H\ddot{o}l}$ satisfies all properties of a metric except that $d_{\alpha-H\ddot{o}l}(x,y) = 0$ does not imply $x = y$, but only $x = c \otimes y$ for $c = x_0^{-1} \otimes y_0 \in G^2(\mathbb{R}^d)$. The distance

$$\widetilde{d}_{\alpha-H\ddot{o}l}(x,y) := d(x_0, y_0) + d_{\alpha-H\ddot{o}l}(x,y)$$

is a metric on $G^2(\mathbb{R}^d)$.

Lemma 2.54. *$\left(C^{\alpha-H\ddot{o}l}([0,1], G^2(\mathbb{R}^d)),\ \widetilde{d}_{\alpha-H\ddot{o}l}\right)$ is a complete metric space.*

$C^{\alpha-H\ddot{o}l}([0,1], G^2(\mathbb{R}^d))$ is not separable. For establishing invariance principles, it is advantageous to work with a separable state space. We will therefore focus on a separable subspace of $C^{\alpha-H\ddot{o}l}([0,1], G^2(\mathbb{R}^d))$, namely the $d_{\alpha-H\ddot{o}l}$-closure of the step-2 lifts of smooth paths $[0,1] \to \mathbb{R}^d$. Note that the starting point of any such step-2 lift is $\mathbf{1}$[6] . To allow for arbitrary starting points, we therefore have to use a transformation.

Definition 2.55. Let $C_1^{0,\alpha-H\ddot{o}l}([0,1], G^2(\mathbb{R}^d))$ be the set of continuous paths $x : [0,1] \to G^2(\mathbb{R}^d)$ for which there exists a sequence of smooth paths $x_n : [0,1] \to \mathbb{R}^d$ such that

$$\lim_{n \to \infty} d_{\alpha-H\ddot{o}l}(x, S_2(x_n)_{0,\,.}) = \lim_{n \to \infty} \sup_{0 \leq s < t \leq 1} \frac{d(x_{s,t}, S_2(x_n)_{s,t})}{|t-s|^\alpha} = 0.$$

By $C^{0,\alpha-H\ddot{o}l}([0,1], G^2(\mathbb{R}^d))$, we denote the set of paths x satisfying $x_{0,\,.} = x_0^{-1} x_. \in C_1^{0,\alpha-H\ddot{o}l}([0,1], G^2(\mathbb{R}^d))$.

[6] as $S_2(x)_{0,0} = \mathbf{1}$ for any path $x : [0,1] \to \mathbb{R}^d$.

Proposition 2.56. $C^{0,\alpha-Höl}\left([0,1], G^2\left(\mathbb{R}^d\right)\right)$ *is a Polish space for the metric* $\widetilde{d}_{\alpha-Höl}$.

As $C^{0,\alpha-Höl}\left([0,1], G^2\left(\mathbb{R}^d\right)\right)$ is a closed subspace of $C^{\alpha-Höl}\left([0,1], G^2\left(\mathbb{R}^d\right)\right)$, completeness follows immediately from completeness of $C^{\alpha-Höl}\left([0,1], G^2\left(\mathbb{R}^d\right)\right)$. The proof of separability is more involved (see [5]).

Remark 2.57. In the sequel, we will always endow $C^{\alpha-Höl}\left([0,1], G^2\left(\mathbb{R}^d\right)\right)$ and $C^{0,\alpha-Höl}\left([0,1], G^2\left(\mathbb{R}^d\right)\right)$ with the topology induced by the metric $\widetilde{d}_{\alpha-Höl}$. We call this topology **rough path topology**.

The next result is a useful characterization of the subspace $C^{0,\alpha-Höl}\left([0,1], G^2\left(\mathbb{R}^d\right)\right)$ of $C^{\alpha-Höl}\left([0,1], G^2\left(\mathbb{R}^d\right)\right)$ for $\alpha \neq 1$.

Theorem 2.58 (Wiener's characterization, [5], Th. 8.22)**.** *Let* $\alpha \in (0,1)$ *and* $x \in C^{\alpha-Höl}\left([0,1], G^2\left(\mathbb{R}^d\right)\right)$. *Then,* $x \in C^{0,\alpha-Höl}\left([0,1], G^2\left(\mathbb{R}^d\right)\right)$ *if and only if*

$$\lim_{\delta\to 0} \sup_{\substack{s,t\in[0,1] \\ 0<|t-s|<\delta}} \frac{d(x_s, x_t)}{|t-s|^{\alpha}} = 0.$$

We will distinguish two types of α-Hölder rough paths.

Definition 2.59. Let $\alpha \in (1/3, 1/2]$. Elements of $C^{\alpha-Höl}\left([0,1], G^2\left(\mathbb{R}^d\right)\right)$ are called **weak geometric α-Hölder rough paths** and elements of $C^{0,\alpha-Höl}\left([0,1], G^2\left(\mathbb{R}^d\right)\right)$ are called **geometric α-Hölder rough paths**.

We now give a condition which guarantees that the sample paths of a stochastic process with values in $G^2\left(\mathbb{R}^d\right)$ are a.s. weak geometric α-Hölder rough paths. A similar condition is used in Kolmogorov's continuity criterion, which ensures that a stochastic process with values in $\mathbb{R}^d$ has a continuous version. We will apply the following result to enhanced Brownian motion in the next section.

Theorem 2.60 ([5], Th. A.10)**.** *Let* $(X_t : t \in [0,1])$ *be a stochastic process with values in* $G^2\left(\mathbb{R}^d\right)$. *Assume there exist constants* $a, b, C > 0$, *satisfying* $a > 1 + b$, *such that*

$$E\left[d(X_s, X_t)^a\right] \le C\,|t-s|^{1+b}$$

for all $s, t \in [0,1]$.
Then, for any $\gamma \in (0, \frac{b}{a})$, *the sample paths of* X *are a.s. in* $C^{\gamma-Höl}\left([0,1], G^2\left(\mathbb{R}^d\right)\right)$.

2.8 Enhanced Brownian motion

This section deals with the rough path lift of Brownian motion to the Lie group $G^2\left(\mathbb{R}^d\right)$, called enhanced Brownian motion (EBM). We will see that EBM has several properties which are similar to the properties of Brownian motion. The notion of

Brownian motion can be generalized to Lie groups (see for example [15], p. 115f.) and EBM is in fact a Brownian motion on the Lie group $G^2\left(\mathbb{R}^d\right)$.
To simplify the formulation of certain results, we will define EBM on the time interval $[0,\infty)$, although we will only need the restriction to the time interval $[0,1]$ in the sequel.

Definition 2.61. Given a d-dimensional Brownian motion $B=(B^1,\dots,B^d)$, the d-dimensional **Lévy area** $A=(A^{ij}:i,j\in\{1,\dots,d\})$ is defined as the continuous version of the process

$$t\in[0,\infty)\mapsto A_t^{ij}=\frac{1}{2}\left(\int_0^t B_u^i\ dB_u^j-\int_0^t B_u^j\ dB_u^i\right).$$

Definition 2.62. Let B be a d-dimensional Brownian motion and A its Lévy area. Then, the continuous version of the $G^2\left(\mathbb{R}^d\right)$-valued process $\mathbf{B}$, defined by

$$\mathbf{B}_t:=\exp\left((0,B_t,A_t)\right)$$

for $t\geq 0$ is called **enhanced Brownian motion** (EBM).

We will establish that almost every sample path realization of enhanced Brownian motion is a weak geometric α-Hölder rough path for any $\alpha\in(1/3,1/2)$. In fact, the term "weak" can be omitted in the last sentence, but we will not prove this here (for a proof, see [5], Cor. 13.14).
First, we give some basic properties of enhanced Brownian motion. Some of these properties will be useful in the proof of rough path regularity.
Let us consider the increments of enhanced Brownian motion, Brownian motion and Lévy's area: For $0\leq s\leq t<\infty$,

$$\begin{aligned}
\mathbf{B}_{s,t}&=\mathbf{B}_s^{-1}\otimes\mathbf{B}_t;\\
B_{s,t}&:=B_t-B_s;\\
A_{s,t}^{ij}&:=A_t^{ij}-A_s^{ij}-\frac{1}{2}\left(B_s^iB_{s,t}^j-B_s^jB_{s,t}^i\right)\\
&=\frac{1}{2}\left(\int_s^t B_{s,r}^i\ dB_r^j-\int_s^t B_{s,r}^j\ dB_r^i\right)\text{ a.s..}
\end{aligned}$$

One can then check by an easy computation that

$$\mathbf{B}_{s,t}=\exp\left((0,B_{s,t},A_{s,t})\right).$$

Proposition 2.63 ([5], Prop. 13.11)**.** *Let* $\mathbf{B}$ *be an enhanced Brownian motion. Then, we have:*

(i) $\mathbf{B}_0(\omega)=\mathbf{1}$ *for all* ω.

(ii) For every $t,h\in[0,\infty)$, $\mathbf{B}_{t,t+h}$ *is independent of* $\sigma(\mathbf{B}_u:u\leq t)$.

(iii) $\mathbf{B}$ *has stationary increments, i.e. for every* $s \geq 0$

$$(\mathbf{B}_{s,s+t})_{t\geq 0} \stackrel{D}{=} (\mathbf{B}_t)_{t\geq 0}\,.$$

(i) is easy to see and (ii) is left to the reader (or see [5], p. 334-335).

Proof of (iii). Brownian motion has stationary increments, that is for each $s \geq 0$

$$(B_{s,s+t})_{t\geq 0} \stackrel{D}{=} (B_t)_{t\geq 0}\,.$$

Now, fix $s \geq 0$. Then, we have

$$\begin{aligned}
\left(A^{ij}_{s,s+t}\right)_{t\geq 0} &= \left(\frac{1}{2}\left(\int_s^{s+t} B^i_{s,r}\, dB^j_r - \int_s^{s+t} B^j_{s,r}\, dB^i_r\right)\right)_{t\geq 0} \\
&= \left(\frac{1}{2}\left(\int_s^{s+t} B^i_{s,r}\, dB^j_{s,r} - \int_s^{s+t} B^j_{s,r}\, dB^i_{s,r}\right)\right)_{t\geq 0} \\
&\stackrel{D}{=} \left(\frac{1}{2}\left(\int_0^{t} B^i_{r}\, dB^j_r - \int_0^{t} B^j_{r}\, dB^i_r\right)\right)_{t\geq 0} \\
&= \left(A^{ij}_t\right)_{t\geq 0}\,.
\end{aligned}$$

Consequently,

$$(B_{s,s+t}, A_{s,s+t})_{t\geq 0} \stackrel{D}{=} (B_t, A_t)_{t\geq 0}\,.$$

Now, the result follows from the definition of enhanced Brownian motion and the fact that the map $\exp : g^2\left(\mathbb{R}^2\right) \to G^2\left(\mathbb{R}^d\right)$ is continuous and thus measurable. $\square$

It is not hard to deduce the following scaling result for EBM from the scaling of Brownian motion (i.e. $(B_{\lambda^2 t})_{t\geq 0} \stackrel{D}{=} (\lambda B_t)_{t\geq 0}$ for $\lambda > 0$).

Lemma 2.64 (Scaling of EBM, [5], Lem. 13.12)**.** *Let* $\mathbf{B}$ *be an enhanced Brownian motion. Then, for every* $\lambda > 0$*, we have*

$$(\mathbf{B}_{\lambda^2 t})_{t\geq 0} \stackrel{D}{=} (\delta_\lambda(\mathbf{B}_t))_{t\geq 0}\,.$$

We now turn to the proof of rough path regularity.

Proposition 2.65. *Let* $\mathbf{B}$ *be an enhanced Brownian motion defined on the time interval* $[0,1]$*. Then, for any* $\alpha \in (1/3, 1/2)$*,* $\mathbf{B}_{\cdot}(\omega)$ *is a weak geometric* α*-Hölder rough path almost surely.*

Proof. Let $s, t \in [0,1]$, $s < t$. Since EBM has stationary increments and scales as stated in Lemma 2.64, we have

$$\mathbf{B}_{s,t} \stackrel{D}{=} \mathbf{B}_{0,t-s} \stackrel{D}{=} \delta_{\sqrt{t-s}}(\mathbf{B}_{0,1}) = \delta_{\sqrt{t-s}}(\mathbf{B}_1)\,.$$

Consequently, for all $q \in \mathbb{N}$ and all $s, t \in [0,1]$, $s < t$,

$$E\left[d\left(\mathbf{B}_s, \mathbf{B}_t\right)^{2q}\right] = E\left[\left\|\mathbf{B}_{s,t}\right\|^{2q}\right] = E\left[\left\|\mathbf{B}_1\right\|^{2q}\right] |t-s|^q . \tag{2.19}$$

Now, given any $\alpha \in (0, 1/2)$, there exists $q \in \mathbb{N}$ such that $\frac{q-1}{2q} > \alpha$, so that by (2.19) and Theorem 2.60, we have $\mathbf{B}_{\cdot}(\omega) \in C^{\alpha-H\ddot{o}l}\left([0,1], G^2\left(\mathbb{R}^d\right)\right)$ almost surely. $\square$

3 Weak convergence and tightness

3.1 Definitions and basic results

The reference for this section is [1].
Let (S, d) be a metric space and $\mathcal{S}$ the Borel σ-algebra on S.
If P_n, $n \in \mathbb{N}$ and P are probability measures on $(S, \mathcal{S})$ and P_n converges weakly to P, we write $\text{w-lim}_{n\to\infty} P_n = P$.
In the sequel, let Π denote a family of probability measures on $(S, \mathcal{S})$.

Definition 3.1. Π is **tight** if for every $\epsilon > 0$ there exists a compact set $K_\epsilon \subset S$ such that $P(K_\epsilon) > 1 - \epsilon$ for all $P \in \Pi$.

Definition 3.2. Π is **relatively compact** if every sequence of elements of Π contains a weakly convergent subsequence i.e. for every sequence $(P_n)_{n\geq 1}$, $P_n \in \Pi$ for all n, there exists a subsequence $(P_{n_i})_{i\geq 1}$ and a probability measure Q[1] defined on $(S, \mathcal{S})$ such that $\text{w-lim}_{i\to\infty} P_{n_i} = Q$.

Theorem 3.3 (Prohorov's theorem)**.** *If Π is tight, then it is relatively compact.*

Remark 3.4. If the metric space S is separable and complete, then the converse is also true. This result will not be used in the sequel.

Definition 3.5. For $m \in \mathbb{N}$ and $t_1, \dots, t_m \in [0, 1]$, we define the map

$$\begin{aligned} \pi_{t_1,\dots,t_m} : C([0,1], S) &\to S^m \\ x &\mapsto (x(t_1), \dots, x(t_m)). \end{aligned}$$

If P is a probability measure on $C([0, 1], S)$, we call the image measures of the form $P(\pi_{t_1,\dots,t_m})^{-1}$ its **finite-dimensional distributions**.

We now turn to some results related to weak convergence. The next result says in particular that the limit measure of a weakly convergent sequence of probability measures is unique.

Proposition 3.6. *Let P and Q be two probability measures on $(S, \mathcal{S})$. If*

$$\int f \; dP = \int f \; dQ$$

for every bounded, continuous function $f : S \to \mathbb{R}$, then $P = Q$.

[1] not necessarily an element of Π.

The fact that weak convergence is preserved under continuous maps will be used several times in the subsequent proofs.
Let S' be a metric space with Borel σ-algebra $\mathcal{S}'$.

Theorem 3.7 (Continuous mapping theorem)**.** *Let P_n, $n \in \mathbb{N}$ and P be probability measures satisfying* $\text{w-lim}_{n\to\infty} P_n = P$. *Let $h : S \to S'$ be a continuous map. Then,* $\text{w-lim}_{n\to\infty} P_n h^{-1} = Ph^{-1}$.

3.2 Proof method

In this section, we prove the basic weak convergence result on which the proof of the Breuillard-Friz-Huesmann invariance principle relies.
Let $\alpha \in (0,1]$ and let $\mathcal{B}_\alpha$ be the Borel σ-algebra on $C^{0,\alpha-Höl}\left([0,1], G^2\left(\mathbb{R}^d\right)\right)$ with respect to the α-Hölder metric $\widetilde{d}_{\alpha-\text{Höl}}$.

Theorem 3.8. *Let P_n, $n \in \mathbb{N}$, and P be probability measures on $\left(C^{0,\alpha-Höl}\left([0,1], G^2\left(\mathbb{R}^d\right)\right), \mathcal{B}_\alpha\right)$. If the finite-dimensional distributions of P_n converge weakly to those of P and if $(P_n)_{n\geq 1}$ is tight, then P_n converges weakly to P.*

The argumentation in the proof below is adapted from [1], where the corresponding result is established for the space $C\left([0,1], \mathbb{R}\right)$ with uniform topology. We will use the following result, which will be proven at the end of this section.

Lemma 3.9. *A probability measure on $\left(C^{0,\alpha-Höl}\left([0,1], G^2\left(\mathbb{R}^d\right)\right), \mathcal{B}_\alpha\right)$ is completely determined by its finite-dimensional distributions.*

Proof of Theorem 3.8. By Prohorov's theorem, $(P_n)_{n\geq 1}$ is relatively compact i.e. every subsequence $(P_{n_i})_{i\geq 1}$ contains another subsequence $(P_{n_{i_k}})_{k\geq 1}$ converging weakly to some probability measure Q on $C^{0,\alpha-Höl}\left([0,1], G^2\left(\mathbb{R}^d\right)\right)$.
Let $m \in \mathbb{N}$ and $t_1, \ldots, t_m \in [0,1]$. Since the map $\pi_{t_1,\ldots,t_m}$ defined in the previous section (Definition 3.5) is continuous, the continuous mapping theorem (Theorem 3.7) implies

$$\underset{k\to\infty}{\text{w-lim}}\, P_{n_{i_k}} \pi^{-1}_{t_1,\ldots,t_m} = Q\pi^{-1}_{t_1,\ldots,t_m}.$$

As the finite-dimensional distributions of P_n converge weakly to those of P, we have

$$\underset{n\to\infty}{\text{w-lim}}\, P_n \pi^{-1}_{t_1,\ldots,t_m} = P\pi^{-1}_{t_1,\ldots,t_m}.$$

Thus, by Proposition 3.6,

$$P\pi^{-1}_{t_1,\ldots,t_m} = Q\pi^{-1}_{t_1,\ldots,t_m}$$

for all $m \in \mathbb{N}$ and all $t_1, \ldots, t_m \in [0,1]$ i.e. P and Q have the same finite-dimensional distributions. By Lemma 3.9, this implies $P = Q$. Thus, every subsequence $(P_{n_i})_{i\geq 1}$ contains another subsequence $(P_{n_{i_k}})_{k\geq 1}$ satisfying $\text{w-lim}_{k\to\infty} P_{n_{i_k}} = P$. Consequently, the whole sequence $(P_n)_{n\geq 1}$ converges weakly to P. □

Proof of Lemma 3.9. Let $\mathcal{C}$ denote the set of cylinder sets in $C^{0,\alpha-H\ddot{o}l}\left([0,1],G^2\left(\mathbb{R}^d\right)\right)$ i.e.

$$\mathcal{C} := \left\{\pi^{-1}_{i_1,\dots,i_m}(B) : B \text{ is a Borel set in } \left(G^2\left(\mathbb{R}^d\right)\right)^m, i_1,\dots,i_m \in [0,1], m \in \mathbb{N}\right\}.$$

$\mathcal{C}$ is stable under intersections, so that it suffices to prove $\sigma(\mathcal{C}) = \mathcal{B}_\alpha$.
$\sigma(\mathcal{C}) \subset \mathcal{B}_\alpha$ is clear since the maps of the form $\pi_{t_1,\dots,t_m}$ are continuous. For the other inclusion, note that the separable metric space $\left(C^{0,\alpha-H\ddot{o}l}\left([0,1],G^2\left(\mathbb{R}^d\right)\right),\widetilde{d}_\alpha\right)$ is second-countable. Hence, it is enough to show that an arbitrary open ball in $\mathcal{B}_\alpha$ is an element of $\sigma(\mathcal{C})$. For this, we will use the fact that $G^2\left(\mathbb{R}^d\right)$ is separable. Let $\{g_i : i \in \mathbb{N}\}$ be a dense subset of $G^2\left(\mathbb{R}^d\right)$. For $x \in C^{0,\alpha-H\ddot{o}l}\left([0,1],G^2\left(\mathbb{R}^d\right)\right) =: C^{0,\alpha}$ and $\epsilon > 0$, we then have

$$\begin{aligned}
B_\epsilon(x) = {} & \left\{y \in C^{0,\alpha} : d(x_0,y_0) + d_{\alpha-H\ddot{o}l}(x,y) < \epsilon\right\} \\
= {} & \bigcup_{\substack{(p,q)\in\mathbb{Q}_0^+\times\mathbb{Q}_0^+,\\ p+q<\epsilon}} \left\{y \in C^{0,\alpha} : d(x_0,y_0) \le p \ \wedge \sup_{0\le s<t\le 1} \frac{d(x_{s,t},y_{s,t})}{|t-s|^\alpha} \le q\right\} \\
= {} & \bigcup_{\substack{(p,q)\in\mathbb{Q}_0^+\times\mathbb{Q}_0^+,\\ p+q<\epsilon}} \ \bigcap_{\substack{s,t\in\mathbb{Q}\cap[0,1]\\ s<t}} \left\{y \in C^{0,\alpha} : d(x_0,y_0) \le p \ \wedge \ d(x_{s,t},y_{s,t}) \le \underbrace{q\,|t-s|^\alpha}_{=:q_{s,t}}\right\} \qquad (3.1) \\
= {} & \bigcup_{\substack{(p,q)\in\mathbb{Q}_0^+\times\mathbb{Q}_0^+,\\ p+q<\epsilon}} \ \bigcap_{\substack{s,t\in\mathbb{Q}\cap[0,1]\\ s<t}} \ \bigcup_{i\in\mathbb{N}} \bigcap_{m\in\mathbb{N}} \Big\{y \in C^{0,\alpha} : y_0 \in \overline{B_p(x_0)} \ \wedge \ y_t \in \overline{B_{1/m}(g_i)} \\
& \wedge \ y_s \otimes x_{s,t} \in \overline{B_{q_{s,t}+1/m}(g_i)}\Big\} \in \sigma(\mathcal{C}). \qquad (3.2)
\end{aligned}$$

(3.1) is due to the fact that paths in $C^{0,\alpha}$ are continuous. To obtain (3.2), we have used that by left-invariance of the metric d, we have $d(x_{s,t},y_{s,t}) = d(y_s \otimes x_{s,t}, y_t)$ and that tensor multiplication by $x_{s,t}$ is continuous and hence measurable. This completes the proof. □

4 Breuillard-Friz-Huesmann (BFH) invariance principle for independent random variables with values in $G^2(\mathbb{R}^d)$

4.1 Results

We will need the following

Definition 4.1. A $G^2(\mathbb{R}^d)$-valued random variable $\mathbf{X}$ is called centered if the $\mathbb{R}^d$-valued random variable $\pi_1(\mathbf{X})$ is centered i.e. $E[\pi_1(\mathbf{X})] = 0 \in \mathbb{R}^d$.

Theorem 4.2 (Breuillard-Friz-Huesmann theorem, BFH theorem for short, [3])**.** *Let $(\mathbf{X}_i)_{i\geq 1}$ be a sequence of iid $G^2(\mathbb{R}^d)$-valued random variables which are centered and have finite moments of all orders (i.e. $\forall q \in [1,\infty) : E[\|\mathbf{X}_i\|^q] < \infty$). Consider the rescaled random walk $\mathbf{W}^{(n)} = \{\mathbf{W}^{(n)}_t\}_{t\in[0,1]}$ defined by $\mathbf{W}^{(n)}_0 = \mathbf{1}$,*

$$\mathbf{W}^{(n)}_t = \delta_{n^{-1/2}}(\mathbf{X}_1 \otimes \cdots \otimes \mathbf{X}_{nt})$$

for $t \in \left\{\frac{1}{n}, \frac{2}{n}, \ldots, 1\right\}$ and piecewise geodesically connected in between (i.e. $\mathbf{W}^{(n)}|_{[\frac{i}{n},\frac{i+1}{n}]}$ is a geodesic connecting $\mathbf{W}^{(n)}_{i/n}$ and $\mathbf{W}^{(n)}_{(i+1)/n}$).
Then, for any $\alpha \in (0, 1/2)$, $\mathbf{W}^{(n)}$ converges weakly to an enhanced Brownian motion $\mathbf{B}$ in $C^{0,\alpha-H\ddot{o}l}\left([0,1], G^2(\mathbb{R}^d)\right)$ i.e. for any bounded, continuous function $f : C^{0,\alpha-H\ddot{o}l}\left([0,1], G^2(\mathbb{R}^d)\right) \to \mathbb{R}$, we have

$$\lim_{n\to\infty} E\left[f\left(\mathbf{W}^{(n)}\right)\right] = E[f(\mathbf{B})].$$

In particular, the above theorem implies

Corollary 4.3 (Donsker's theorem for enhanced Brownian motion, [3])**.** *Let $(X_i)_{i\geq 1}$ be a sequence of iid $\mathbb{R}^d$-valued random variables which are centered and have finite moments of all orders (i.e. $\forall q \in [1,\infty) : E[|X_i|^q] < \infty$). Let $S_0 = 0$ and $S_n = \sum_{i=1}^n X_i$ for $n \in \mathbb{N}$ and consider the rescaled and piecewise linearly connected random walk $Y^{(n)} = \{Y^{(n)}_t\}_{t\in[0,1]}$ defined by*

$$Y^{(n)}_t = \frac{1}{\sqrt{n}}(S_{\lfloor nt\rfloor} + (nt - \lfloor nt\rfloor)X_{\lfloor nt\rfloor+1}).$$

Then, for any $\alpha \in (0, 1/2)$, $S_2(Y^{(n)})_{0,\cdot}$ converges weakly to an enhanced Brownian motion in $C^{0,\alpha-H\ddot{o}l}\left([0,1], G^2(\mathbb{R}^d)\right)$.

Proof of Corollary 4.3. By Chen's theorem (Theorem 2.29), the signature of a piecewise linear path in $\mathbb{R}^d$ equals the tensor product of the signatures of the linear pieces. From Example 2.18, we know that the signature of a linear path in $\mathbb{R}^d$ with slope $a \in \mathbb{R}^d$ is $\exp(a)$. Consequently, we have

$$S_2(Y^{(n)})_{0,t} = \begin{cases} \delta_{n^{-1/2}} \left(\exp\left(ntX_1\right)\right) & \text{for } t \in [0, \frac{1}{n}] \\ \delta_{n^{-1/2}} \left(\bigotimes_{i=1}^{\lfloor nt \rfloor} \exp(X_i) \otimes \exp\left((nt - \lfloor nt \rfloor) X_{\lfloor nt \rfloor + 1}\right)\right) & \text{for } t \in (\frac{1}{n}, 1]. \end{cases}$$

Now, we set $\mathbf{X}_i := \exp(X_i)$ for $i \in \mathbb{N}$. Then, $(\mathbf{X}_i)_{i \geq 1}$ is a sequence of centered[1] iid random variables with values in $G^2\left(\mathbb{R}^d\right)$.
In addition, $\mathbf{X}_i$ has finite moments of all orders: For each $\omega \in \Omega$, the shortest path in $\mathbb{R}^d$ with signature $\mathbf{X}_i(\omega)$ is a linear path with slope $X_i(\omega)$. Hence, by the definition of the Carnot-Carathéodory norm, we have

$$E\left[\|\mathbf{X}_i\|^q\right] = E\left[|X_i|^q\right] < \infty$$

for all $q \in [1, \infty)$.
For $a \in \mathbb{R}^d$ and $g \in G^2\left(\mathbb{R}^d\right)$, one can check that the path $t \in [0,1] \mapsto g \otimes \exp(ta)$ satisfies property (2.12) and thus is a geodesic connecting g and $g \otimes \exp(a)$. Hence, the stochastic process $S_2(Y^{(n)})_{0,\cdot}$ is geodesically interpolated between the time points $\frac{k}{n}$ and $\frac{k+1}{n}$, $k = 0, \ldots, n-1$.
Consequently, we can apply the BFH theorem to the sequence $(\mathbf{X}_i)_{i \geq 1}$ to obtain the conclusion of the corollary. $\square$

4.2 Proof

The first thing to do is showing that the sample paths of $\mathbf{W}^{(n)}$ are a.s. in $C^{0,\alpha-H\ddot{o}l}\left([0,1], G^2\left(\mathbb{R}^d\right)\right)$. It is easy to see that the sample paths of $\mathbf{W}^{(n)}$ are a.s. in $C^{\alpha-H\ddot{o}l}\left([0,1], G^2\left(\mathbb{R}^d\right)\right)$. Then, one can use Wiener's characterization (Theorem 2.58) to obtain the desired result. The details are left to the reader.
By Theorem 3.8, it is enough to prove that the finite-dimensional distributions of $(P(\mathbf{W}^{(n)})^{-1})_{n \geq 1}$ converge weakly to those of $P\mathbf{B}^{-1}$ and that the sequence $(P(\mathbf{W}^{(n)})^{-1})_{n \geq 1}$ is tight in $C^{0,\alpha-H\ddot{o}l}\left([0,1], G^2\left(\mathbb{R}^d\right)\right)$ for any $\alpha \in (0, 1/2)$.

Remark 4.4. In the sequel, we will just write "the sequence $(\mathbf{W}^{(n)})_{n \geq 1}$ is tight in $C^{0,\alpha-H\ddot{o}l}\left([0,1], G^2\left(\mathbb{R}^d\right)\right)$" for the latter fact.

[1] as $E[\pi_1(\mathbf{X}_i)] = E[X_i] = 0$.

4.2.1 Step 1: Convergence of the finite-dimensional distributions

We will use the following properties of $G^2\left(\mathbb{R}^d\right)$.

Lemma 4.5. *The Lie group $G^2\left(\mathbb{R}^d\right)$ is simply connected and nilpotent.*

Proof. $G^2\left(\mathbb{R}^d\right)$ is simply connected:
We endow the vector space $g^2\left(\mathbb{R}^d\right)$ with the topology induced by the norm $|.|_{T^2\left(\mathbb{R}^d\right)}$. Then, the map $\exp|_{g^2\left(\mathbb{R}^d\right)} : g^2\left(\mathbb{R}^d\right) \to G^2\left(\mathbb{R}^d\right)$ is a homeomorphism. Hence, the claim follows from the fact that the topological vector space $g^2\left(\mathbb{R}^d\right)$ is contractible and therefore simply connected.
$G^2\left(\mathbb{R}^d\right)$ is nilpotent[2]:
Let $U := \left\{g \in G^2\left(\mathbb{R}^d\right) : \pi_1(g) = 0 \ \wedge \ \pi_2(g) \text{ is skew-symmetric}\right\}$. It is easy to check that $\left[G^2\left(\mathbb{R}^d\right), G^2\left(\mathbb{R}^d\right)\right] = U$ and $\left[G^2\left(\mathbb{R}^d\right), U\right] = \{\mathbf{1}\}$. Thus, the descending central series of $G^2\left(\mathbb{R}^d\right)$ is $(G^2\left(\mathbb{R}^d\right), U, \mathbf{1}, \mathbf{1}, \dots)$ and $G^2\left(\mathbb{R}^d\right)$ is nilpotent. $\square$

Consequently, the following weak convergence result holds for centered iid random variables on $G^2\left(\mathbb{R}^d\right)$.

Theorem 4.6 (Central limit theorem for centered iid random variables on a nilpotent Lie group, [3])**.** *Let N be a simply connected nilpotent Lie group and let $(\mathbf{X}_i)_{i\geq 1}$ be a sequence of iid random variables with values in N. Let $\mathbf{X}_i$ be centered (i.e. their projection $\pi_1(\mathbf{X}_i)$ on the abelianization of N has mean 0) and have a finite second moment i.e. $E[\|\mathbf{X}_i\|^2] < \infty$. Then we have*

$$\delta_{n^{-1/2}}(\mathbf{X}_1 \otimes \dots \otimes \mathbf{X}_n) \overset{D}{\to} \mathbf{B}_1,$$

where $\mathbf{B}_1$ is the time 1 value of the Brownian motion on N associated to $\mathbf{X}_1$.

For a proof of the above theorem, see [4].
We now explain how convergence of the finite-dimensional distributions can be obtained from the above central limit theorem.
We first consider the case of a single time point $s \in [0,1]$. We have

$$d\left(\mathbf{W}_s^{(n)}, \mathbf{W}_{\lfloor ns\rfloor/n}^{(n)}\right) \leq \left\|\delta_{n^{-1/2}}\left(\mathbf{X}_{\lfloor ns\rfloor+1}\right)\right\| = n^{-1/2}\left\|\left(\mathbf{X}_{\lfloor ns\rfloor+1}\right)\right\| \overset{P}{\to} 0. \tag{4.1}$$

Since the metric space $(C^{0,\alpha-H\ddot{o}l}\left([0,1], G^2\left(\mathbb{R}^d\right)\right), \ \widetilde{d}_{\alpha-H\ddot{o}l})$ is separable, we can apply Proposition A.1, which yields that it is enough to show

$$\mathbf{W}_{\lfloor ns\rfloor/n}^{(n)} \overset{D}{\to} \mathbf{B}_s. \tag{4.2}$$

We have

$$\begin{aligned}\mathbf{W}_{\lfloor ns\rfloor/n}^{(n)} &= \delta_{n^{-1/2}}\left(\mathbf{X}_1 \otimes \dots \otimes \mathbf{X}_{\lfloor ns\rfloor}\right) \\ &= \delta_{(\lfloor ns\rfloor/n)^{1/2}}\left(\delta_{\lfloor ns\rfloor^{-1/2}}\left(\mathbf{X}_1 \otimes \dots \otimes \mathbf{X}_{\lfloor ns\rfloor}\right)\right) \overset{D}{\to} \delta_{s^{1/2}}\left(\mathbf{B}_1\right) = \mathbf{B}_s,\end{aligned}$$

[2] see Definition A.5.

where the convergence in distribution is due to the above central limit theorem, the continuous mapping theorem (the dilation map is continuous) and the fact that $\lim_{n\to\infty} \lfloor ns \rfloor / n = s$.
Now, we consider two time points $s, t \in [0,1]$, $s < t$. The aim is to show that $(\mathbf{W}_s^{(n)}, \mathbf{W}_t^{(n)}) \xrightarrow{D} (\mathbf{B}_s, \mathbf{B}_t)$. It follows from (4.1) that it suffices to prove

$$\left(\mathbf{W}^{(n)}_{\lfloor ns \rfloor/n}, \mathbf{W}^{(n)}_{\lfloor nt \rfloor/n}\right) \xrightarrow{D} (\mathbf{B}_s, \mathbf{B}_t) .$$

By the continuous mapping theorem, it is enough to show

$$\left(\mathbf{W}^{(n)}_{\lfloor ns \rfloor/n}, \left(\mathbf{W}^{(n)}_{\lfloor ns \rfloor/n}\right)^{-1} \otimes \mathbf{W}^{(n)}_{\lfloor nt \rfloor/n}\right) \xrightarrow{D} (\mathbf{B}_s, \mathbf{B}_{s,t}) . \tag{4.3}$$

As the components of the left are independent, we can apply Theorem A.2. Thus, (4.3) follows from (4.2), identical distribution of the $\mathbf{X}_i$ and the fact that $\mathbf{B}_{t-s} \overset{D}{=} \mathbf{B}_{s,t}$. Analogous arguments also apply to three or more time points, so that we obtain convergence of the finite-dimensional distributions.

4.2.2 Step 2: Tightness

We will take advantage of the following result:

Theorem 4.7 (Kolmogorov-Lamperti tightness criterion, [5], Cor. A.11)**.** *Let $\left(\{\mathbf{Y}_t^{(n)} : t \in [0,1]\}\right)_{n\geq 1}$ be a sequence of continuous $G^2(\mathbb{R}^d)$-valued stochastic processes. Assume there exist constants $a, b, c > 0$, satisfying $a > 1 + b$, such that*

$$\sup_{n\geq 1} E\left[d\left(\mathbf{Y}_s^{(n)}, \mathbf{Y}_t^{(n)}\right)^a\right] \leq c\,|t-s|^{1+b}$$

for all $s, t \in [0,1]$.
Then, $\left(\mathbf{Y}^{(n)}\right)_{n\geq 1}$ is tight in $C^{\gamma-Höl}\left([0,1], G^2(\mathbb{R}^d)\right)$ for any $\gamma \in \left(0, \frac{b}{a}\right)$.

Remark 4.8. Note that for a sequence of stochastic processes $\left(\mathbf{Y}^{(n)}\right)_{n\geq 1}$ whose sample paths are a.s. in $C^{0,\gamma-Höl}\left([0,1], G^2(\mathbb{R}^d)\right)$, tightness in $C^{\gamma-Höl}\left([0,1], G^2(\mathbb{R}^d)\right)$ implies tightness in $C^{0,\gamma-Höl}\left([0,1], G^2(\mathbb{R}^d)\right)$.

We will prove the following:
There exist constants $a, b, c > 0$, $a > 1 + b$, such that b/a lies arbitrarily close to $1/2$ and

$$E\left[d\left(\mathbf{W}_s^{(n)}, \mathbf{W}_t^{(n)}\right)^a\right] \leq c\,|t-s|^{1+b} \tag{4.4}$$

holds for all $s, t \in [0,1]$ and all $n \in \mathbb{N}$. By the above tightness criterion, $\left(\mathbf{W}^{(n)}\right)_{n\geq 1}$ is then tight in $C^{0,\gamma-Höl}\left([0,1], G^2(\mathbb{R}^d)\right)$ for any $\gamma \in (0, 1/2)$.

We now fix $n \in \mathbb{N}$ and simplify condition (4.4) in several steps. First, we claim that it is enough to show the bound (4.4) for $s = 0$ and $t = \frac{k}{n}$, $k = 1, \ldots, n$. This will be established in two steps:

Claim 1: If (4.4) holds for $s = 0$ and $t = \frac{k}{n}$, $k = 1, \ldots, n$, then (4.4) holds for all $s, t \in \left\{0, \frac{1}{n}, \frac{2}{n}, \ldots, 1\right\}$.

Claim 2: If (4.4) holds for all $s, t \in \left\{0, \frac{1}{n}, \frac{2}{n}, \ldots, 1\right\}$, then (4.4) holds for all $s, t \in [0, 1]$.

Proof of Claim 1. Let $s, t \in \left\{\frac{1}{n}, \frac{2}{n}, \ldots, 1\right\}$, $s < t$. By left-invariance of the Carnot-Carathéodory metric, we have

$$d\left(\mathbf{W}_s^{(n)}, \mathbf{W}_t^{(n)}\right) = \ d\left(\mathbf{1}, \delta_{n^{-1/2}}\left(\mathbf{X}_{ns+1} \otimes \cdots \otimes \mathbf{X}_{nt}\right)\right).$$

We set $k := nt - ns \in \{1, \ldots, n-1\}$. As the $\mathbf{X}_i$ are identically distributed, $\delta_{n^{-1/2}}(\mathbf{X}_{ns+1} \otimes \cdots \otimes \mathbf{X}_{nt})$ has the same distribution as $\mathbf{W}_{k/n}^{(n)} = \delta_{n^{-1/2}}(\mathbf{X}_1 \otimes \cdots \otimes \mathbf{X}_{nt-ns})$.

Thus, as the map $g \in G^2\left(\mathbb{R}^d\right) \mapsto d(\mathbf{1}, g)$ is continuous and thus measurable, we have

$$E\left[d\left(\mathbf{W}_s^{(n)}, \mathbf{W}_t^{(n)}\right)^a\right] = E\left[d\left(\mathbf{W}_0^{(n)}, \mathbf{W}_{k/n}^{(n)}\right)^a\right],$$

which implies Claim 1. □

Proof of Claim 2. By assumption, there exists a constant $\overline{c} > 0$ such that

$$E\left[d\left(\mathbf{W}_s^{(n)}, \mathbf{W}_t^{(n)}\right)^a\right] \leq \overline{c}\,|t - s|^{1+b}.$$

holds for all $s, t \in \left\{0, \frac{1}{n}, \frac{2}{n}, \ldots, 1\right\}$.

Let $s, t \in [0, 1]$.

Case 1: $|s - t| \leq \frac{1}{n}$

(i) $\exists k \in \{0, 1, \ldots, n-1\} : s, t \in \left[\frac{k}{n}, \frac{k+1}{n}\right]$.

Without loss of generality (W.l.o.g.), we can assume k=0. Hence, there exist $y, z \in [0, 1]$ such that $s = \frac{y}{n}$ and $t = \frac{z}{n}$. We use the defining property (2.12) for geodesics and the assumption $a > 1 + b$ to get

$$\begin{aligned} E\left[d\left(\mathbf{W}_{y/n}^{(n)}, \mathbf{W}_{z/n}^{(n)}\right)^a\right] &= |y - z|^a\, E\left[d\left(\mathbf{W}_0^{(n)}, \mathbf{W}_{1/n}^{(n)}\right)^a\right] \\ \leq\ |y - z|^a\, \overline{c}\left(\frac{1}{n}\right)^{1+b} &\leq \overline{c}\left|\frac{1}{n}(y - z)\right|^{1+b} = \overline{c}\,|s - t|^{1+b}. \end{aligned}$$

(ii) $\exists k \in \{0,1,\ldots,n-2\} : s \in \left[\frac{k}{n}, \frac{k+1}{n}\right] \ \wedge \ t \in \left[\frac{k+1}{n}, \frac{k+2}{n}\right]$.

W.l.o.g., we can assume k=0. Using the triangle inequality, convexity of the function x^a on $\mathbb{R}_{\geq 0}$ for $a > 1$ and the bound obtained in Case (i) yields[3]

$$\begin{aligned} E\left[d\left(\mathbf{W}_s^{(n)}, \mathbf{W}_t^{(n)}\right)^a\right] &\leq E\left[\left(d\left(\mathbf{W}_s^{(n)}, \mathbf{W}_{1/n}^{(n)}\right) + d\left(\mathbf{W}_{1/n}^{(n)}, \mathbf{W}_t^{(n)}\right)\right)^a\right] \\ &\leq 2^{a-1}\left(E\left[d\left(\mathbf{W}_s^{(n)}, \mathbf{W}_{1/n}^{(n)}\right)^a\right] + E\left[d\left(\mathbf{W}_{1/n}^{(n)}, \mathbf{W}_t^{(n)}\right)^a\right]\right) \\ &\leq 2^{a-1}\left(\overline{c}\left|\frac{1}{n} - s\right|^{1+b} + \overline{c}\left|t - \frac{1}{n}\right|^{1+b}\right) \leq 2^a \overline{c}\,|s-t|^{1+b}. \end{aligned}$$

Case 2: $|s-t| > \frac{1}{n}$

Then, there exist $k, l \in \{0,1,\ldots,n-1\}$, $k < l$ such that $s \in \left[\frac{k}{n}, \frac{k+1}{n}\right)$ and $t \in \left[\frac{l}{n}, \frac{l+1}{n}\right)$.

Due to geodesic interpolation between time points $\frac{k}{n}$, $k = 0,1,\ldots,n$, we have

$$d\left(\mathbf{W}_s^{(n)}, \mathbf{W}_t^{(n)}\right) \leq d\left(\mathbf{W}_{k/n}^{(n)}, \mathbf{W}_{(l+1)/n}^{(n)}\right).$$

It follows that

$$\begin{aligned} E\left[d\left(\mathbf{W}_s^{(n)}, \mathbf{W}_t^{(n)}\right)^a\right] &\leq E\left[d\left(\mathbf{W}_{k/n}^{(n)}, \mathbf{W}_{(l+1)/n}^{(n)}\right)^a\right] \\ &\leq \overline{c}\left|\frac{l+1-k}{n}\right|^{1+b} \leq 3^{1+b}\overline{c}\,|t-s|^{1+b}, \end{aligned}$$

where in the last inequality we have used that

$$\left|\frac{l+1-k}{n}\right| \leq |t-s| + \frac{2}{n} \leq 3\,|t-s|.$$

Consequently, if we choose $c := \max\left\{2^a\overline{c}, 3^{1+b}\overline{c}\right\}$, then (4.4) will hold for all $s, t \in [0,1]$. Note that the constant c does not depend on $n \in \mathbb{N}$. $\square$

Claim 1 and 2 imply that it suffices to show that there exist positive constants a, b, c, satisfying $a > 1 + b$, such that b/a lies arbitrarily close to $1/2$ and for all $n \in \mathbb{N}$ and all $k \in \{1,\ldots,n\}$

$$E\left[d\left(\mathbf{W}_0^{(n)}, \mathbf{W}_{k/n}^{(n)}\right)^a\right] \leq c\left|\frac{k}{n}\right|^{1+b}. \tag{4.5}$$

Now, we have

$$d\left(\mathbf{W}_0^{(n)}, \mathbf{W}_{k/n}^{(n)}\right) = \left\|\delta_{n^{-1/2}}\left(\mathbf{X}_1 \otimes \cdots \otimes \mathbf{X}_k\right)\right\| = n^{-1/2}\left\|\mathbf{X}_1 \otimes \cdots \otimes \mathbf{X}_k\right\|,$$

[3] The idea for this argument is taken from [6].

which implies that (4.5) is equivalent to

$$n^{-a/2} E[\|\mathbf{X}_1 \otimes \cdots \otimes \mathbf{X}_k\|^a] \le c \left|\frac{k}{n}\right|^{1+b}$$

for all $n \in \mathbb{N}$ and all $k \in \{1, \ldots, n\}$.
This will follow if we prove that for all $p \in \mathbb{N}$

$$E[\|\mathbf{X}_1 \otimes \cdots \otimes \mathbf{X}_k\|^{4p}] = O(k^{2p}) \quad \text{for } k \to \infty, \tag{4.6}$$

as we can then choose $a_p = 4p$ and $b_p = 2p-1$, $p \in \mathbb{N}$, and clearly $\lim_{p\to\infty} \frac{2p-1}{4p} = \frac{1}{2}$, so that we obtain tightness of $\left(\mathbf{W}^{(n)}\right)_{n\ge1}$ in $C^{0,\alpha-H\ddot{o}l}\left([0,1], G^2\left(\mathbb{R}^d\right)\right)$ for any $\alpha \in (0, 1/2)$.

The rest of the proof aims at establishing (4.6) by using polynomial functions on $G^2\left(\mathbb{R}^d\right)$, which will now be introduced. Let $a = (a^{1;k}, a^{2;ij}; 1 \le k \le d, 1 \le i < j \le d) \in g^2(\mathbb{R}^d)$ be the log-chart of $G^2\left(\mathbb{R}^d\right)$, $g \mapsto a = \log(g)$, and let P be a polynomial function on $G^2\left(\mathbb{R}^d\right)$ i.e. a polynomial in $a^{1;k}, a^{2;ij}$. In the sequel, we will use the term "polynomial" to mean "polynomial function on $G^2\left(\mathbb{R}^d\right)$".
The degree $d^\circ P$ of P is defined as follows: If P is a monomial of the form

$$P = \prod_{r=1}^{n} \prod_{s=1}^{m} (a^{1;k_r})^{\alpha_{k_r}} (a^{2;i_s j_s})^{\alpha_{i_s j_s}},$$

where $k_r \in \{1, 2, \ldots, d\}$ and $i_s, j_s \in \{1, 2, \ldots, d\}$, $i_s < j_s$, then

$$d^\circ P := \sum_{r=1}^{n} \alpha_{k_r} + 2 \sum_{s=1}^{m} \alpha_{i_s j_s}.$$

If P is an arbitrary polynomial, then $d^\circ P$ is the maximum of the degrees of its monomials, as usual.
Next, we will express (4.6) in terms of a polynomial function on $G^2\left(\mathbb{R}^d\right)$. For this, let $\widetilde{P}$ be the polynomial defined by

$$\widetilde{P}(\exp(a)) := \sum_{i=1}^{d} \left(a^{1;i}\right)^{4p} + \sum_{\substack{i,j \in \{1,\ldots,d\} \\ i<j}} \left(a^{2;ij}\right)^{2p}$$

for $a \in g^2\left(\mathbb{R}^d\right)$. Clearly, $d^\circ \widetilde{P} = 4p$. We claim that

Lemma 4.9. *For each $p \in \mathbb{N}$, there exists a constant $C_p > 0$ such that for all $a \in g^2(\mathbb{R}^d)$,*

$$\|\exp(a)\|^{4p} \le C_p\, \widetilde{P}(\exp(a)). \tag{4.7}$$

Proof. Consider the map

$$\begin{aligned}\|.\|_1 : G^2\left(\mathbb{R}^d\right) &\rightarrow \mathbb{R}_{\geq 0} \\ g &\mapsto |\pi_1(\log(g))| + |\pi_2(\log(g))|^{1/2}.\end{aligned}$$

It is easy to see that $\|.\|_1$ is a homogenous norm on $G^2\left(\mathbb{R}^d\right)$. By equivalence of homogenous norms on $G^2\left(\mathbb{R}^d\right)$ (Theorem 2.44), $\|.\|_1$ is equivalent to the Carnot-Carathéodory norm. Hence, there exists a positive constant $\overline{C}$ such that for all $a = (0, a^1, a^2) \in g^2\left(\mathbb{R}^d\right)$,

$$\|\exp(a)\| \leq \overline{C}\,\|\exp(a)\|_1.$$

We now fix $a \in g^2\left(\mathbb{R}^d\right)$. Then, we have

$$\begin{aligned}\|\exp(a)\|_1^{4p} &= \left(\left|a^1\right| + \left|a^2\right|^{1/2}\right)^{4p} \\ &\leq 2^{4p}\left(\left|a^1\right|^{4p} + \left|a^2\right|^{2p}\right) \\ &= 2^{4p}\left(\left(\sum_{i=1}^{d}\left(a^{1;d}\right)^2\right)^{2p} + \left(\sum_{i,j\in\{1,\ldots,d\}}\left(a^{2;ij}\right)^2\right)^{p}\right) \\ &\leq 2^{4p}d^{4p}\left(\sum_{i=1}^{d}\left(a^{1;d}\right)^{4p} + \sum_{i,j\in\{1,\ldots,d\}}\left(a^{2;ij}\right)^{2p}\right) \\ &\leq 2^{4p+1}d^{4p}\left(\sum_{i=1}^{d}\left(a^{1;d}\right)^{4p} + \sum_{\substack{i,j\in\{1,\ldots,d\}\\ i<j}}\left(a^{2;ij}\right)^{2p}\right),\end{aligned}$$

where the last inequality is due to the fact that $a^{2;ji} = -a^{2;ij}$ for all $i, j \in \{1, \ldots, d\}$. It follows that (4.7) holds with $C_p = \overline{C}^{4p}2^{4p+1}d^{4p}$. $\square$

To obtain (4.6), it is thus enough to prove that

$$E[\widetilde{P}(\mathbf{X}_1 \otimes \cdots \otimes \mathbf{X}_k)] = O(k^{2p}). \tag{4.8}$$

We now introduce the key object of this proof. It is a map T on polynomial functions on $G^2\left(\mathbb{R}^d\right)$. For this, let $\mathbf{X}$ be a $G^2\left(\mathbb{R}^d\right)$-valued random variable with finite moments of all orders and let P be a polynomial function on $G^2\left(\mathbb{R}^d\right)$. Then, we define the polynomial function TP on $G^2\left(\mathbb{R}^d\right)$ by

$$TP(g) := E[P(g \otimes \mathbf{X})] - P(g).$$

We will take advantage of the fact that applying the map T to a polynomial yields a polynomial of at least two degrees lower.

Lemma 4.10. *If P is a polynomial function of degree $d^\circ P \geq 2$ and the random variable $\mathbf{X}$ (introduced in the above definition of the map T) is centered, we have*

$$d^\circ(TP) \leq d^\circ P - 2. \tag{4.9}$$

The proof of this lemma heavily relies on the Campbell-Baker-Hausdorff formula and the assumption that $\mathbf{X}$ is centered.

Proof. Clearly, it is enough to show (4.9) for an arbitrary monomial. Let $g \in G^2\left(\mathbb{R}^d\right)$ be given. We set $a := \log(g) \in g^2\left(\mathbb{R}^d\right)$ and $\varphi := \log(\mathbf{X})$. Then, φ is a $g^2\left(\mathbb{R}^d\right)$-valued random variable. We write $a = (0, a^1, a^2)$ and $\varphi = (0, \varphi^1, \varphi^2)$. Then, we have $\varphi^1 = \pi_1(\mathbf{X})$, so that $E[\varphi^1] = 0$ i.e. φ is also centered. This fact will be central to the following argumentation.
Then, we have

$$\begin{aligned} TP(g) &= E[P(\exp(a) \otimes \exp(\varphi))] - P(\exp(a)) \\ &= E\left[P\left(\exp\left(a + \varphi + \frac{1}{2}[a,\varphi]\right)\right)\right] - P(\exp(a)) \qquad (4.10) \\ &= E\left[P\left(\exp\left(\left(0, a^1 + \varphi^1, a^2 + \varphi^2 + \frac{1}{2}\left(a^1 \otimes \varphi^1 - \varphi^1 \otimes a^1\right)\right)\right)\right)\right] \\ &\quad - P(\exp(a)), \qquad (4.11) \end{aligned}$$

where in (4.10), we have applied the Campbell-Baker-Hausdorff formula.
We now study the expression (4.11) for two examples of simple monomials and then we discuss arbitrary monomials.

Example 1:

$$P(g) = (a^{1;i})^m$$

for some $i \in \{1, \ldots, d\}$ and $m \in \mathbb{N} \setminus \{1\}$. Then, $d^\circ P = m$ and we have

$$\begin{aligned} TP(g) &= E\left[\left(a^{1;i} + \varphi^{1;i}\right)^m\right] - \left(a^{1;i}\right)^m \\ &= \left(a^{1;i}\right)^m + m \underbrace{E\left[\varphi^{1;i}\right]}_{=0} \left(a^{1;i}\right)^{m-1} + \binom{m}{2} E\left[(\varphi^{1;i})^2\right] \left(a^{1;i}\right)^{m-2} \\ &\quad + E[\text{ terms of lower degree }] - \left(a^{1;i}\right)^m \\ &= r\left(a^{1;i}\right)^{m-2} + \text{ terms of lower degree,} \end{aligned}$$

where r is a real number. Thus, $d^\circ(TP) \leq m - 2 = d^\circ P - 2$.

Example 2:

$$P(g) = (a^{2;ij})^m$$

for some $i, j \in \{1, \dots, d\},\ i < j$ and $m \in \mathbb{N}$. Then, $d^\circ P = 2m$. We assume now that $m \geq 2$. Then, similarly to Example 1, we have

$$\begin{aligned}
TP(g) = {} & E\left[\left(a^{2;ij} + \varphi^{2;ij} + \frac{1}{2}\left(a^{1;i}\varphi^{1;j} - \varphi^{1;i}a^{1;j}\right)\right)^m\right] - \left(a^{2;ij}\right)^m \\
= {} & \left(a^{2;ij}\right)^m + mE\left[\varphi^{2;ij}\right]\left(a^{2;ij}\right)^{m-1} \\
& + \frac{m}{2}\underbrace{E\left[\left(a^{1;i}\varphi^{1;j} - \varphi^{1;i}a^{1;j}\right)\right]}_{=0}\left(a^{2;ij}\right)^{m-1} \\
& + \frac{1}{2}\binom{m}{2}E\left[\left(a^{1;i}\varphi^{1;j} - \varphi^{1;i}a^{1;j}\right)^2\right]\left(a^{2;ij}\right)^{m-2} \\
& + E[\text{ terms of lower degree }] - \left(a^{2;ij}\right)^m \\
= {} & r_1\left(a^{2;ij}\right)^{m-1} + \left(r_2\left(a^{1;i}\right)^2 + r_3 a^{1;i}a^{1;j} + r_4\left(a^{1;j}\right)^2\right)\left(a^{2;ij}\right)^{m-2} \\
& + \text{ terms of lower degree,}
\end{aligned}$$

where $r_1, \dots, r_4$ are real numbers.
In the case $m = 1$, it is easy to see that $TP(g) = E[\varphi^{2;ij}] \in \mathbb{R}$, which is a polynomial of degree 0.
Thus, we conclude that $d^\circ(TP) \leq 2m - 2 = d^\circ P - 2$.

The reasoning in Examples 1 and 2 basically carries over to arbitrary monomials, which are finite products of powers of $a^{1;k}$, $k \in \{1, \dots, d\}$ and $a^{2;ij}$, $i, j \in \{1, \dots, d\}$, $i < j$. To see this, note that for an arbitrary monomial P, we have:

1. The term of degree $d^\circ P$ in $E[P(g \otimes \mathbf{X})]$ equals $P(g)$ and therefore it cancels out in $TP(g)$.

2. All terms of degree $d^\circ P - 1$ in $P(g \otimes \mathbf{X})$ contain exactly one component of the centered random variable φ^1 and thus they have expectation zero.

This implies that $TP(g)$ is a polynomial of degree $\leq d^\circ P - 2$. □

Now, for any polynomial P on $G^2\left(\mathbb{R}^d\right)$ and $k \in \mathbb{N}$, we have

$$\begin{aligned}
& E[P(\mathbf{X}_1 \otimes \cdots \otimes \mathbf{X}_k)] && (4.12)\\
= {} & E[E[P((\mathbf{X}_1 \otimes \cdots \otimes \mathbf{X}_{k-1}) \otimes \mathbf{X}_k)|\mathbf{X}_1, \dots, \mathbf{X}_{k-1}]] \\
= {} & E[TP(\mathbf{X}_1 \otimes \cdots \otimes \mathbf{X}_{k-1}) + P(\mathbf{X}_1 \otimes \cdots \otimes \mathbf{X}_{k-1})] && (4.13)\\
= {} & E[(T + Id)(P(\mathbf{X}_1 \otimes \cdots \otimes \mathbf{X}_{k-1}))], && (4.14)
\end{aligned}$$

where (4.13) follows from the definition of the map T and the fact that $\mathbf{X}_k$ is independent of $\mathbf{X}_1, \dots, \mathbf{X}_{k-1}$.

By iteratively applying the equality between (4.12) and (4.14), we obtain

$$E[\widetilde{P}(\mathbf{X}_1 \otimes \cdots \otimes \mathbf{X}_k)] = (T + Id)^k \widetilde{P}(\mathbf{1}) = \sum_{l=1}^{k} \binom{k}{l} T^l \widetilde{P}(\mathbf{1}). \tag{4.15}$$

In fact, the number of summands in (4.15) does not depend on k. To see this, we use that applying the map T to a polynomial gives a polynomial of at least two degrees lower (Lemma 4.10) to obtain

$$d^\circ(T^l \widetilde{P}) \leq d^\circ \widetilde{P} - 2l = 2(2p - l),$$

which implies

$$E[\widetilde{P}(\mathbf{X}_1 \otimes \cdots \otimes \mathbf{X}_k)] = \sum_{l=0}^{2p} \binom{k}{l} T^l \widetilde{P}(\mathbf{1}). \tag{4.16}$$

Now, for all $k \in \mathbb{N}$, for all $l \in \mathbb{N} \cup \{0\}$, we have $\binom{k}{l} \leq \frac{k^l}{l!}$, so that $\binom{k}{l} = O(k^l)$ as $k \to \infty$. Hence, for all $l \in \{0, 1, \ldots, 2p\}$

$$\binom{k}{l} T^l \widetilde{P}(\mathbf{1}) = O(k^{2p}).$$

Combining this with (4.16) yields the estimate (4.8). □

5 Some probabilistic tools

In this chapter, we introduce some notions and give some results that will be used to prove the invariance principles in the subsequent chapters.

5.1 Strong mixing sequences of random variables

The strong mixing coefficient is a means to quantify the dependence between random variables.

Definition 5.1. Let $(\Omega, \mathcal{F}, P)$ be a probability space. For two σ-fields $\mathcal{A}, \mathcal{B} \subseteq \mathcal{F}$, we define the **strong mixing coefficient** of $\mathcal{A}$ and $\mathcal{B}$ by

$$\alpha(\mathcal{A},\mathcal{B}) := \sup_{(A,B)\in\mathcal{A}\times\mathcal{B}} |P(A\cap B) - P(A)P(B)|\,.$$

Definition 5.2. Let $(X_i)_{i\geq 1}$ be a sequence of random variables defined on the same probability space $(\Omega, \mathcal{F}, P)$. For $k \in \mathbb{N}$ and $l \in \mathbb{N}$, we define the σ-algebras

$$\begin{aligned} \mathcal{F}_k^l &:= \sigma(X_i : k \leq i \leq l) \quad \text{and} \\ \mathcal{F}_k^\infty &:= \sigma(X_i : i \geq k). \end{aligned}$$

For $n \in \mathbb{N}$

$$\alpha_n := \sup_{m\in\mathbb{N}} \alpha(\mathcal{F}_1^m, \mathcal{F}_{n+m}^\infty).$$

If $\lim_{n\to\infty} \alpha_n = 0$, we say that the sequence $(X_i)_{i\geq 1}$ is **$\boldsymbol{\alpha}$-mixing** or **strong mixing**.

Thus, random variables that lie far apart in a strong mixing sequence of random variables are "almost independent".
R. Yokoyama proved a moment bound for partial sums of strong mixing sequences of random variables, which will be applied several times in the proofs of the following invariance principles. In particular, this bound will be used to establish tightness for a real-valued Markov chain (which will be shown to be strong mixing).

Theorem 5.3 (Yokoyama's moment bound, [22])**.** *Let $(X_i)_{i\geq 1}$ be a stationary*[1]*, strong mixing sequence of real-valued centered random variables satisfying $E[|X_1|^{r+\delta}] < \infty$ for some $r > 2$ and some $\delta > 0$. If*

$$\sum_{i=0}^{\infty} (i+1)^{r/2-1} \alpha_i^{\delta/(r+\delta)} < \infty, \tag{5.1}$$

[1]see Definition A.3.

then there exists a constant $C > 0$ *such that for all* $n \in \mathbb{N}$

$$E\left[\left|\sum_{i=1}^{n} X_i\right|^r\right] \leq C n^{r/2}.$$

5.2 The martingale problem

To establish convergence of the finite-dimensional distributions, Breuillard et al. used the central limit theorem for independent random variables on nilpotent Lie groups. To our knowledge, no similar result has been established for dependent random variables on nilpotent Lie groups. We will thus take a different approach, which is based on the concept introduced in this section.

Introduction and definition

The martingale problem formulation – a concept introduced by D. Stroock and S.R.S. Varadhan – is a way to specify Markov processes. Other approaches are partial differential equations and stochastic differential equations. All three approaches were shown to be equivalent. We will state this more precisely below for the martingale problem and stochastic differential equations. The martingale problem formulation has the advantage that one can use the following powerful techniques: the theory of weak convergence of probability measures on complete separable metric spaces, the theory of regular conditional probabilities and localization arguments of martingale theory. The main reference for this section is [20].
First, we introduce some notation and some basic objects. Let d_∞ be the metric on $\Omega := C([0,1], \mathbb{R}^d)$ induced by the supremum norm and let $\mathcal{M}$ be the Borel σ-algebra on the metric space (Ω, d_∞).
The coordinate process on Ω is the map $x : [0,1] \times \Omega \to \mathbb{R}^d$ defined by $x(t,\omega) = \omega(t)$. One can show that $\mathcal{M} = \sigma(x(s) : 0 \leq s \leq 1)$ (see [20], p. 30). For $t \in [0,1]$, we define the sigma-algebra $\mathcal{M}_t := \sigma(x(s) : 0 \leq s \leq t)$.
For a $d \times d$ real matrix $a = (a_{ij})_{1 \leq i,j \leq d}$, let $\|a\|_{\text{op}} := \sup_{x \in \mathbb{R}^d, |x|=1} |Ax|$ denote its operator norm. In the sequel, spaces of real matrices will always be endowed with the topology induced by this norm. Let S_d be the set of symmetric, positive semidefinite $d \times d$ real matrices. We assume that the maps $a : \mathbb{R}^d \to S_d$ and $b : \mathbb{R}^d \to \mathbb{R}^d$ are Borel measurable. For such a pair (a,b) of maps, we define the second-order differential operator $L_{a,b}$ by

$$L_{a,b} f(y) := \frac{1}{2} \sum_{i,j=1}^{d} a_{ij}(y) \frac{\partial^2}{\partial y_i \partial y_j} f(y) + \sum_{i=1}^{d} b_i(y) \frac{\partial}{\partial y_i} f(y)$$

for $y \in \mathbb{R}^d$ and $f : \mathbb{R}^d \to \mathbb{R}$ a two times continuously partially differentiable function. Let $C_0^\infty\left(\mathbb{R}^d\right)$ denote the set of maps $\mathbb{R}^d \to \mathbb{R}$ having compact support and continuous partial derivatives of all orders. We are now ready to state the basic definition of this section:

Definition 5.4. A probability measure P on $\big(C([0,1],\mathbb{R}^d),\mathcal{M}\big)$ is a **solution to the martingale problem for $(\mathbf{a},\mathbf{b})$ starting at $\mathbf{y_0}$** $\in \mathbb{R}^d$ if the following two conditions hold:

(*i*) $P(x(0) = y_0) = 1$;

(*ii*) For any $f \in C_0^\infty\left(\mathbb{R}^d\right)$, the stochastic process

$$\left(f(x(t)) - f(x(0)) - \int_0^t L_{a,b} f(x(u))du\right)_{t\in[0,1]}$$

is a P-martingale with respect to the filtration $(\mathcal{M}_t)_{t\in[0,1]}$.

We say that the martingale problem for (a,b) is **well-posed** if for every $y_0 \in \mathbb{R}^d$, there exists a unique solution to the martingale for (a,b) starting at y_0.

A useful characterization of martingales

The next lemma will be useful for working with the martingale problem formulation in the subsequent chapters.

Lemma 5.5. *Let $(M_t)_{t\in[0,1]}$ be a family of real-valued random variables defined on $\big(C([0,1],\mathbb{R}^d),\mathcal{M}\big)$ and let P be a probability measure on this function space. Then, the following two statements are equivalent:*

(*i*) *$(M_t)_{t\in[0,1]}$ is a P-martingale with respect to the filtration $(\mathcal{M}_t)_{t\in[0,1]}$.*

(*ii*) *For all s, $t \in [0,1]$, $s < t$ and for any continuous, bounded, $\mathcal{M}_s$-measurable map $\phi : C([0,1],\mathbb{R}^d) \to \mathbb{R}$, we have*

$$E_P[(M_t - M_s)\phi] = 0.$$

In the proof of the above lemma, the following result will be applied twice.

Lemma 5.6. *Let ξ be a (real-valued) random variable in $\mathcal{L}_1(\Omega,\mathcal{F},P)$ and let $\mathcal{D}$ be a generating system of $\mathcal{F}$ which is stable under intersections. If for all $A \in \mathcal{D}$*

$$\int_A \xi \; dP = 0,$$

then $\xi = 0$ almost surely.

The proof of the above lemma is left to the reader.

Proof of Lemma 5.5. (*i*) $\Rightarrow$ (*ii*): This can be established by a standard argument from measure theory.

$(ii) \Rightarrow (i)$: (i) is equivalent to the following fact: For all $s,\, t \in [0,1]$, $s < t$ and all $A \in \mathcal{M}_s$, we have

$$\int_A (M_t - M_s)\; dP = 0.$$

We fix $s,\, t \in [0,1]$ satisfying $s < t$. Observe that

$$\mathcal{D}_s := \left(\bigcup_{\substack{I \subset [0,s];\\ I \text{ finite}, I \neq \emptyset}} \sigma\left(x(u) : u \in I\right) \right) \cup \emptyset$$

is stable under intersections and a generating system of $\mathcal{M}_s = \sigma\left(x(u) : u \in [0,s]\right)$. Hence, by Lemma 5.6, it suffices to show that for all $D \in \mathcal{D}_s$

$$\int_D (M_t - M_s)\; dP = 0. \tag{5.2}$$

Note that for $I = \{i_1, \dots, i_k\}$, we have

$$\sigma\left(x(i_1), \dots, x(i_k)\right) = \sigma\left(\pi^{-1}_{i_1,\dots,i_k}(B) : B \in \mathcal{B}_k\right),$$

where $\mathcal{B}_k$ denotes the Borel σ-algebra on $\left(\mathbb{R}^d\right)^k \cong \mathbb{R}^{dk}$. Hence, (5.2) is equivalent to the following fact: For $k \in \mathbb{N}$, $0 \leq i_1 < i_2 < \cdots < i_k \leq s$ and $B \in \mathcal{B}_k$, we have

$$\int (M_t - M_s)\; \mathbf{1}_B(\pi_{i_1,\dots,i_k})\; dP = 0. \tag{5.3}$$

Now, we apply Lemma 5.6 again: Since the set of compact subsets of $\mathbb{R}^{dk}$ is stable under intersections and generates $\mathcal{B}_k$, it is enough to prove (5.3) for all compact sets $B \subset \mathbb{R}^{dk}$.
We now fix a compact $B \subset \mathbb{R}^{dk}$. For $n \in \mathbb{N}$, we define the function $f_n : \mathbb{R}^{dk} \to [0,1]$ by[2]

$$f_n(x) = \begin{cases} 1 & \text{if } x \in B \\ 0 & \text{if } d(x,B) \geq 1/n \\ 1 - n\; d(x,B) & \text{if } 0 < d(x,B) < 1/n. \end{cases}$$

Then, for all $x \in \mathbb{R}^{dk}$, we have $\lim_{n\to\infty} f_n(x) = \mathbf{1}_B(x)$. Note that $f_n(\pi_{i_1,\dots,i_k}) : C([0,1], \mathbb{R}^{dk}) \to [0,1]$ is a continuous, bounded, $\mathcal{M}_s$-measurable function, so that by (ii), we have

$$\int (M_t - M_s)\; f_n(\pi_{i_1,\dots,i_k})\; dP = 0$$

for all $n \in \mathbb{N}$. By Lebesgue's dominated convergence theorem, this implies

$$\int (M_t - M_s)\; \mathbf{1}_B(\pi_{i_1,\dots,i_k})\; dP = \lim_{n\to\infty} \int (M_t - M_s)\; f_n(\pi_{i_1,\dots,i_k})\; dP = 0.$$

$\square$

[2] For $x \in \mathbb{R}^{dk}$, the distance to the set B is defined as $d(x,B) := \inf\left\{\|x - y\| : y \in B\right\}$.

Equivalence to the weak solution formulation for stochastic differential equations

In the proofs of the following invariance principles, we will characterize the law of the limit process as a solution to a certain martingale problem. We are therefore interested in conditions which ensure that a given martingale problem is well-posed. We will rely on a simple condition that comes from the theory of stochastic differential equations (abbreviated as SDEs). Since the relation between SDEs and the martingale problem will also be useful for obtaining alternative characterizations of the limit process, we now study this relation in some detail. The rest of this section is based on [18].
First, we need to introduce some notions from the theory of SDEs. The following stochastic integrals are Itô integrals.
Let M_d denote the space of $d \times d$ real matrices and let $\sigma : \mathbb{R}^d \to M_d$ and $b : \mathbb{R}^d \to \mathbb{R}^d$ be Borel measurable maps. Let μ be a probability measure on $\mathbb{R}^d$.

Definition 5.7. We say that the SDE

$$X_t = X_0 + \int_0^t \sigma(X_u)\ dB_u + \int_0^t b(X_u)\ du, \quad t \in [0,1], \tag{5.4}$$

has a **weak solution with initial distribution $\boldsymbol{\mu}$** if there exists a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,1]}, P)$ satisfying the usual conditions, together with continuous semi-martingales X and B with values in $\mathbb{R}^d$ such that the following conditions hold:

(i) B is an $(\mathcal{F}_t)$-Brownian motion.

(ii) X_0 has law μ.

(iii) $\int_0^t \Big(\text{trace}\,(\sigma(X_u)(\sigma(X_u))^T) + |b(X_u)|\Big)\, du < \infty$ a.s. for all $t \in [0,1]$.

(iv) $X_t = X_0 + \int_0^t \sigma(X_u)\ dB_u + \int_0^t b(X_u)\ du \quad$ for $t \in [0,1]$.

If for every probability measure μ on $\mathbb{R}^d$, there exists a weak solution to the SDE (5.4) with μ as its initial distribution, we say that the SDE (5.4) has a **weak solution**.

Remark 5.8. Let us write $b = (b_1, \dots, b_d)$, $\sigma = (\sigma_{ij})_{i,j\in\{1,\dots,d\}}$, $X = (X^1, \dots, X^d)$ and $B = (B^1, \dots, B^d)$. The vector notation used in (5.4) is defined as follows:

$$X_t^i = X_0^i + \sum_{j=1}^d \int_0^t \sigma_{ij}(X_u)\ dB_u^j + \int_0^t b_i(X_u)\ du, \quad i = 1, \dots, d.$$

Definition 5.9. The solution of the SDE (5.4) is said to be **unique in law for the initial distribution $\boldsymbol{\mu}$** if any two solutions (perhaps on different filtered probability spaces) with initial distribution μ have the same law.
The solution of (5.4) is said to be **unique in law** if it is unique in law for any initial distribution μ on $\mathbb{R}^d$.

Now, we can state the equivalence results rigorously.

Theorem 5.10. *If X is a weak solution of the SDE (5.4) starting at $y \in \mathbb{R}^d$, then the law of X solves the martingale problem for $(\sigma\sigma^T, b)$ starting at y.*
Conversely, if P_y is a solution to the martingale problem for $(\sigma\sigma^T, b)$ starting at $y \in \mathbb{R}^d$, then there exists a weak solution to the SDE (5.4) starting at y whose law is P_y.

Remark 5.11. The first implication is easy to prove. The second is more involved[3]. We will only use the second implication.

Theorem 5.12. *Uniqueness in law holds for the SDE (5.4) starting at $y \in \mathbb{R}^d$ if and only if there exists at most one solution to the martingale problem for $(\sigma\sigma^T, b)$ starting at y.*

Now, we assume that the maps σ and b are Lipschitz continuous i.e. there exists a positive constant L such that for all $x, y \in \mathbb{R}^d$

$$\|\sigma(x) - \sigma(y)\|_{\mathrm{op}} \leq L\,|x - y| \quad \text{and} \quad |b(x) - b(y)| \leq L\,|x - y|\,. \tag{5.5}$$

If condition (5.5) is satisfied, then the SDE (5.4) has a weak solution and uniqueness in law holds for this SDE. It is now a consequence of Theorem 5.10 and Theorem 5.12 that

Theorem 5.13. *If the maps σ and b are Lipschitz continuous, then the martingale problem for $(\sigma\sigma^T, b)$ is well-posed.*

[3]see for example [18], p. 160f..

6 Invariance principle for a real-valued Markov chain

In this chapter, we will use the martingale problem formulation to prove an invariance principle for a stochastic process constructed from a Markov chain which takes only two real values. This simple setting is suitable for demonstrating how the dependencies between the random variables can be handled.

6.1 Preliminaries and result

First, we introduce the α-Hölder topology for real-valued paths. The definitions and remarks in this section are taken from [6]. We assume $\alpha \in (0,1]$.

Definition 6.1. Let $C^{\alpha-H\ddot{o}l}([0,1],\mathbb{R})$ denote the space of paths $x:[0,1]\to\mathbb{R}$ satisfying $x_0 = 0$ and

$$\|x\|_\alpha = \sup_{\substack{s,t\in[0,1]\\ s\neq t}} \frac{|x_t - x_s|}{|t-s|^\alpha} + \sup_{t\in[0,1]} |x_t| < \infty.$$

Definition 6.2. The **Hölderian modulus of continuity** of a path $x:[0,1]\to\mathbb{R}$ is defined as

$$w_\alpha(x,\delta) := \sup_{\substack{s,t\in[0,1]\\ 0<|t-s|<\delta}} \frac{|x_t - x_s|}{|t-s|^\alpha}.$$

$C^{0,\alpha-H\ddot{o}l}([0,1],\mathbb{R}) := \big\{x \in C^{\alpha-H\ddot{o}l}([0,1],\mathbb{R}) : \lim_{\delta\to 0} w_\alpha(x,\delta) = 0\big\}$.

Remark 6.3. $(C^{0,\alpha-H\ddot{o}l}([0,1],\mathbb{R}), \|.\|_\alpha)$ is a separable closed subspace of the non-separable Banach space $(C^{\alpha-H\ddot{o}l}([0,1],\mathbb{R}), \|.\|_\alpha)$.

Theorem 6.4. *Let $(X_i)_{i\geq 1}$ be a Markov chain such that each X_i takes values in $\{-1,1\}$. Assume $P(X_1 = 1) = P(X_1 = -1) = 1/2$ and for some $p \in (0,1)$*

$$\begin{aligned} &P(X_{i+1} = 1 \mid X_i = 1) = p \quad \textit{and} \\ &P(X_{i+1} = -1 \mid X_i = -1) = p. \end{aligned}$$

For $n \in \mathbb{N}$, define the stochastic process $Y^{(n)} = \{Y_t^{(n)}\}_{t\in[0,1]}$ by $Y_0^{(n)} = 0$,

$$Y_t^{(n)} = \frac{1}{\sqrt{n}} \sum_{i=1}^{nt} X_i$$

for $t \in \big\{\frac{1}{n}, \frac{2}{n}, \dots, 1\big\}$ and piecewise linearly interpolated in between.
Let $\sigma = \sqrt{\frac{p}{1-p}}$. Then, for any $\alpha \in (0,1/2)$, $Y^{(n)}$ converges weakly to σB in $C^{0,\alpha-H\ddot{o}l}([0,1],\mathbb{R})$, where B is a standard Brownian motion.

6.2 Calculation of the strong mixing coefficients and an exponential covariance bound

We now establish two important auxiliary results for the proof of Theorem 6.4. In the whole section, $(X_i)_{i\geq 1}$ denotes the Markov chain introduced in that theorem. We will first calculate the strong mixing coefficients of the Markov chain $(X_i)_{i\geq 1}$. This result will then be used to obtain an exponential covariance bound.

Lemma 6.5. *Let $(\alpha_n)_{n\geq 1}$ denote the sequence of strong mixing coefficients of the Markov chain $(X_i)_{i\geq 1}$. Then, for all $n \in \mathbb{N}$, we have*

$$\alpha_n = \frac{1}{4}\left|2p-1\right|^n . \tag{6.1}$$

As $|2p-1| < 1$, this implies that the sequence $(X_i)_{i\geq 1}$ is strong mixing.

Proof of (6.1). For $k, l \in \mathbb{N}$, we set $\mathcal{G}_k^l := \sigma(X_i : k \leq i \leq l)$ and $\mathcal{G}_k^\infty := \sigma(X_i : i \geq k)$. Then, by the Markov property and stationarity, we have

$$\alpha_n = \sup_{m\in\mathbb{N}} \alpha(\mathcal{G}_1^m, \mathcal{G}_{m+n}^\infty) = \sup_{m\in\mathbb{N}} \alpha(\mathcal{G}_m^m, \mathcal{G}_{m+n}^{m+n}) = \alpha(\mathcal{G}_1^1, \mathcal{G}_{n+1}^{n+1}).$$

For $i \in \mathbb{N}$, we set $\mathcal{G}_i := \mathcal{G}_i^i$. Then, $\mathcal{G}_i = \{\{X_i = 1\}, \{X_i = -1\}\}$ and these events both have probability $1/2$.
For $n \in \mathbb{N}$, we define

$$\widehat{n} := \begin{cases} \left\lfloor \frac{n}{2} \right\rfloor & \text{if } n \text{ is odd} \\ \left\lfloor \frac{n}{2} \right\rfloor - 1 & \text{otherwise.} \end{cases}$$

Note that X_1 and X_{n+1} take the same value if and only if there is an even number of changes (which happen with probability $1-p$) between $+1$ and -1 in the Markov chain up to X_{n+1}. The probability that X_1 and X_{n+1} take the same value/different values are thus

$$G(n) := \sum_{i=0}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{2i} (1-p)^{2i} p^{n-2i} \qquad \text{and}$$

$$U(n) := 1 - G(n) = \sum_{i=0}^{\widehat{n}} \binom{n}{2i+1} (1-p)^{2i+1} p^{n-2i-1}.$$

Consequently, we obtain

$$\begin{aligned} \alpha_n &= \sup_{(A,B)\in\mathcal{G}_1\times\mathcal{G}_{n+1}} \left| P(A\cap B) - \frac{1}{4} \right| \\ &= \frac{1}{2}\max\left\{ \left| G(n) - \frac{1}{2} \right|, \left| U(n) - \frac{1}{2} \right| \right\} = \frac{1}{2}\left| G(n) - \frac{1}{2} \right|. \end{aligned}$$

Now, we have

$$\left| G(n) - \frac{1}{2} \right| = \frac{1}{2}\left| G(n) - U(n) \right| = \frac{1}{2}\left|2p-1\right|^n ,$$

where the last equation is due to the binomial theorem. The claim now follows. □

Proposition 6.6. *Let $n \in \mathbb{N}$, $1 \le i_1 \le \cdots \le i_n$ and $l \in \{1,\dots,n-1\}$ and let $f : \{-1,1\}^l \to \mathbb{R}$ and $g : \{-1,1\}^{n-l} \to \mathbb{R}$. We set $F := \|f\|_\infty$ and $G := \|g\|_\infty$. Then, there exists a constant $\beta > 0$ such that*

$$\begin{aligned}&\left|E\left[f(X_{i_1},\dots,X_{i_l})g(X_{i_{l+1}},\dots,X_{i_n})\right] - E\left[f(X_{i_1},\dots,X_{i_l})\right]E\left[g(X_{i_{l+1}},\dots,X_{i_n})\right]\right|\\ &\le FG\exp(-\beta(i_{l+1}-i_l)). \qquad (6.2)\end{aligned}$$

In particular, for $1 \le i_1 \le i_2$, we have

$$|E\left[X_{i_1}X_{i_2}\right]| \le \exp(-\beta(i_2-i_1)). \qquad (6.3)$$

For the proof of Proposition 6.6, we will use the following bound for the covariance of two bounded random variables:

Lemma 6.7 ([2], Th. 1.11)**.** *Let X and Y be two real-valued random variables satisfying $|X| \le C$ a.s. and $|Y| \le D$ a.s. for two constants $C \ge 0$, $D \ge 0$. Then, we have*[1]

$$|E[XY] - E[X]E[Y]| \le 4CD\ \alpha(\sigma(X),\sigma(Y)).$$

Proof of Proposition 6.6. Let $n \in \mathbb{N}$ and $l \in \{1,\dots,n-1\}$ be fixed. Applying Lemma 6.7 to our setting gives

$$\begin{aligned}&\left|E\left[f(X_{i_1},\dots,X_{i_l})g(X_{i_{l+1}},\dots,X_{i_n})\right] - E\left[f(X_{i_1},\dots,X_{i_l})\right]E\left[g(X_{i_{l+1}},\dots,X_{i_n})\right]\right|\\ &\le 4FG\ \alpha\left(\sigma\left(f(X_{i_1},\dots,X_{i_l})\right),\sigma\left(g(X_{i_{l+1}},\dots,X_{i_n})\right)\right). \qquad (6.4)\end{aligned}$$

We have

$$\begin{aligned}&\alpha\left(\sigma\left(f(X_{i_1},\dots,X_{i_l})\right),\sigma\left(g(X_{i_{l+1}},\dots,X_{i_n})\right)\right)\\ &\le \alpha(\sigma(X_{i_1},\dots,X_{i_l}),\sigma(X_{i_{l+1}},\dots,X_{i_n})) \qquad (6.5)\\ &= \alpha(\sigma(X_{i_l}),\sigma(X_{i_{l+1}})) \qquad (6.6)\\ &= \alpha(\sigma(X_1),\sigma(X_{i_{l+1}-i_l+1})) \qquad (6.7)\\ &= \frac{1}{4}\,|2p-1|^{i_{l+1}-i_l}. \qquad (6.8)\end{aligned}$$

In (6.5), we have used measurability of the functions f and g and the fact that for σ-algebras $\mathcal{A}_0 \subset \mathcal{A}$, $\mathcal{B}_0 \subset \mathcal{B}$, we have $\alpha(\mathcal{A}_0,\mathcal{B}_0) \le \alpha(\mathcal{A},\mathcal{B})$. (6.6) follows from the Markov property and (6.7) is due to stationarity of the Markov chain $(X_i)_{i\ge1}$.
Combining (6.4) and (6.8) and setting $\beta := -\log(|2p-1|)$ gives the claim. □

[1] α denotes the strong mixing coefficient.

6.3 Proof

The strategy is as follows: We will first prove that the sequence $(Y^{(n)})_{n\geq 1}$ is tight in $C^{0,\alpha-H\ddot{o}l}([0,1],\mathbb{R})$ for any $\alpha \in (0,1/2)$. By Prohorov's theorem, $(Y^{(n)})_{n\geq 1}$ is thus relatively compact in $C^{0,\alpha-H\ddot{o}l}([0,1],\mathbb{R})$. Hence, it suffices to show that an arbitrary weakly convergent subsequence of $(Y^{(n)})_{n\geq 1}$ has limit σB. In this step of the proof, the martingale problem formulation will come into play.

6.3.1 Step 1: Tightness

The following argumentation is taken from [6][2]. We will take advantage of

Proposition 6.8 (Lamperti's tightness criterion, [6], Cor. 1)**.** *Let* $\left(\{Y_t^{(n)} : t \in [0,1]\}\right)_{n\geq 1}$ *be a sequence of real-valued stochastic processes vanishing at* 0*. Assume there exist positive constants* a, b *and* c *such that for all* s*,* $t \in [0,1]$ *and for all* $n \in \mathbb{N}$

$$E\left[\left|Y_t^{(n)} - Y_s^{(n)}\right|^a\right] \leq c\,|t-s|^{1+b}. \tag{6.9}$$

Then, the sequence $(Y^{(n)})_{n\geq 1}$ *is tight in* $C^{0,\gamma-H\ddot{o}l}([0,1],\mathbb{R})$ *for any* $\gamma \in \left(0, \frac{b}{a}\right)$*.*

Remark 6.9. Note that the above proposition includes the statement that the sample paths of $Y^{(n)}$ are a.s. in $C^{0,\gamma-H\ddot{o}l}([0,1],\mathbb{R})$. Hence, we do not have to show this fact separately.

Our aim is now to establish (6.9). We fix $a > 2$. We first consider the case where s and t lie in the same interpolation interval.

Case 1: $\exists j \in \{0,1,\ldots,n-1\}$: $\frac{j}{n} \leq s \leq t \leq \frac{j+1}{n}$.

We have

$$\left|Y_t^{(n)} - Y_s^{(n)}\right| = \left|\frac{1}{\sqrt{n}}(nt-ns)X_{j+1}\right| = \sqrt{n}\,|t-s|\,|X_{j+1}|.$$

This implies that for $\epsilon > 0$

$$\begin{aligned} E\left[\left|Y_t^{(n)} - Y_s^{(n)}\right|^a\right] &= n^{a/2}\,|t-s|^a\,E\left[|X_1|^a\right] \\ &\leq |t-s|^{a/2}\,E\left[|X_1|^{a+\epsilon}\right]^{a/(a+\epsilon)}, \end{aligned}$$

where the last inequality is due to Hölder's inequality and the fact that $n\,|t-s| \leq 1$.

[2] proof of Theorem 17.

Case 2: $s \neq t \wedge \ (\exists j \in \{1, \dots, n-1\}, \exists k \in \{0, 1, \dots, n-1\} : j+k+1 \leq n$
$\wedge \ \frac{j-1}{n} \leq s \leq \frac{j}{n} \ \wedge \ \frac{j+k}{n} \leq t \leq \frac{j+k+1}{n})$.
Then,

$$\begin{aligned}
& E\left[\left|Y_t^{(n)} - Y_s^{(n)}\right|^a\right] \\
\leq\ & E\left[\left(\left|Y_t^{(n)} - Y_{(j+k)/n}^{(n)}\right| + \left|Y_{(j+k)/n}^{(n)} - Y_{j/n}^{(n)}\right| + \left|Y_{j/n}^{(n)} - Y_s^{(n)}\right|\right)^a\right] \\
\leq\ & 3^{a-1}\left(E\left[\left|Y_t^{(n)} - Y_{(j+k)/n}^{(n)}\right|^a\right] + E\left[\left|Y_{(j+k)/n}^{(n)} - Y_{j/n}^{(n)}\right|^a\right] + E\left[\left|Y_{j/n}^{(n)} - Y_s^{(n)}\right|^a\right]\right).
\end{aligned} \tag{6.10}$$

In (6.10), we have used convexity of the function x^a on $\mathbb{R}_{\geq 0}$ for $a > 2$. For the first and the last expectation in (6.10), we have the upper bound derived in Case 1. To find an upper bound for the expectation in the middle if $k \neq 0$, we will apply Yokoyama's moment bound (Theorem 5.3). In view of the assumptions of this theorem, note that we have already seen in Lemma 6.5 that $\alpha_i = \frac{1}{4}|2p-1|^i$ and the sequence $(X_i)_{i\geq 1}$ is strong mixing. One can check that

$$\sum_{i=0}^{\infty}(i+1)\left(\frac{1}{4}|2p-1|^i\right)^{1/5} < \infty,$$

i.e. for $r = 4$ and $\delta = 1$, assumption (5.1) holds.
Consequently, there exists a constant $C > 0$ such that

$$\begin{aligned}
E\left[\left|Y_{(j+k)/n}^{(n)} - Y_{j/n}^{(n)}\right|^a\right] &= E\left[\left|\frac{1}{\sqrt{n}}(X_{j+1} + X_{j+2} + \cdots + X_{j+k})\right|^a\right] \\
&\leq C\left(\frac{k}{n}\right)^{a/2} \leq C\,|t-s|^{a/2},
\end{aligned} \tag{6.11}$$

where the first inequality in (6.11) is due to Yokoyama's moment bound.

Hence, if we choose $b = \frac{a}{2} - 1$ and $c = 3^{a-1}\left(2E\left[|X_1|^{a+\epsilon}\right]^{a/(a+\epsilon)} + C\right)$, then (6.9) will hold for all $s, t \in [0,1]$, so that $\left(Y^{(n)}\right)_{n\geq 1}$ is tight in $C^{0,\alpha-Höl}([0,1],\mathbb{R})$ for any $\alpha \in (0, 1/2)$. $\square$

6.3.2 Step 2: Characterization of the limit process

In this step[3], we use some ideas presented in [19] and [20], Chapter 11.
Let $\left(Y^{(n_i)}\right)_{i\geq 1}$ be subsequence of $(Y^{(n)})_{n\geq 1}$ that converges weakly to a limit process Y. In the sequel, we will always work with this subsequence, but we will denote it by $(Y^{(n)})_{n\geq 1}$ to simplify the notation.
As already mentioned in the introduction in connection with Donsker's theorem, the stochastic process $Y^{(n)}$ can be seen as a map $\Omega \to (C([0,1],\mathbb{R})$. Since the Borel σ-algebra $\mathcal{M}$ on $(C([0,1],\mathbb{R})$ is generated by the one-dimensional cylinder sets and $Y_t^{(n)}$

[3]and in the respective steps of the subsequent proofs.

is a random variable for each $t \in [0,1]$, $Y^{(n)}$ is $\mathcal{F}-\mathcal{M}$-measurable i.e. it is a random variable taking values in $(C([0,1],\mathbb{R}),\mathcal{M})$. It is important to understand that in the following proof, we will switch between the two probability spaces $(\Omega,\mathcal{F},P)$ and $\big(C([0,1],\mathbb{R}),\mathcal{M},P(Y^{(n)})^{-1}\big)$ (or $\big(C([0,1],\mathbb{R}),\mathcal{M},PY^{-1}\big)$, respectively).

Let $a,b:\mathbb{R}\to\mathbb{R}$ be given by $a(x)=\sigma^2=\frac{p}{1-p}$ and $b(x)=0$ for all $x\in\mathbb{R}$.
Our aim is to prove the following

Claim: The probability measure PY^{-1} is the unique solution to the martingale problem for (a,b) starting at 0 i.e. $P(Y_0=0)=1$ and for each $f\in C_0^\infty(\mathbb{R})$

$$\left(f(x(t))-f(x(0))-\frac{1}{2}\sigma^2\int_s^t f''(x(u))\,du\right)_{t\in[0,1]} \tag{6.12}$$

is a PY^{-1}-martingale with respect to $(\mathcal{M}_t)_{t\in[0,1]}$.

It then follows from Theorem 5.10 that PY^{-1} is the law of σB, where B is a standard Brownian motion.

Proof of the claim:

By Theorem 5.13, uniqueness is clear. As weak convergence in $C([0,1],\mathbb{R})$ implies convergence of the finite-dimensional distributions[4], $P(Y_0=0)=1$ follows from $P(Y_0^{(n)}=0)=1$ for all $n\in\mathbb{N}$.
We turn to the proof of (6.12). We fix $f\in C_0^\infty(\mathbb{R})$, $s,t\in[0,1]$, $s<t$ and a continuous, bounded, $\mathcal{M}_s$-measurable map $\phi: C([0,1],\mathbb{R})\to\mathbb{R}$. By Lemma 5.5, it is enough to show that

$$E_{PY^{-1}}\left[\left(f(x(t))-f(x(s))-\frac{1}{2}\sigma^2\int_s^t f''(x(u))\,du\right)\phi\right]=0.$$

By the transformation theorem for image measures (Theorem A.4), this is equivalent to

$$E_P\left[\left(f(Y_t)-f(Y_s)-\frac{1}{2}\sigma^2\int_s^t f''(Y_u)\,du\right)\phi(Y)\right]=0. \tag{6.13}$$

We will prove that for $\epsilon\in(0,t-s)$, we have

$$E_P\left[\left(f(Y_t)-f(Y_{s+\epsilon})-\frac{1}{2}\sigma^2\int_{s+\epsilon}^t f''(Y_u)\,du\right)\phi(Y)\right]=0. \tag{6.14}$$

Since the stochastic process Y is continuous and the functions f and f'' are bounded and continuous, Lebesgue's dominated convergence theorem implies that the left-hand side of (6.14) converges to the left-hand side of (6.13) as $\epsilon\to 0$. The following argumentation thus aims at establishing (6.14).
For the rest of the proof, we fix $\epsilon\in(0,t-s)$ and we set $\bar{s}:=s+\epsilon$.

[4]This is due to the continuous mapping theorem, as the maps of the form $\pi_{t_1,\dots,t_m}$ introduced in Definition 3.5 are continuous.

It is easy to see that for fixed $n \in \mathbb{N}$

$$f\left(x\left(\frac{k}{n}\right)\right) - \sum_{j=1}^{k} \left(E_{P(Y^{(n)})^{-1}} \left[f\left(x\left(\frac{j}{n}\right)\right) \,\middle|\, \mathcal{M}_{(j-1)/n} \right] - f\left(x\left(\frac{j-1}{n}\right)\right)\right),$$

$1 \leq k \leq n$, is a (discrete-parameter) $P\left(Y^{(n)}\right)^{-1}$-martingale with respect to the filtration $(\mathcal{M}_{k/n})_{1\leq k\leq n}$. As ϕ is $\mathcal{M}_{\lceil n\bar{s}\rceil/n}$-measurable, this implies (we use Lemma 5.5 again)

$$\begin{aligned} &E_{P(Y^{(n)})^{-1}} \left[\left(f\left(x\left(\frac{\lceil nt\rceil}{n}\right)\right) - f\left(x\left(\frac{\lceil n\bar{s}\rceil}{n}\right)\right) \right.\right. \\ &\left.\left. - \sum_{j=\lceil n\bar{s}\rceil+1}^{\lceil nt\rceil} \left(E_{P(Y^{(n)})^{-1}} \left[f\left(x\left(\frac{j}{n}\right)\right) \,\middle|\, \mathcal{M}_{(j-1)/n} \right] - f\left(x\left(\frac{j-1}{n}\right)\right)\right)\right)\phi \right] = 0. \end{aligned} \tag{6.15}$$

Now, we will apply the transformation theorem for image measures again. For all the above terms except the conditional expectations, this is easy. We now focus on the above conditional expectations. We define $\mathcal{F}_0 := \{\emptyset, \Omega\}$ and

$$\mathcal{F}_j := \sigma\left(X_i : 1 \leq i \leq j\right) \quad \text{for } j \in \{1, \ldots, n\}.$$

Note that we have

$$\mathcal{F}_j = \sigma\left(Y_u^{(n)} : 0 \leq u \leq \frac{j}{n}\right).$$

Let $j \in \{\lceil n\bar{s}\rceil + 1, \ldots, \lceil nt\rceil\}$. We will use that ϕ is $\mathcal{M}_{(j-1)/n}$-measurable and $\phi\left(Y^{(n)}\right)$ is $\mathcal{F}_{j-1}$-measurable to obtain

$$\begin{aligned} &\int E_{P(Y^{(n)})^{-1}} \left[f\left(x\left(\frac{j}{n}\right)\right) \,\middle|\, \mathcal{M}_{(j-1)/n} \right] \phi \, dP\left(Y^{(n)}\right)^{-1} \\ &= \int f\left(x\left(\frac{j}{n}\right)\right) \phi \, dP\left(Y^{(n)}\right)^{-1} \\ &= \int f\left(Y_{j/n}^{(n)}\right) \phi\left(Y^{(n)}\right) dP \\ &= \int E_P\left[f\left(Y_{j/n}^{(n)}\right) \,\middle|\, \mathcal{F}_{j-1}\right] \phi\left(Y^{(n)}\right) dP. \end{aligned}$$

Consequently, (6.15) is equivalent to

$$\begin{aligned} &E_P\left[\left(f\left(Y_{\lceil nt\rceil/n}^{(n)}\right) - f\left(Y_{\lceil n\bar{s}\rceil/n}^{(n)}\right) \right.\right. \\ &\left.\left. - \sum_{j=\lceil n\bar{s}\rceil+1}^{\lceil nt\rceil} \left(E_P\left[f\left(Y_{j/n}^{(n)}\right) \,\middle|\, \mathcal{F}_{j-1}\right] - f\left(Y_{(j-1)/n}^{(n)}\right)\right)\right) \phi\left(Y^{(n)}\right)\right] = 0. \end{aligned} \tag{6.16}$$

In the rest of the proof, we will show that the left-hand side of (6.16) converges to the left-hand side of (6.14) as $n \to \infty$, which will imply that (6.14) holds.
In the following, all (conditional and unconditional) expectations will be taken with respect to the probability measure P. We will thus omit P in the notation for expectations.
Since $Y^{(n)}$ converges weakly to Y and for $u \in [0,1]$, the map $h \in C([0,1],\mathbb{R}) \mapsto f(h_u)\phi(h)$ is continuous, $f(Y_u^{(n)})\phi\left(Y^{(n)}\right)$ converges weakly to $f(Y_u)\phi(Y)$ by the continuous mapping theorem. Thus, as f and ϕ are bounded functions, we have

$$\lim_{n\to\infty} E\left[\left(f\left(Y^{(n)}_{\lceil nt\rceil/n}\right) - f\left(Y^{(n)}_{\lceil n\bar{s}\rceil/n}\right)\right)\phi\left(Y^{(n)}\right)\right] = E\left[\left(f\left(Y_t\right) - f\left(Y_{\bar{s}}\right)\right)\phi(Y)\right].$$

We define $\phi_Y := \phi(Y)$ and $\phi_n := \phi\left(Y^{(n)}\right)$ for $n \in \mathbb{N}$. What remains to be shown is

$$\lim_{n\to\infty} E\left[\sum_{j=\lceil n\bar{s}\rceil+1}^{\lceil nt\rceil}\left(E\left[f\left(Y^{(n)}_{j/n}\right)\;\middle|\;\mathcal{F}_{j-1}\right] - f\left(Y^{(n)}_{(j-1)/n}\right)\right)\phi_n\right]$$
$$= \frac{1}{2}\sigma^2 E\left[\left(\int_{\bar{s}}^{t} f''(Y_u)\,du\right)\phi_Y\right]. \tag{6.17}$$

For this, we define $S_k := \sum_{i=1}^k X_i$ for $k \in \mathbb{N}$, $S_0 := 0$ and we consider the Taylor expansion of $E[f(Y^{(n)}_{j/n})|\mathcal{F}_{j-1}]$ for $j \in \mathbb{N}$. Using that S_{j-1} is $\mathcal{F}_{j-1}$-measurable, we obtain

$$E\left[f\left(Y^{(n)}_{j/n}\right)\;\middle|\;\mathcal{F}_{j-1}\right] = E\left[f\left(\frac{S_{j-1}}{\sqrt{n}} + \frac{X_j}{\sqrt{n}}\right)\;\middle|\;\mathcal{F}_{j-1}\right]$$
$$= f\left(\frac{S_{j-1}}{\sqrt{n}}\right) + \frac{1}{\sqrt{n}}f'\left(\frac{S_{j-1}}{\sqrt{n}}\right)E\left[X_j \mid \mathcal{F}_{j-1}\right] + \frac{1}{2n}f''\left(\frac{S_{j-1}}{\sqrt{n}}\right)E\left[X_j^2 \mid \mathcal{F}_{j-1}\right]$$
$$+ E\left[R_3\left(\frac{S_j}{\sqrt{n}}\right)\;\middle|\;\mathcal{F}_{j-1}\right] \quad \text{a.s.,} \tag{6.18}$$

where R_3 denotes the remainder term[5] of the Taylor expansion. Observe that f''' is bounded and $\left|X_i^3\right| = 1$, so that we have

$$\left|E\left[E\left[R_3\left(\frac{S_j}{\sqrt{n}}\right)\;\middle|\;\mathcal{F}_{j-1}\right]\phi_n\right]\right| \le \frac{\|f'''\|_\infty\,\|\phi\|_\infty}{n^{3/2}} = O\left(n^{-3/2}\right).$$

More generally, if for a summand $s(j)$ in the Taylor expansion, we have $E[s(j)\phi_n] = o\left(1/n\right)$ for all $j \in \{1,\dots,n\}$, then $\lim_{n\to\infty}\sum_{j=\lceil n\bar{s}\rceil+1}^{\lceil nt\rceil} E[s(j)\phi_n] = 0$. This fact will be used frequently in the sequel. Here, we only need to consider the first three terms in (6.18).
We clearly have $E[X_j^2|\mathcal{F}_{j-1}] = 1$. Using the Markov property of $(X_i)_{i\ge1}$ and the given transition probabilities for this Markov chain, one can check that

$$E[X_j|\mathcal{F}_{j-1}] = (2p-1)X_{j-1}.$$

[5] we consider its Lagrange form.

Hence, we have

$$\begin{aligned}
&\lim_{n\to\infty} E\left[\sum_{j=\lceil n\bar{s}\rceil+1}^{\lceil nt\rceil}\left(E\left[f\left(Y^{(n)}_{j/n}\right)\,\Big|\,\mathcal{F}_{j-1}\right]-f\left(Y^{(n)}_{(j-1)/n}\right)\right)\phi_n\right]\\
&=(2p-1)\lim_{n\to\infty}\sum_{j=\lceil n\bar{s}\rceil+1}^{\lceil nt\rceil} E\left[\frac{1}{\sqrt{n}}f'\left(\frac{S_{j-1}}{\sqrt{n}}\right)X_{j-1}\ \phi_n\right] \qquad (6.19)\\
&\quad+\frac{1}{2}\lim_{n\to\infty}\sum_{j=\lceil n\bar{s}\rceil+1}^{\lceil nt\rceil} E\left[\frac{1}{n}f''\left(\frac{S_{j-1}}{\sqrt{n}}\right)\phi_n\right].
\end{aligned}$$

The difficulty in expression (6.19) is that S_{j-1} and X_{j-1} are dependent random variables. To handle this difficulty, we set $L(n) := \lfloor sn^{1/10}\rfloor$ and we will do Taylor expansion of $f'\left(\frac{S_{j-1}}{\sqrt{n}}\right)$ with expansion point $\frac{S_{j-L(n)}}{\sqrt{n}}$. Then, we will apply the covariance estimate (6.2) to the resulting summands in such a way that applying $\lim_{n\to\infty}\sum_{j=\lceil n\bar{s}\rceil+1}^{\lceil nt\rceil}$ to the error terms yields 0. We claim that

$$\begin{aligned}
&\lim_{n\to\infty}\sum_{j=\lceil n\bar{s}\rceil+1}^{\lceil nt\rceil} E\left[\frac{1}{\sqrt{n}}f'\left(\frac{S_{j-1}}{\sqrt{n}}\right)X_{j-1}\ \phi_n\right]\\
&=\frac{1}{2(1-p)}\lim_{n\to\infty}\sum_{j=\lceil n\bar{s}\rceil+1}^{\lceil nt\rceil} E\left[\frac{1}{n}f''\left(\frac{S_{j-1}}{\sqrt{n}}\right)\phi_n\right]. \qquad (6.20)
\end{aligned}$$

To establish (6.20), we will use the two auxiliary results below. The first says that shifting indices by $L(n)$ gives an error term which tends to 0 as $n\to\infty$. As the functions f'' and ϕ are bounded, it is easy to see that

Lemma 6.10.

$$\left|\sum_{j=\lceil n\bar{s}\rceil+1}^{\lceil nt\rceil}\frac{1}{n}E\left[f''\left(\frac{S_{j-L(n)}}{\sqrt{n}}\right)\phi_n\right]-\sum_{j=\lceil n\bar{s}\rceil+1}^{\lceil nt\rceil}\frac{1}{n}E\left[f''\left(\frac{S_{j-1}}{\sqrt{n}}\right)\phi_n\right]\right| = O\left(\frac{L(n)}{n}\right).$$

For the second auxiliary result, we will use the fact that the stationary Markov chain $(X_i)_{i\geq 1}$ can be extended to a two-sided stationary Markov chain $(X_i)_{-\infty<i<\infty}$.

Lemma 6.11. *For $j\in\mathbb{N}$, we have*

$$\lim_{n\to\infty}\sum_{k=\lfloor L(n)/2\rfloor}^{L(n)-1} E\left[X_{j-L(n)+k}X_{j-1}\right]=\frac{1}{2(1-p)}.$$

Proof. We have

$$\sum_{k=\lfloor L(n)/2\rfloor}^{L(n)-1} E\left[X_{j-L(n)+k}X_{j-1}\right] = \sum_{i=1}^{L(n)-\lfloor L(n)/2\rfloor} E\left[X_1X_i\right] \tag{6.21}$$

$$= \sum_{i=0}^{\infty} E[X_1X_i] - \sum_{i=L(n)-\lfloor L(n)/2\rfloor+1}^{\infty} E[X_1X_i],$$

where in (6.21), we have used stationarity. Now, it follows from the covariance estimate (6.3) that

$$\lim_{n\to\infty}\left|\sum_{i=L(n)-\lfloor L(n)/2\rfloor+1}^{\infty} E[X_1X_i]\right| \le \lim_{n\to\infty}\sum_{i=L(n)-\lfloor L(n)/2\rfloor}^{\infty} \exp(-\beta i) = 0.$$

This implies

$$\lim_{n\to\infty}\sum_{k=\lfloor L(n)/2\rfloor}^{L(n)-1} E\left[X_{j-L(n)+k}X_{j-1}\right] = \sum_{i=1}^{\infty} E[X_1X_i].$$

For $i \in \mathbb{N}$, we have

$$\begin{aligned} E[X_1X_{i+1}] =& 2P(X_{i+1}=1|X_1=1)P(X_1=1) - 2P(X_{i+1}=-1|X_1=1)P(X_1=1) \\ =& (2p-1)^i, \end{aligned}$$

where the last equality is obtained by a similar argument as in the proof of (6.1) on page 56. Consequently, as $|2p-1|<1$, we get

$$\sum_{i=1}^{\infty} E[X_1X_i] = \sum_{i=0}^{\infty}(2p-1)^i = \frac{1}{2(1-p)}.$$

$\square$

Proof of (6.20). As already mentioned above, we first do Taylor expansion, which gives:

$$\begin{aligned} & f'\left(\frac{S_{j-1}}{\sqrt{n}}\right)\frac{X_{j-1}}{\sqrt{n}} \\ =& \left[f'\left(\frac{S_{j-L(n)}}{\sqrt{n}}\right) + f''\left(\frac{S_{j-L(n)}}{\sqrt{n}}\right)\left(\frac{X_{j-L(n)+1}+\cdots+X_{j-1}}{\sqrt{n}}\right) + R_2\left(\frac{S_{j-1}}{\sqrt{n}}\right)\right]\frac{X_{j-1}}{\sqrt{n}}. \end{aligned} \tag{6.22}$$

We have

$$\left|E\left[R_2\left(\frac{S_{j-1}}{\sqrt{n}}\right)\frac{X_{j-1}}{\sqrt{n}}\ \phi_n\right]\right| \le \|f'''\|_\infty \|\phi\|_\infty \frac{(L(n))^2}{n^{3/2}} = o\left(\frac{1}{n}\right),$$

so that we do not have to consider this term.
We now show how the second summand in (6.22) can be handled. The argument for the first summand is similar, but simpler. We split the sum

$$\sum_{j=\lceil n\bar{s}\rceil+1}^{\lceil nt\rceil} E\left[f''\left(\frac{S_{j-L(n)}}{\sqrt{n}}\right)\left(\frac{X_{j-L(n)+1}+\cdots+X_{j-1}}{\sqrt{n}}\right)\frac{X_{j-1}}{\sqrt{n}}\ \phi_n\right] \tag{6.23}$$

into two parts, to which we apply the covariance estimate (6.2) in different ways. We would like to point out that the following argumentation would not work if $\bar{s}=s+\epsilon$ were replaced by s.
Note that the function $\phi_n=\phi\circ Y^{(n)}$ depends on the random variables $X_1,\ldots,X_{\lceil ns\rceil}$. For n large enough, we have $L(n)\leq\lceil n\bar{s}\rceil-\lceil ns\rceil$, so that ϕ_n only depends on the random variables $X_1,\ldots,X_{\lceil n\bar{s}\rceil-L(n)}$. We will use this fact in the following covariance bound applications.
We set $K:=\|f''\|_\infty\|\phi\|_\infty$.
We begin with the lower part of the sum (6.23). For n large enough, we have[6]

$$\begin{aligned}
&\sum_{j=\lceil n\bar{s}\rceil+1}^{\lceil nt\rceil}\ \sum_{k=1}^{\lfloor L(n)/2\rfloor-1}\frac{1}{n}\left|E\left[f''\left(\frac{S_{j-L(n)}}{\sqrt{n}}\right)X_{j-L(n)+k}\ X_{j-1}\ \phi_n\right]\right|\\
\leq&\sum_{j=\lceil n\bar{s}\rceil+1}^{\lceil nt\rceil}\ \sum_{k=1}^{\lfloor L(n)/2\rfloor-1}\frac{1}{n}\left(\left|E\left[f''\left(\frac{S_{j-L(n)}}{\sqrt{n}}\right)X_{j-L(n)+k}\ \phi_n\right]\underbrace{E\left[X_{j-1}\right]}_{=0}\right|\right.\\
&\left.+K\exp(-\beta(L(n)-k-1))\right)\\
\leq&\frac{(\lceil nt\rceil-\lceil n\bar{s}\rceil)}{n}\sum_{k=L(n)-\lfloor L(n)/2\rfloor}^{L(n)-2}K\exp(-\beta k)\xrightarrow[n\to\infty]{}0.
\end{aligned}\tag{6.25}$$

We now turn to the upper part of the sum (6.23). For n large enough and $k\in\{\lfloor L(n)/2\rfloor,\ldots,L(n)-1\}$, we have

$$\begin{aligned}
&\left|\frac{1}{n}E\left[f''\left(\frac{S_{j-L(n)}}{\sqrt{n}}\right)X_{j-L(n)+k}\ X_{j-1}\ \phi_n\right]\right.\\
&\left.-\frac{1}{n}E\left[f''\left(\frac{S_{j-L(n)}}{\sqrt{n}}\right)\phi_n\right]E\left[X_{j-L(n)+k}\ X_{j-1}\right]\right|\leq\frac{K}{n}\exp(-\beta k).
\end{aligned}$$

Note that

$$\lim_{n\to\infty}\sum_{j=\lceil n\bar{s}\rceil+1}^{\lceil nt\rceil}\ \sum_{k=\lfloor L(n)/2\rfloor}^{L(n)-1}\frac{K}{n}\exp(-\beta k)=0.$$

[6]A similar argument applies to the first summand in (6.22) and yields

$$\lim_{n\to\infty}\frac{1}{\sqrt{n}}E\left[\sum_{j=\lceil n\bar{s}\rceil+1}^{\lceil nt\rceil}f'\left(\frac{S_{j-L(n)}}{\sqrt{n}}\right)X_{j-1}\ \phi_n\right]=0. \tag{6.24}$$

Consequently,

$$
\begin{aligned}
&\lim_{n\to\infty} \sum_{j=\lceil n\bar{s}\rceil+1}^{\lceil nt\rceil} \sum_{k=\lfloor L(n)/2\rfloor}^{L(n)-1} \frac{1}{n} E\left[f''\left(\frac{S_{j-L(n)}}{\sqrt{n}}\right) X_{j-L(n)+k}\ X_{j-1}\ \phi_n\right] \\
=\ &\lim_{n\to\infty} \sum_{j=\lceil n\bar{s}\rceil+1}^{\lceil nt\rceil} \frac{1}{n} E\left[f''\left(\frac{S_{j-L(n)}}{\sqrt{n}}\right) \phi_n\right] \sum_{k=\lfloor L(n)/2\rfloor}^{L(n)-1} E\left[X_{j-L(n)+k}\ X_{j-1}\right] \\
=\ &\frac{1}{2(p-1)} \lim_{n\to\infty} \sum_{j=\lceil n\bar{s}\rceil+1}^{\lceil nt\rceil} \frac{1}{n} E\left[f''\left(\frac{S_{j-1}}{\sqrt{n}}\right) \phi_n\right], \qquad (6.26)
\end{aligned}
$$

where in (6.26), we have applied Lemma 6.10 and Lemma 6.11.
The claim is now obtained by combining (6.24), (6.25) and (6.26).

□

In the next step, we will replace the expectation of the sum of the random variables $f''\left(\frac{S_{j-1}}{\sqrt{n}}\right)$, $j = \lceil n\bar{s}\rceil + 1, \ldots, \lceil nt\rceil$ by the expectation of the integral of the corresponding linearly interpolated process $Y^{(n)}$ and an error term which tends to 0 as $n \to \infty$. More precisely, we will use the following result:

Lemma 6.12.

$$
E\left[\left|\frac{1}{n} \sum_{j=\lceil n\bar{s}\rceil+1}^{\lceil nt\rceil} f''\left(\frac{S_{j-1}}{\sqrt{n}}\right) - \int_{\bar{s}}^{t} f''\left(Y_u^{(n)}\right)\ du\right|\right] = O\left(\frac{1}{\sqrt{n}}\right).
$$

Proof. We fix $n \in \mathbb{N}$ and define the stochastic process $Z^{(n)}$ on the interval $(\frac{\lceil n\bar{s}\rceil}{n}, \frac{\lceil nt\rceil}{n}]$ by

$$
Z_u^{(n)} = \frac{S_{\lceil n\bar{s}\rceil+k}}{\sqrt{n}} \quad \text{for } u \in \left(\frac{\lceil n\bar{s}\rceil + k}{n}, \frac{\lceil n\bar{s}\rceil + k + 1}{n}\right],\ k = 0, \ldots, \lceil nt\rceil - \lceil n\bar{s}\rceil - 1.
$$

Let $K := \sup_{t\in\mathbb{R}} |f'''(t)|$, which is a Lipschitz constant for f''. We will use that for $u \in (\bar{s}, \bar{s} + \frac{1}{n}]$, $Y_u^{(n)}$ is the linear interpolation between $\frac{S_{\lceil n\bar{s}\rceil}}{\sqrt{n}}$ and $\frac{S_{\lceil n\bar{s}\rceil+1}}{\sqrt{n}}$ to obtain

$$
\left|f''\left(Z_u^{(n)}\right) - f''\left(Y_u^{(n)}\right)\right| \le K \left|\frac{S_{\lceil n\bar{s}\rceil+1}}{\sqrt{n}} - \frac{S_{\lceil n\bar{s}\rceil}}{\sqrt{n}}\right| = \frac{K}{\sqrt{n}}.
$$

For the other intervals of the form $(\bar{s} + \frac{k}{n}, \bar{s} + \frac{k+1}{n}]$, $k \in \{1, \ldots, \lceil nt\rceil - \lceil n\bar{s}\rceil - 1\}$, the same argumentation holds true.

Hence,

$$\begin{aligned}
&\left|\frac{1}{n}\sum_{j=\lceil n\bar{s}\rceil+1}^{\lceil nt\rceil} f''\left(\frac{S_{j-1}}{\sqrt{n}}\right) - \int_{\bar{s}}^{t} f''\left(Y_u^{(n)}\right) du\right| \\
= &\left|\int_{\lceil n\bar{s}\rceil/n}^{\lceil nt\rceil/n} f''\left(Z_u^{(n)}\right) du - \int_{\bar{s}}^{t} f''\left(Y_u^{(n)}\right) du\right| \\
\leq &\left|\int_{\lceil n\bar{s}\rceil/n}^{(\lceil nt\rceil-1)/n} \left(f''\left(Z_u^{(n)}\right) - f''\left(Y_u^{(n)}\right)\right) du\right| + \left|\int_{(\lceil nt\rceil-1)/n}^{\lceil nt\rceil/n} f''\left(Z_u^{(n)}\right) du\right| \\
&+ \left|\int_{\bar{s}}^{\lceil n\bar{s}\rceil/n} f''\left(Y_u^{(n)}\right) du\right| + \left|\int_{(\lceil nt\rceil-1)/n}^{t} f''\left(Y_u^{(n)}\right) du\right| \\
\leq &\sum_{k=\lceil n\bar{s}\rceil}^{\lceil nt\rceil-2} \left|\int_{k/n}^{(k+1)/n} \left(f''(Z_u^{(n)}) - f''\left(Y_u^{(n)}\right)\right) du\right| + \frac{3}{n}\left\|f''\right\|_\infty \\
\leq &\ (\lceil nt\rceil - \lceil n\bar{s}\rceil)\frac{K}{n\sqrt{n}} + \frac{3}{n}\left\|f''\right\|_\infty \leq \frac{K + 3\left\|f''\right\|_\infty}{\sqrt{n}}.
\end{aligned}$$

This implies the claim. □

Let us write

$$R(f) := \frac{1}{n}\sum_{j=\lceil n\bar{s}\rceil+1}^{\lceil nt\rceil} f''\left(\frac{S_{j-1}}{\sqrt{n}}\right) - \int_{\bar{s}}^{t} f''\left(Y_u^{(n)}\right) du.$$

By Lemma 6.12,

$$|E[R(f)\phi_n]| \leq \|\phi\|_\infty E[|R(f)|] = O(n^{-1/2}).$$

Consequently, we obtain (6.17):

$$\begin{aligned}
&\lim_{n\to\infty} E\left[\sum_{j=\lceil n\bar{s}\rceil+1}^{\lceil nt\rceil} \left(E\left[f\left(Y_{j/n}^{(n)}\right) \,\middle|\, \mathcal{F}_{j-1}\right] - f\left(Y_{(j-1)/n}^{(n)}\right)\right)\phi_n\right] \\
&= \frac{1}{2}\underbrace{\left(1 + \frac{2p-1}{1-p}\right)}_{=\sigma^2} \lim_{n\to\infty} E\left[\sum_{j=\lceil n\bar{s}\rceil+1}^{\lceil nt\rceil} \frac{1}{n} f''\left(\frac{S_{j-1}}{\sqrt{n}}\right)\phi_n\right] \qquad (6.27) \\
&= \frac{1}{2}\sigma^2 \lim_{n\to\infty}\left(E\left[\left(\int_{\bar{s}}^{t} f''\left(Y_u^{(n)}\right) du\right)\phi_n\right] + E[R(f)\phi_n]\right) \\
&= \frac{1}{2}\sigma^2 \lim_{n\to\infty} E\left[\left(\int_{\bar{s}}^{t} f''\left(Y_u^{(n)}\right) du\right)\phi_n\right] \\
&= \frac{1}{2}\sigma^2 E\left[\left(\int_{\bar{s}}^{t} f''\left(Y_u\right) du\right)\phi_Y\right], \qquad (6.28)
\end{aligned}$$

where (6.27) follows from (6.20) and (6.28) is a consequence of weak convergence of $Y^{(n)}$ to Y and the continuous mapping theorem. For this, note that the map $h \in C([0,1],\mathbb{R}) \mapsto \left(\int_{\bar{s}}^{t} f''(h(u))\,du\right)\phi(h)$ is continuous and bounded.
This completes the proof. $\square$

7 Alternative proof for a special case of the BFH invariance principle

In this chapter, we will consider random variables taking values in $G^2\left(\mathbb{R}^2\right)$. As we will often work with single real-valued components of such random variables in the sequel, we define the following projections:

Definition 7.1. For $i \in \{1, 2\}$,

$$\begin{aligned} \pi_{1;i} : T^2\left(\mathbb{R}^2\right) &\to \mathbb{R} \\ \left(a_0, \begin{pmatrix} a^{1;1} \\ a^{1;2} \end{pmatrix}, \begin{pmatrix} a^{2;11} & a^{2;12} \\ a^{2;21} & a^{2;22} \end{pmatrix}\right) &\mapsto a^{1;i}, \end{aligned}$$

and for $i, j \in \{1, 2\}$,

$$\begin{aligned} \pi_{2;ij} : T^2\left(\mathbb{R}^2\right) &\to \mathbb{R} \\ \left(a_0, \begin{pmatrix} a^{1;1} \\ a^{1;2} \end{pmatrix}, \begin{pmatrix} a^{2;11} & a^{2;12} \\ a^{2;21} & a^{2;22} \end{pmatrix}\right) &\mapsto a^{2;ij}. \end{aligned}$$

7.1 Result

We now use the martingale problem formulation to prove a special case of the BFH theorem (Theorem 4.2) without using the central limit theorem for centered iid variables on a nilpotent Lie group.

Theorem 7.2. *Let $(\mathbf{X}_i)_{i\geq 1}$ be a sequence of iid $G^2\left(\mathbb{R}^2\right)$-valued random variables which are centered. We assume that each real component of $\pi_1(\mathbf{X}_i)$ and $\pi_2(\mathbf{X}_i)$ is bounded i.e. there exists a constant $M \geq 1$ such that*

$$|\pi_{1;j}(\mathbf{X}_i)| \leq M \quad \text{for } j = 1, 2 \quad \text{and} \quad |\pi_{2;12}(\mathbf{X}_i)| \leq M. \tag{7.1}$$

In addition, we assume that $\pi_{2;12}(\mathbf{X}_i)$ is centered and that $\pi_{1;1}(\mathbf{X}_i)$, $\pi_{1;2}(\mathbf{X}_i)$ and $\pi_{2;12}(\mathbf{X}_i)$ are independent and have variance 1. Consider the rescaled random walk $\mathbf{W}^{(n)} = \{\mathbf{W}_t^{(n)}\}_{t\in[0,1]}$ defined by $\mathbf{W}_0^{(n)} = \mathbf{1}$,

$$\mathbf{W}_t^{(n)} = \delta_{n^{-1/2}}(\mathbf{X}_1 \otimes \cdots \otimes \mathbf{X}_{nt})$$

for $t \in \left\{\frac{1}{n}, \frac{2}{n}, \ldots, 1\right\}$ and interpolated by curves of the form $\left[\frac{i}{n}, \frac{i+1}{n}\right] \ni t \mapsto g \otimes \exp(ta)$,

where $g \in G^2\left(\mathbb{R}^2\right)$ and $a \in g^2\left(\mathbb{R}^2\right)$, in between [1].
Then, for any $\alpha \in (0,1/2)$, $\mathbf{W}^{(n)}$ converges weakly to an enhanced Brownian motion in $C^{0,\alpha-H\ddot{o}l}\left([0,1], G^2\left(\mathbb{R}^2\right)\right)$.

7.2 Proof

The strategy is the same as for the real-valued Markov chain in the previous chapter. Tightness was already established in the proof of the BFH theorem. The only thing left to do is showing that the limit process is an enhanced Brownian motion.
We define $\xi_i := \log(\mathbf{X}_i)$ for $i \in \mathbb{N}$.
As the map $\exp : g^2\left(\mathbb{R}^2\right) \to G^2\left(\mathbb{R}^2\right)$ is continuous, it is enough to prove that the limit process (in the sense of weak convergence) of $(\log(\mathbf{W}^{(n)}))_{n\geq 1}$ is

$$\left(0, \begin{pmatrix} B^1 \\ B^2 \end{pmatrix}, A\right),$$

where (B^1, B^2) is a 2-dimensional Brownian motion and A its Lévy area. An element of $g^2\left(\mathbb{R}^2\right)$ is determined by three real numbers (a 2-dimensional vector and one entry in the skew-symmetric 2×2 matrix), so that we can identify $g^2\left(\mathbb{R}^2\right)$ with $\mathbb{R}^3$ (as sets). We will thus work with the following 3-dimensional stochastic process:

$$\begin{aligned} \widetilde{S}^{(n)} &= \left(Y_1^{(n)}, Y_2^{(n)}, Z^{(n)}\right) \\ &:= \left(\pi_{1;1}\left(\log\left(\mathbf{W}^{(n)}\right)\right), \pi_{1;2}\left(\log\left(\mathbf{W}^{(n)}\right)\right), \pi_{2;12}\left(\log\left(\mathbf{W}^{(n)}\right)\right)\right). \end{aligned} \tag{7.2}$$

We clearly have $\widetilde{S}^{(n)}(0) = (0,0,0)$. We will now express the components of $\widetilde{S}^{(n)}$ in terms of the real-valued entries of the $g^2\left(\mathbb{R}^2\right)$-valued random variables ξ_i. If ξ and η are two $g^2\left(\mathbb{R}^2\right)$-valued random variables, then by the Campbell-Baker-Hausdorff formula, we have

$$\exp(\xi) \otimes \exp(\eta) = \exp\left(\xi + \eta + \frac{1}{2}[\xi, \eta]\right).$$

We iteratively apply this formula and use the fact that iterated bracket expressions

[1] that is

$$\begin{aligned} \mathbf{W}_t^{(n)}|_{\left[\frac{i}{n}, \frac{i+1}{n}\right]} =& \mathbf{W}_{i/n}^{(n)} \otimes \exp\left(-i \log\left(\left(\mathbf{W}_{i/n}^{(n)}\right)^{-1} \otimes \mathbf{W}_{(i+1)/n}^{(n)}\right)\right) \\ & \otimes \exp\left(tn \log\left(\left(\mathbf{W}_{i/n}^{(n)}\right)^{-1} \otimes \mathbf{W}_{(i+1)/n}^{(n)}\right)\right) \end{aligned}$$

for $i = 0, 1, \ldots, n-1$.

equal 0 in $g^2\left(\mathbb{R}^2\right)$ to obtain

$$Y_m^{(n)}\left(\frac{j}{n}\right) = \pi_{1;m}\left(\log\left(\mathbf{W}_{j/n}^{(n)}\right)\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^{j}\xi_i^{1;m} \quad \text{for } m \in \{1,2\} \text{ and} \tag{7.3}$$

$$Z^{(n)}\left(\frac{j}{n}\right) = \pi_{2;12}\left(\log\left(\mathbf{W}_{j/n}^{(n)}\right)\right) = \frac{1}{n}\left(\sum_{i=1}^{j}\xi_i^{2;12} + \frac{1}{2}\sum_{1\le k<i\le j}\left(\xi_k^{1;1}\xi_i^{1;2} - \xi_k^{1;2}\xi_i^{1;1}\right)\right) \tag{7.4}$$

for $j \in \{1,\ldots,n\}$. In between, the process is linearly interpolated (due to the form of the chosen interpolating curves in $G^2\left(\mathbb{R}^2\right)$).
We know (by tightness and Prohorov's theorem) that $\left(\mathbf{W}^{(n)}\right)_{n\ge1}$ is relatively compact. It follows from the continuous mapping theorem that $(\widetilde{S}^{(n)})_{n\ge1}$ is also relatively compact. To characterize the limit process of an arbitrary weakly convergent subsequence of $(\widetilde{S}^{(n)})_{n\ge1}$ by the martingale problem formulation, we will need several auxiliary results, which we establish next.

7.2.1 Step 1: Auxiliary results

We will need moment bounds for partial sums of the real-valued random variables $\xi_i^{1;m}$, $i \in \mathbb{N}$, $m \in \{1,2\}$. The bound for the first moments will follow from the one for the second moments, which we give first.

Lemma 7.3. *For $n \in \mathbb{N}$, $k \in \{1,\ldots,n\}$ and l, $m \in \{1,2\}$, we have*

$$E\left[\frac{1}{n}\left(\sum_{i=1}^{k}\xi_i^{1;m}\right)\left(\sum_{i=1}^{k}\xi_i^{1;l}\right)\right] \le \frac{k}{n} \le 1.$$

Proof. Since ξ_i and ξ_j are independent for $i \neq j$, the components $\xi_i^{1;1}$ and $\xi_i^{1;2}$ are independent and $\xi_i^{1;m}$, $m \in \{1,2\}$, has variance 1, we have

$$E\left[\frac{1}{n}\left(\sum_{i=1}^{k}\xi_i^{1;m}\right)\left(\sum_{i=1}^{k}\xi_i^{1;l}\right)\right] = \frac{1}{n}\sum_{i=1}^{k}E[\xi_i^{1;m}\xi_i^{1;l}] \le \frac{k}{n}.$$

□

Lemma 7.4. *For $n \in \mathbb{N}$, $k \in \{1,\ldots,n\}$ and $m \in \{1,2\}$, we have*

$$E\left[\frac{1}{\sqrt{n}}\left|\sum_{i=1}^{k}\xi_i^{1;m}\right|\right] \le \sqrt{\frac{k}{n}} \le 1.$$

Proof. Applying first Schwarz's inequality and then Lemma 7.3 yields

$$E\left[\frac{1}{n}\left|\sum_{i=1}^{k}\xi_i^{1;m}\right|\right] \leq E\left[\frac{1}{n}\left(\sum_{i=1}^{k}\xi_i^{1;m}\right)^2\right]^{1/2} \leq \sqrt{\frac{k}{n}}.$$

□

The proof of the bound for the third-order moments relies on Yokoyama's moment bound (Theorem 5.3).

Lemma 7.5. *There exists a constant $\overline{C} > 0$ such that for all $n \in \mathbb{N}$, for $k \in \{1, \ldots, n\}$ and for all $m_1, m_2 \in \mathbb{N}_0$ satisfying $m_1 + m_2 = 3$, we have*

$$E\left[\frac{1}{n^{3/2}}\left|\sum_{i=1}^{k}\xi_i^{1;1}\right|^{m_1}\left|\sum_{i=1}^{k}\xi_i^{1;2}\right|^{m_2}\right] \leq \overline{C}\left(\frac{k}{n}\right)^{3/2} \leq \overline{C}. \tag{7.5}$$

Proof. **Case 1**: $m_1 = 0$ or $m_2 = 0$
Let $j \in \{1, 2\}$. We consider the stationary sequence $(\xi_i^{1;j})_{i\geq 1}$ of independent, centered, bounded random variables. This sequence of random variables is strong mixing and satisfies assumption (5.1) in Yokoyama's moment bound (since we have $\alpha_i = 0$ for all $i \in \mathbb{N}$). Applying this bound yields that there exists a constant $C_j > 0$ such that for all $n \in \mathbb{N}$ and $k \in \{1, \ldots, n\}$

$$E\left[\left|\sum_{i=1}^{k}\xi_i^{1;j}\right|^3\right] \leq C_j k^{3/2}.$$

Case 2: $\min\{m_1, m_2\} \geq 1$
W.l.o.g., we can assume $m_1 = 1$ and $m_2 = 2$. We will deduce the desired bound from the bound obtained in Case 1.
Let $n \in \mathbb{N}$ and $k \in \{1, \ldots, n\}$. Then, we have

$$\left|\sum_{i=1}^{k}\xi_i^{1;1}\right|\left|\sum_{i=1}^{k}\xi_i^{1;2}\right|^2 \leq \max\left\{\left|\sum_{i=1}^{k}\xi_i^{1;1}\right|^3, \left|\sum_{i=1}^{k}\xi_i^{1;2}\right|^3\right\} \leq \left|\sum_{i=1}^{k}\xi_i^{1;1}\right|^3 + \left|\sum_{i=1}^{k}\xi_i^{1;2}\right|^3,$$

which implies

$$E\left[\left|\sum_{i=1}^{k}\xi_i^{1;1}\right|\left|\sum_{i=1}^{k}\xi_i^{1;2}\right|^2\right] \leq E\left[\left|\sum_{i=1}^{k}\xi_i^{1;1}\right|^3\right] + E\left[\left|\sum_{i=1}^{k}\xi_i^{1;2}\right|^3\right] \leq (C_1 + C_2)k^{3/2}.$$

Thus, (7.5) holds for $\overline{C} = C_1 + C_2$. □

As in the previous chapter, we will need the fact that the expectation of the difference between certain Riemann sums involving $\widetilde{S}^{(n)}$ and the corresponding integrals tends to 0 as $n \to \infty$.

Lemma 7.6. *Let $g : \mathbb{R}^3 \to \mathbb{R}$ be a continuously partially differentiable function with compact support. Then,*

$$E\left[\left|\frac{1}{n}\sum_{j=\lceil ns\rceil+1}^{\lceil nt\rceil} g\left(\widetilde{S}^{(n)}\left(\frac{j-1}{n}\right)\right) - \int_s^t g\left(\widetilde{S}^{(n)}(u)\right)du\right|\right] = O\left(\frac{1}{\sqrt{n}}\right).$$

Proof. It follows from the mean value theorem and the fact that first partial derivatives of g are bounded that there exists a constant $L > 0$ such that for all $x, y \in \mathbb{R}^3$

$$|g(x) - g(y)| \leq L\,\|y - x\|_1,$$

where $\|.\|_1$ denotes the L^1-norm on $\mathbb{R}^3$.
We fix $n \in \mathbb{N}$ and define the stochastic process $V^{(n)}$ on the interval $(\frac{\lceil ns\rceil}{n}, \frac{\lceil nt\rceil}{n}]$ by

$$V^{(n)}(u) = \widetilde{S}^{(n)}\left(\frac{\lceil ns\rceil + m}{n}\right) \quad \text{for } u \in \left(\frac{\lceil ns\rceil + m}{n}, \frac{\lceil ns\rceil + m + 1}{n}\right], \quad m = 0, \ldots, \lceil nt\rceil - \lceil ns\rceil - 1.$$

Now, let $k \in \{\lceil ns\rceil, \lceil ns\rceil + 1, \ldots, \lceil nt\rceil - 1\}$. $\widetilde{S}^{(n)}|_{(k/n,(k+1)/n]}$ is the linear interpolation between $\widetilde{S}^{(n)}(\frac{k}{n})$ and $\widetilde{S}^{(n)}(\frac{k+1}{n})$ and we have

$$\begin{aligned}
&\widetilde{S}^{(n)}\left(\frac{k+1}{n}\right) - \widetilde{S}^{(n)}\left(\frac{k}{n}\right)\\
&= \left(\frac{1}{\sqrt{n}}\xi_{k+1}^{1;1}, \frac{1}{\sqrt{n}}\xi_{k+1}^{1;2}, \frac{1}{n}\left(\xi_{k+1}^{2;12} + \frac{1}{2}\xi_{k+1}^{1;2}\left(\sum_{i=1}^{k}\xi_i^{1;1}\right) - \frac{1}{2}\xi_{k+1}^{1;1}\left(\sum_{i=1}^{k}\xi_i^{1;2}\right)\right)\right).
\end{aligned}$$

Note that it follows from assumption (7.1) that $\left|\xi_i^{1;m}\right| = \left|\mathbf{X}_i^{1;m}\right|$ and $\left|\xi_i^{2;12}\right| = \left|\mathbf{X}_i^{2;12} - \frac{1}{2}\mathbf{X}_i^{1;1}\mathbf{X}_i^{1;2}\right|$ are bounded by $\overline{M} := M + M^2$. Hence,

$$\begin{aligned}
&E\left[\left\|V^{(n)}(u) - \widetilde{S}^{(n)}(u)\right\|_1\right]\\
&\leq E\left[\left\|\left(\frac{1}{\sqrt{n}}\xi_{k+1}^{1;1}, \frac{1}{\sqrt{n}}\xi_{k+1}^{1;2}, \frac{1}{n}\left(\xi_{k+1}^{2;12} + \frac{1}{2}\xi_{k+1}^{1;2}\left(\sum_{i=1}^{k}\xi_i^{1;1}\right)\right.\right.\right.\right.\\
&\qquad \left.\left.\left.\left. -\frac{1}{2}\xi_{k+1}^{1;1}\left(\sum_{i=1}^{k}\xi_i^{1;2}\right)\right)\right)\right\|_1\right]\\
&\leq \frac{2\overline{M}}{\sqrt{n}} + \frac{\overline{M}}{n} + \frac{\overline{M}}{2}E\left[\frac{1}{n}\left|\sum_{i=1}^{k}\xi_i^{1;1}\right|\right] + \frac{\overline{M}}{2}E\left[\frac{1}{n}\left|\sum_{i=1}^{k}\xi_i^{1;2}\right|\right] \qquad (7.6)\\
&\leq \frac{4\overline{M}}{\sqrt{n}}, \qquad (7.7)
\end{aligned}$$

where (7.7) is due to Lemma 7.4.
We have

$$\begin{aligned}
&E\left[\left|\int_{\frac{k}{n}}^{\frac{k+1}{n}}\left(g\left(V^{(n)}(u)\right)-g\left(\widetilde{S}^{(n)}(u)\right)\right)du\right|\right] \\
\leq\ &\int_{\frac{k}{n}}^{\frac{k+1}{n}} E\left[\left|g\left(V^{(n)}(u)\right)-g\left(\widetilde{S}^{(n)}(u)\right)\right|\right]du \\
\leq\ &\frac{4L\overline{M}}{n\sqrt{n}},
\end{aligned}$$

where the last inequality is due to Lipschitz continuity of g and (7.7). Finally, by a similar argument as in the proof of Lemma 6.12, we obtain

$$\begin{aligned}
&E\left[\left|\frac{1}{n}\sum_{j=\lceil ns\rceil+1}^{\lceil nt\rceil} g\left(\widetilde{S}^{(n)}\left(\frac{j-1}{n}\right)\right)-\int_s^t g\left(\widetilde{S}^{(n)}(u)\right)du\right|\right] \\
=\ &E\left[\left|\int_{\lceil ns\rceil/n}^{\lceil nt\rceil/n} g\left(V^{(n)}(u)\right)du-\int_s^t g\left(\widetilde{S}^{(n)}(u)\right)du\right|\right] \\
=\ &E\left[\left|\int_{\lceil ns\rceil/n}^{(\lceil nt\rceil-1)/n}\left(g\left(V^{(n)}(u)\right)-g\left(\widetilde{S}^{(n)}(u)\right)\right)du\right|\right]+\frac{3}{n}\|g\|_\infty \\
\leq\ &\sum_{k=\lceil ns\rceil}^{\lceil nt\rceil-2} E\left[\left|\int_{\frac{k}{n}}^{\frac{k+1}{n}}\left(g\left(V^{(n)}(u)\right)-g\left(\widetilde{S}^{(n)}(u)\right)\right)du\right|\right]+\frac{3}{n}\|g\|_\infty \\
\leq\ &(\lceil nt\rceil-\lceil ns\rceil)\frac{4L\overline{M}}{n\sqrt{n}}+\frac{3}{n}\|g\|_\infty\leq\frac{4L\overline{M}+3\|g\|_\infty}{\sqrt{n}}.
\end{aligned}$$

□

7.2.2 Step 2: Characterization of the limit process

Let $\widetilde{S}=(Y_1,Y_2,Z)$ be the limit process of an arbitrary weakly convergent subsequence of $(\widetilde{S}^{(n)})_{n\geq 1}$. As in the last chapter, we will omit the indices for the subsequence.
Let $a:\mathbb{R}^3\to S_3$ be defined by

$$a(x,y,z)=\begin{pmatrix}1 & 0 & -y/2\\ 0 & 1 & x/2\\ -y/2 & x/2 & (x^2+y^2)/4\end{pmatrix}$$

and let $b:\mathbb{R}^3\to\mathbb{R}^3; x\mapsto(0,0,0)$.
The aim is to establish the following
Claim: $P\widetilde{S}^{-1}$ is the unique solution to the martingale problem for (a,b) starting at $(0,0,0)$.

Using equivalence of the martingale problem formulation and the weak solution formulation for SDEs, we can easily deduce the desired characterization of the limit process $\widetilde{S}$ from the above claim:
Note that the map $\sigma : \mathbb{R}^3 \to M_3$ given by

$$\sigma(x,y,z) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -y/2 & x/2 & 0 \end{pmatrix}$$

is Lipschitz continuous and we have $\sigma(x,y,z)\sigma(x,y,z)^T = a(x,y,z)$. Consequently, uniqueness in law holds for the 3-dimensional SDE

$$X_t = X_0 + \int_0^t \sigma(X_u)\; dB_u \tag{7.8}$$

and due to Theorem 5.10, every weak solution of (7.8) starting at $(0,0,0)$ has law $P\widetilde{S}^{-1}$. It is easy to see that[2] (B^1, B^2, A^{12}) is such a weak solution.

Proof of the claim:

Uniqueness is due to the fact that the map σ is Lipschitz continuous.
$P(\widetilde{S}(0) = (0,0,0)) = 1$ follows from $P(\widetilde{S}^{(n)}(0) = (0,0,0)) = 1$ for all $n \in \mathbb{N}$.
The second-order differential operator corresponding to the above martingale problem is

$$\begin{aligned} L_{a,b}F(x,y,z) = \frac{1}{2}\;\partial_{11}F(x,y,z) + \frac{1}{2}\;\partial_{22}F(x,y,z) + \frac{1}{8}(x^2+y^2)\;\partial_{33}F(x,y,z) \\ - \frac{1}{2}y\;\partial_{13}F(x,y,z) + \frac{1}{2}x\;\partial_{23}F(x,y,z). \end{aligned}$$

The first part of the argumentation is analogous to the one for the real-valued Markov chain in the previous chapter (see p. 60ff.). We define the σ-algebras $\mathcal{F}_0 := \{\emptyset, \Omega\}$ and

$$\mathcal{F}_j := \sigma\left(\mathbf{X}_i : 1 \le i \le j\right) \qquad \text{for } j \in \{1,\ldots,n\}.$$

Observe that

$$\mathcal{F}_j := \sigma\left(\mathbf{W}_u^{(n)} : 0 \le u \le \frac{j}{n}\right).$$

Further note that since $\mathbf{W}^{(n)}$ is continuous, the process $\widetilde{S}^{(n)}$ and the limit process $\widetilde{S}$ are also continuous. Let $F \in C_0^\infty\left(\mathbb{R}^3\right)$, $s,t \in [0,1]$, $s < t$ and let $\phi : C([0,1],\mathbb{R}^3) \to \mathbb{R}$ be a continuous, bounded, $\mathcal{M}_s$-measurable map. We define $\phi_n := \phi\left(\widetilde{S}^{(n)}\right)$ and $\phi_S := \phi\left(\widetilde{S}\right)$. The aim is to show that

$$E\left[\left(F\left(\widetilde{S}(t)\right) - F\left(\widetilde{S}(s)\right) - \int_s^t L_{a,b}F\left(\widetilde{S}(u)\right)du\right)\phi_S\right] = 0. \tag{7.9}$$

[2]Recall that A denotes the Lévy area of the 2-dimensional Brownian motion (B^1, B^2).

Note that for fixed $n \in \mathbb{N}$

$$F\left(\widetilde{S}^{(n)}\left(\frac{k}{n}\right)\right) - \sum_{j=1}^{k}\left(E\left[F\left(\widetilde{S}^{(n)}\left(\frac{j}{n}\right)\right) \,\middle|\, \mathcal{F}_{j-1}\right] - F\left(\widetilde{S}^{(n)}\left(\frac{j-1}{n}\right)\right)\right),$$

$1 \leq k \leq n$, is a (discrete-parameter) $P\left(\widetilde{S}^{(n)}\right)^{-1}$-martingale with respect to the filtration $(\mathcal{M}_{k/n})_{1\leq k\leq n}$. We now aim at proving that

$$\begin{aligned}
&\lim_{n\to\infty} E\Bigg[\Bigg(F\left(\widetilde{S}^{(n)}\left(\frac{\lceil nt\rceil}{n}\right)\right) - F\left(\widetilde{S}^{(n)}\left(\frac{\lceil ns\rceil}{n}\right)\right) \\
&\qquad - \sum_{j=\lceil ns\rceil+1}^{\lceil nt\rceil}\left(E\left[F\left(\widetilde{S}^{(n)}\left(\frac{j}{n}\right)\right) \,\middle|\, \mathcal{F}_{j-1}\right] - F\left(\widetilde{S}^{(n)}\left(\frac{j-1}{n}\right)\right)\right)\Bigg)\phi_n\Bigg] \\
&= E\left[\left(F\left(\widetilde{S}(t)\right) - F\left(\widetilde{S}(s)\right) - \int_s^t L_{a,b}F\left(\widetilde{S}(u)\right)du\right)\phi_S\right], \qquad (7.10)
\end{aligned}$$

since by Lemma 5.5, the above limit equals 0.
It follows from weak convergence of $\widetilde{S}^{(n)}$ to $\widetilde{S}$ and the continuous mapping theorem that

$$\begin{aligned}
&\lim_{n\to\infty} E\left[\left(F\left(\widetilde{S}^{(n)}\left(\frac{\lceil nt\rceil}{n}\right)\right) - F\left(\widetilde{S}^{(n)}\left(\frac{\lceil ns\rceil}{n}\right)\right)\right)\phi_n\right] \\
&= E\left[\left(F\left(\widetilde{S}(t)\right) - F\left(\widetilde{S}(s)\right)\right)\phi_S\right]. \qquad (7.11)
\end{aligned}$$

Now, we do Taylor expansion for the argument of the conditional expectation in (7.10). Let $j \in \{1,\dots,n\}$. Then,

$$\begin{aligned}
&F\left(\widetilde{S}^{(n)}\left(\frac{j}{n}\right)\right) \\
&= F\left(\widetilde{S}^{(n)}\left(\frac{j-1}{n}\right) + \right. \\
&\left.\left(\frac{1}{\sqrt{n}}\xi_j^{1;1}, \frac{1}{\sqrt{n}}\xi_j^{1;2}, \frac{1}{\sqrt{n}}\underbrace{\left(\frac{1}{\sqrt{n}}\xi_j^{2;12} + \frac{1}{2}\xi_j^{1;2}\overbrace{\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{j-1}\xi_i^{1;1}\right)}^{=Y_1^{(n)}\left(\frac{j-1}{n}\right)} - \frac{1}{2}\xi_j^{1;1}\overbrace{\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{j-1}\xi_i^{1;2}\right)}^{=Y_2^{(n)}\left(\frac{j-1}{n}\right)}\right)}_{=:U(j,n)}\right)\right) \\
&= F\left(\widetilde{S}^{(n)}\left(\frac{j-1}{n}\right)\right) \\
&+ \frac{1}{\sqrt{n}}\xi_j^{1;1}\,\partial_1 F\left(\widetilde{S}^{(n)}\left(\frac{j-1}{n}\right)\right) + \frac{1}{\sqrt{n}}\xi_j^{1;2}\,\partial_2 F\left(\widetilde{S}^{(n)}\left(\frac{j-1}{n}\right)\right)
\end{aligned}$$

$$
\begin{aligned}
&+\ \frac{1}{\sqrt{n}}\, U(j,n)\,\partial_3 F\left(\widetilde{S}^{(n)}\left(\frac{j-1}{n}\right)\right)\\
&+\ \frac{1}{2n}\left(\xi_j^{1;1}\right)^2 \partial_{11}F\left(\widetilde{S}^{(n)}\left(\frac{j-1}{n}\right)\right)+\frac{1}{2n}\left(\xi_j^{1;2}\right)^2 \partial_{22}F\left(\widetilde{S}^{(n)}\left(\frac{j-1}{n}\right)\right)\\
&+\ \frac{1}{2n}\,(U(j,n))^2\,\partial_{33}F\left(\widetilde{S}^{(n)}\left(\frac{j-1}{n}\right)\right)\\
&+\ \frac{1}{n}\xi_j^{1;1}\xi_j^{1;2}\,\partial_{12}F\left(\widetilde{S}^{(n)}\left(\frac{j-1}{n}\right)\right)\\
&+\ \frac{1}{n}\xi_j^{1;1}\,U(j,n)\,\partial_{13}F\left(\widetilde{S}^{(n)}\left(\frac{j-1}{n}\right)\right)+\frac{1}{n}\xi_j^{1;2}\,U(j,n)\,\partial_{23}F\left(\widetilde{S}^{(n)}\left(\frac{j-1}{n}\right)\right)\\
&+\ R_3\left(\widetilde{S}^{(n)}\left(\frac{j}{n}\right)\right),
\end{aligned}
$$

where R_3 denotes the remainder term.
We use that $\widetilde{S}^{(n)}\left(\frac{j-1}{n}\right)$ and $Y_m^{(n)}\left(\frac{j-1}{n}\right)$, $m=1,2$ are $\mathcal{F}_{j-1}$-measurable to obtain[3]

$$
\begin{aligned}
&E\left[F\left(\widetilde{S}^{(n)}\left(\frac{j}{n}\right)\right)\,\middle|\,\mathcal{F}_{j-1}\right]-F\left(\widetilde{S}^{(n)}\left(\frac{j-1}{n}\right)\right)\\
=&\ \frac{1}{\sqrt{n}}\,\partial_1 F\left(\widetilde{S}^{(n)}\left(\frac{j-1}{n}\right)\right)E\left[\xi_j^{1;1}\,\middle|\,\mathcal{F}_{j-1}\right]\\
+&\ \frac{1}{\sqrt{n}}\,\partial_2 F\left(\widetilde{S}^{(n)}\left(\frac{j-1}{n}\right)\right)E\left[\xi_j^{1;2}\,\middle|\,\mathcal{F}_{j-1}\right]\\
+&\ \frac{1}{\sqrt{n}}\,\partial_3 F\left(\widetilde{S}^{(n)}\left(\frac{j-1}{n}\right)\right)\left(\frac{1}{\sqrt{n}}E\left[\xi_j^{2;12}\,\middle|\,\mathcal{F}_{j-1}\right]+\ \frac{1}{2}\,Y_1^{(n)}\left(\frac{j-1}{n}\right)E\left[\xi_j^{1;2}\,\middle|\,\mathcal{F}_{j-1}\right]\right.\\
&\qquad\left.-\ \frac{1}{2}\,Y_2^{(n)}\left(\frac{j-1}{n}\right)E\left[\xi_j^{1;1}\,\middle|\,\mathcal{F}_{j-1}\right]\right)\\
+&\ \frac{1}{2n}\,\partial_{11}F\left(\widetilde{S}^{(n)}\left(\frac{j-1}{n}\right)\right)E\left[\left(\xi_j^{1;1}\right)^2\,\middle|\,\mathcal{F}_{j-1}\right]\\
+&\ \frac{1}{2n}\,\partial_{22}F\left(\widetilde{S}^{(n)}\left(\frac{j-1}{n}\right)\right)E\left[\left(\xi_j^{1;2}\right)^2\,\middle|\,\mathcal{F}_{j-1}\right]\\
+&\ \frac{1}{2n}\,\partial_{33}F\left(\widetilde{S}^{(n)}\left(\frac{j-1}{n}\right)\right)\left(\frac{1}{n}E\left[\left(\xi_j^{2;12}\right)^2\,\middle|\,\mathcal{F}_{j-1}\right]\right.\\
&\qquad+\ \frac{1}{4}\left(Y_1^{(n)}\left(\frac{j-1}{n}\right)\right)^2E\left[\left(\xi_j^{1;2}\right)^2\,\middle|\,\mathcal{F}_{j-1}\right]\\
&\qquad+\ \frac{1}{4}\left(Y_2^{(n)}\left(\frac{j-1}{n}\right)\right)^2E\left[\left(\xi_j^{1;1}\right)^2\,\middle|\,\mathcal{F}_{j-1}\right]\\
&\qquad+\frac{1}{\sqrt{n}}\,Y_1^{(n)}\left(\frac{j-1}{n}\right)E\left[\xi_j^{2;12}\xi_j^{1;2}\,\middle|\,\mathcal{F}_{j-1}\right]-\ \frac{1}{\sqrt{n}}\,Y_2^{(n)}\left(\frac{j-1}{n}\right)E\left[\xi_j^{2;12}\xi_j^{1;1}\,\middle|\,\mathcal{F}_{j-1}\right]
\end{aligned}
$$

[3] Writing down the whole expansion in this general form will be useful for the next chapter. We will see below that in the setting of this chapter, most of the summands actually equal 0.

$$
\begin{aligned}
&\quad - \frac{1}{2}\, Y_1^{(n)}\left(\frac{j-1}{n}\right) Y_2^{(n)}\left(\frac{j-1}{n}\right) E\left[\xi_j^{1;1}\xi_j^{1;2} \;\middle|\; \mathcal{F}_{j-1}\right]\Bigg) \\
&+ \frac{1}{n}\, \partial_{12}F\left(\widetilde{S}^{(n)}\left(\frac{j-1}{n}\right)\right) E\left[\xi_j^{1;1}\xi_j^{1;2} \;\middle|\; \mathcal{F}_{j-1}\right] \\
&+ \frac{1}{n}\, \partial_{13}F\left(\widetilde{S}^{(n)}\left(\frac{j-1}{n}\right)\right) \Bigg(\frac{1}{\sqrt{n}} E\left[\xi_j^{1;1}\xi_j^{2;12} \;\middle|\; \mathcal{F}_{j-1}\right] \\
&\quad + \frac{1}{2}\, Y_1^{(n)}\left(\frac{j-1}{n}\right) E\left[\xi_j^{1;1}\xi_j^{1;2} \;\middle|\; \mathcal{F}_{j-1}\right] - \frac{1}{2}\, Y_2^{(n)}\left(\frac{j-1}{n}\right) E\left[\left(\xi_j^{1;1}\right)^2 \;\middle|\; \mathcal{F}_{j-1}\right]\Bigg) \\
&+ \frac{1}{n}\, \partial_{23}F\left(\widetilde{S}^{(n)}\left(\frac{j-1}{n}\right)\right) \Bigg(\frac{1}{\sqrt{n}} E\left[\xi_j^{1;2}\xi_j^{2;12} \;\middle|\; \mathcal{F}_{j-1}\right] \\
&\quad + \frac{1}{2}\, Y_1^{(n)}\left(\frac{j-1}{n}\right) E\left[\left(\xi_j^{1;2}\right)^2 \;\middle|\; \mathcal{F}_{j-1}\right] - \frac{1}{2}\, Y_2^{(n)}\left(\frac{j-1}{n}\right) E\left[\xi_j^{1;1}\xi_j^{1;2} \;\middle|\; \mathcal{F}_{j-1}\right]\Bigg) \\
&+ E\left[R_3\left(\widetilde{S}^{(n)}\left(\frac{j}{n}\right)\right) \middle|\; \mathcal{F}_{j-1}\right] \quad \text{a.s..}
\end{aligned} \tag{7.12}
$$

Note that

$$
\left| E\left[E\left[R_3\left(\widetilde{S}^{(n)}\left(\frac{j}{n}\right)\right) \middle|\; \mathcal{F}_{j-1}\right] \phi_n \right] \right| = O\left(n^{-3/2}\right) \tag{7.13}
$$

is a consequence of the moment bounds given in Step 1 (Lemma 7.3 - 7.5) and the fact that the third-order partial derivatives of F are bounded.
Now, $\xi_j = \log(\mathbf{X}_j)$ is independent of $\mathcal{F}_{j-1}$, so that for any measurable function $f : g^2\left(\mathbb{R}^2\right) \to \mathbb{R}$, we have $E[f(\xi_j)|\mathcal{F}_{j-1}] = E[f(\xi_j)]$ a.s.. In addition, we use that $\mathbf{X}_j^{1;1} = \xi_j^{1;1}$, $\mathbf{X}_j^{1;2} = \xi_j^{1;2}$ and $\mathbf{X}_j^{2;12}$ are independent, centered and have variance 1, and hence $E\left[\xi_j^{1;1}\xi_j^{1;2}\right] = E\left[\xi_j^{1;1}\xi_j^{2;12}\right] = E\left[\xi_j^{1;2}\xi_j^{2;12}\right] = 0$ and $E\left[\xi_j^{2;12}\right] = E\left[\mathbf{X}_j^{2;12} - \frac{1}{2}\mathbf{X}_j^{1;1}\mathbf{X}_j^{1;2}\right] = 0$. Thus, we obtain

$$
\begin{aligned}
&E\left[\sum_{j=\lceil ns\rceil+1}^{\lceil nt\rceil} \left(E\left[F\left(\widetilde{S}^{(n)}\left(\frac{j}{n}\right)\right) \;\middle|\; \mathcal{F}_{j-1}\right] - F\left(\widetilde{S}^{(n)}\left(\frac{j-1}{n}\right)\right)\right)\phi_n\right] \\
&= \frac{1}{2n}\sum_{j=\lceil ns\rceil+1}^{\lceil nt\rceil} E\left[\partial_{11}F\left(\widetilde{S}^{(n)}\left(\frac{j-1}{n}\right)\right)\phi_n\right] \\
&+ \frac{1}{2n}\sum_{j=\lceil ns\rceil+1}^{\lceil nt\rceil} E\left[\partial_{22}F\left(\widetilde{S}^{(n)}\left(\frac{j-1}{n}\right)\right)\phi_n\right] \\
&+ \frac{1}{8n}\sum_{j=\lceil ns\rceil+1}^{\lceil nt\rceil} E\Bigg[\partial_{33}F\left(\widetilde{S}^{(n)}\left(\frac{j-1}{n}\right)\right)\Bigg(\left(Y_1^{(n)}\left(\frac{j-1}{n}\right)\right)^2 \\
&\quad + \left(Y_2^{(n)}\left(\frac{j-1}{n}\right)\right)^2\Bigg)\phi_n\Bigg]
\end{aligned}
$$

$$
\begin{aligned}
&- \frac{1}{2n} \sum_{j=\lceil ns\rceil+1}^{\lceil nt\rceil} E\left[\partial_{13}F\left(\widetilde{S}^{(n)}\left(\frac{j-1}{n}\right)\right) Y_2^{(n)}\left(\frac{j-1}{n}\right)\phi_n\right] \\
&+ \frac{1}{2n} \sum_{j=\lceil ns\rceil+1}^{\lceil nt\rceil} E\left[\partial_{23}F\left(\widetilde{S}^{(n)}\left(\frac{j-1}{n}\right)\right) Y_1^{(n)}\left(\frac{j-1}{n}\right)\phi_n\right] \\
&+ o\left(\frac{1}{n}\right).
\end{aligned}
$$

Next, we will apply Lemma 7.6 to the above sums of expectations. For this, let $g : \mathbb{R}^3 \to \mathbb{R}$ be a continuously partially differentiable function with compact support. We set

$$
R(g) := \frac{1}{n} \sum_{j=\lceil ns\rceil+1}^{\lceil nt\rceil} g\left(\widetilde{S}^{(n)}\left(\frac{j-1}{n}\right)\right) - \int_s^t g\left(\widetilde{S}^{(n)}(u)\right) du.
$$

By Lemma 7.6, we then have

$$
|E[R(g)\phi_n]| \leq \|\phi\|_\infty E[|R(g)|] = O\left(n^{-1/2}\right).
$$

In addition, it follows from weak convergence of $\widetilde{S}^{(n)}$ to $\widetilde{S}$ and the continuous mapping theorem that

$$
\lim_{n\to\infty} E\left[\left(\int_s^t g\left(\widetilde{S}^{(n)}(u)\right) du\right)\phi_n\right] = E\left[\left(\int_s^t g\left(\widetilde{S}(u)\right) du\right)\phi_S\right].
$$

For example, if we apply the above results to $g = \partial_{13}F(.)p_2(.)$, where $p_2 : \mathbb{R}^3 \to \mathbb{R}$ denotes the projection to the second component, we get

$$
\begin{aligned}
&\lim_{n\to\infty} \sum_{j=\lceil ns\rceil+1}^{\lceil nt\rceil} -\frac{1}{2n} E\left[\partial_{13}F\left(\widetilde{S}^{(n)}\left(\frac{j-1}{n}\right)\right) Y_2^{(n)}\left(\frac{j-1}{n}\right)\phi_n\right] \\
&= -\frac{1}{2}\lim_{n\to\infty}\left(E\left[\left(\int_s^t \partial_{13}F\left(\widetilde{S}^{(n)}(u)\right) Y_2^{(n)}(u)\,du\right)\phi_n\right] + O\left(n^{-1/2}\right)\right) \\
&= -\frac{1}{2} E\left[\left(\int_s^t \partial_{13}F\left(\widetilde{S}(u)\right) Y_2(u)\,du\right)\phi_S\right].
\end{aligned}
$$

Using analogous arguments for the other summands, we thus obtain

$$
\begin{aligned}
&\lim_{n\to\infty} E\left[\sum_{j=\lceil ns\rceil+1}^{\lceil nt\rceil}\left(E\left[F\left(\widetilde{S}^{(n)}\left(\frac{j}{n}\right)\right) \middle| \mathcal{F}_{j-1}\right] - F\left(\widetilde{S}^{(n)}\left(\frac{j-1}{n}\right)\right)\right)\phi_n\right] \\
&= \frac{1}{2}E\left[\left(\int_s^t \partial_{11}F\left(\widetilde{S}(u)\right) du\right)\phi_S\right] + \frac{1}{2}E\left[\left(\int_s^t \partial_{22}F\left(\widetilde{S}(u)\right) du\right)\phi_S\right] \\
&+ \frac{1}{8}E\left[\left(\int_s^t \partial_{33}F\left(\widetilde{S}(u)\right)\left((Y_1(u))^2 + (Y_2(u))^2\right) du\right)\phi_S\right]
\end{aligned}
$$

$$
\begin{aligned}
&- \frac{1}{2} E\left[\left(\int_s^t \partial_{13} F\left(\widetilde{S}\,(u)\right) Y_2\,(u)\,du\right) \phi_S\right] \\
&+ \frac{1}{2} E\left[\left(\int_s^t \partial_{23} F\left(\widetilde{S}\,(u)\right) Y_1\,(u)\,du\right) \phi_S\right] \\
&= E\left[\left(\int_s^t L_{a,b} F\left(\widetilde{S}(u)\right) du\right) \phi_S\right].
\end{aligned}
$$

Combining this with (7.11) yields (7.10) and completes the proof. □

8 Invariance principle for Markov chains with values in $G^2(\mathbb{R}^2)$

8.1 Result

Setting and notation

Let $(\mathbf{X}_i)_{i\geq 1}$ be a time-homogenous Markov chain with values in $G^2(\mathbb{R}^2)$. We assume that each real component of $\pi_1(\mathbf{X}_i)$ and $\pi_2(\mathbf{X}_i)$ is bounded i.e. there exists a constant $M \geq 1$ such that

$$|\pi_{1;j}(\mathbf{X}_i)| \leq M \quad \text{for } j = 1,2 \quad \text{and} \quad |\pi_{2;12}(\mathbf{X}_i)| \leq M. \tag{8.1}$$

Let the transition kernel $P(x, dy)$ of the Markov chain be absolutely continuous with respect to a probability measure μ with density $p(x,y) := P(x,dy)/\mu(dy)$ such that

$$\frac{1}{C} \leq p(x,y) \leq C$$

for some constant $C \geq 1$ and all $x, y \in G^2(\mathbb{R}^2)$. This implies that there exists a unique stationary distribution[1] ν for the Markov chain $(\mathbf{X}_i)_{i\geq 1}$.
We further assume that the transition kernel $P(x,dy)$ is such that for $j = 1,2$

$$\int \pi_{1;j} \; d\nu = 0.$$

Let ν be the initial distribution of the Markov chain $(\mathbf{X}_i)_{i\geq 1}$, so that this Markov chain is stationary.

Remark 8.1. Combining the last two assumptions gives that the (real-valued) random variables $\pi_{1;j}(\mathbf{X}_i) = \pi_{1;j}(\log(\mathbf{X}_i))$, $j = 1,2$ are centered. This fact will be fundamental in the subsequent argumentation.

We define $\xi_i := \log(\mathbf{X}_i)$ for $i \in \mathbb{N}$ and use the following notation:

$$\begin{aligned}
&\sigma_m^2 := var\left(\xi_1^{1;m}\right) = E\left[\left(\xi_1^{1;m}\right)^2\right] \quad \text{for } m \in \{1,2\};\\
&\gamma := E\left[\xi_1^{1;1}\xi_1^{1;2}\right];\\
&s_{lm} := \sum_{k=2}^{\infty} E\left[\xi_1^{1;l}\xi_k^{1;m}\right] \quad \text{for } l,m \in \{1,2\};\\
&v := \sigma_1^2 + 2s_{11} > 0; \quad w := \sigma_2^2 + 2s_{22} > 0;\\
&c := \gamma + s_{12} + s_{21}; \quad d := E\left[\xi_1^{2;12}\right] + \frac{1}{2}\left(s_{12} - s_{21}\right).
\end{aligned} \tag{8.2}$$

[1] This will be shown in the next section.

We assume that

$$c^2 \leq vw.$$

Remark 8.2. The fact that the series in (8.2) converge will be a consequence of the exponential covariance bound, which will be established in the next section.

In addition, we will need the injection

$$i : C\left([0,1], \mathbb{R}^3\right) \to C\left([0,1], g^2\left(\mathbb{R}^2\right)\right)$$
$$f = \begin{pmatrix} f_1 \\ f_2 \\ f_3 \end{pmatrix} \mapsto \left(0, \begin{pmatrix} f_1 \\ f_2 \end{pmatrix}, \begin{pmatrix} 0 & f_3 \\ -f_3 & 0 \end{pmatrix}\right),$$

and the map[2]

$$\widetilde{exp} : C\left([0,1], g^2\left(\mathbb{R}^2\right)\right) \to C\left([0,1], G^2\left(\mathbb{R}^2\right)\right)$$
$$\left(0, \begin{pmatrix} f_1 \\ f_2 \end{pmatrix}, \begin{pmatrix} 0 & f_3 \\ -f_3 & 0 \end{pmatrix}\right) \mapsto \left(1, \begin{pmatrix} f_1 \\ f_2 \end{pmatrix}, \frac{1}{2}\begin{pmatrix} f_1^2 & 2f_3 + f_1 \circ f_2 \\ -2f_3 + f_2 \circ f_1 & f_2^2 \end{pmatrix}\right).$$

Let $a : \mathbb{R}^3 \to S_3$ be defined by

$$a(x,y,z) = \begin{pmatrix} v & c & (cx - vy)/2 \\ c & w & (wx - cy)/2 \\ (cx - vy)/2 & (wx - cy)/2 & (wx^2 + vy^2 - 2cxy)/4 \end{pmatrix} \tag{8.3}$$

and let

$$b : \mathbb{R}^3 \to \mathbb{R}^3; x \mapsto (0,0,d). \tag{8.4}$$

Under these assumptions, we have

Theorem 8.3. *Consider the stochastic process $\mathbf{W}^{(n)} = \{\mathbf{W}_t^{(n)}\}_{t\in[0,1]}$ as defined in Theorem 7.2 (p. 69f.).*
Then, for any $\alpha \in (0, 1/2)$, $\left(\mathbf{W}^{(n)}\right)_{n\geq 1}$ converges weakly to the probability measure $Q\left(\widetilde{\exp} \circ i\right)^{-1}$ in $C^{0,\alpha-Höl}\left([0,1], G^2\left(\mathbb{R}^2\right)\right)$, where Q is the unique solution to the martingale problem for (a,b) starting at $(0,0,0)$.

8.2 An exponential covariance bound and moment bounds

In the sequel, $(\mathbf{X}_i)_{i\geq 1}$ will always denote a Markov chain satisfying all conditions introduced in the previous section and $P(x, dy)$ will denote its transition kernel. Let $\mathcal{B}_{G^2(\mathbb{R}^2)}$ be the Borel σ-algebra on $G^2\left(\mathbb{R}^2\right)$.

The covariance bound given in this section is fundamental for the proof of Theorem 8.3. To establish that bound, we will need the fact that the n-step transition kernels of the Markov chain $(\mathbf{X}_i)_{i\geq 1}$ converge exponentially fast to the stationary measure ν in total variation distance[3].

[2]This is just the "functional version" of the exponential map $g^2\left(\mathbb{R}^2\right) \to G^2\left(\mathbb{R}^2\right)$.

[3]For two probability measures P and Q on $(\Omega, \mathcal{F})$, the total variation distance is defined by $\|P - Q\|_{\mathrm{tv}} := \sup_{A\in\mathcal{F}} |P(A) - Q(A)|$.

Proposition 8.4. *There exists a constant $\lambda > 0$ such that for all $n \in \mathbb{N}$*

$$\sup_{x \in G^2(\mathbb{R}^2)} \|P^n(x,.) - \nu\|_{tv} \leq \exp(-\lambda n).$$

The idea for the proof below is taken from [21][4].

Proof. First, we establish that it suffices to prove the following
Claim: There exist a constant $\lambda > 0$ such that for all $n \in \mathbb{N}$

$$\eta(n) := sup_{x,y \in G^2(\mathbb{R}^2)} \|P^n(x,.) - P^n(y,.)\|_{\mathrm{tv}} \leq \exp(-\lambda n). \tag{8.5}$$

Let $x \in G^2(\mathbb{R}^2)$ and $A \in \mathcal{B}_{G^2(\mathbb{R}^2)}$. Then, by the Chapman-Kolmogorov equation, we have for all $n, m \in \mathbb{N}$

$$\left|P^n(x,A) - P^{m+n}(x,A)\right| = \left|\int_{G^2(\mathbb{R}^2)} (P^n(x,A) - P^n(y,A))\, dP^m(x,dy)\right| \leq \eta(n).$$

Due to (8.5), we have $\lim_{n\to\infty} \eta(n) = 0$, which implies that $\nu(A) := \lim_{n\to\infty} P^n(x,A)$ exists. It follows that for $n \in \mathbb{N}$

$$\sup_{x \in G^2(\mathbb{R}^2)} \|P^n(x,.) - \nu\|_{\mathrm{tv}} \leq \eta(n) \leq \exp(-\lambda n).$$

Note that due to (8.5), ν does not depend on $x \in G^2(\mathbb{R}^2)$. It is easy to see that ν is a stationary measure for the Markov chain $(\mathbf{X}_i)_{i\geq 1}$. In fact, ν is the unique stationary measure for this Markov chain. This can be deduced from the fact that $\nu(A) = \lim_{n\to\infty} P(\mathbf{X}_n \in A)$ for $A \in \mathcal{B}_{G^2(\mathbb{R}^2)}$.

Proof of the claim:
We will show that the sequence $(\eta(n))_{n\geq 1}$ is submultiplicative i.e.

$$\eta(m+n) \leq \eta(m)\eta(n) \quad \text{for } m,n \in \mathbb{N}. \tag{8.6}$$

From this fact, we can then deduce the claim as follows:
As we assumed that $1/C \leq p(x,y) \leq C$, we have $\theta := \eta(1) \leq 1 - \frac{1}{C^2} < 1$. (To see this, use that $P(x,A) = \int_A p(x,z)\mu(dz)$ and distinguish the case $\mu(A) \leq 1/C$ from the case $\mu(A) > 1/C$.) By submultiplicativity, we obtain

$$\eta(n) \leq (\eta(1))^n = \theta^n = \exp(-\lambda n)$$

for $\lambda = -\log(\theta)$.

[4] proof of Theorem 6.10 and subsequent remarks.

We now establish (8.6):
Let $x, y \in G^2\left(\mathbb{R}^2\right)$, $A \in \mathcal{B}_{G^2(\mathbb{R}^2)}$ and m, $n \in \mathbb{N}$. We have (again by the Chapman-Kolmogorov equation)

$$P^{m+n}(x, A) - P^{m+n}(y, A) = \int_{G^2(\mathbb{R}^2)} P^m(z, A)[P^n(x, dz) - P^n(y, dz)].$$

Now, we assume that $f : G^2\left(\mathbb{R}^2\right) \to \mathbb{R}$ is a function satisfying $\sup_{x,y \in G^2(\mathbb{R}^2)} |f(x) - f(y)| \leq K$ for some constant $K > 0$ and $\mu = \mu_1 - \mu_2$ is the difference of two probability measures μ_1 and μ_2 such that $\|\mu\|_{\text{tv}} \leq \delta$ for some $\delta > 0$. Then, we have

$$\begin{aligned} \left|\int f \; d\mu\right| &= \inf_{c \in \mathbb{R}} \left|\int (f - c) d\mu\right| \qquad (8.7) \\ &\leq 2 \inf_{c \in \mathbb{R}} \left\{ \sup_{x \in G^2(\mathbb{R}^2)} |f(x) - c| \right\} \|\mu\|_{\text{tv}} \leq K\delta, \end{aligned}$$

where in (8.7), we have used that $\int c \; d\mu = 0$ for $c \in \mathbb{R}$.
If we choose $f(z) = P^m(z, A)$ and $\mu = P^n(x, .) - P^n(y, .)$, then $\|\mu\|_{\text{tv}} \leq \eta(n)$ and we obtain the inequality

$$\left|P^{m+n}(x, A) - P^{m+n}(y, A)\right| \leq \sup_{z_1, z_2 \in G^2(\mathbb{R}^2)} |P^m(z_1, A) - P^m(z_2, A)| \, \eta(n).$$

Hence,

$$\left\|P^{m+n}(x, .) - P^{m+n}(y, .)\right\|_{\text{tv}} \leq \eta(m)\eta(n).$$

Since $x, y \in G^2\left(\mathbb{R}^2\right)$ are arbitrary, this implies $\eta(m + n) \leq \eta(m)\eta(n)$. $\square$

Proposition 8.5 (exponential covariance bound)**.** *Let $n \in \mathbb{N}$, $1 \leq i_1 \leq \cdots \leq i_n$ and $l \in \{1, \ldots, n-1\}$ and let $g : \{G^2\left(\mathbb{R}^2\right)\}^l \to \mathbb{R}$ and $h : \{G^2\left(\mathbb{R}^2\right)\}^{n-l} \to \mathbb{R}$ be measurable functions such that $|g(\mathbf{X}_{i_1}, \ldots, \mathbf{X}_{i_l})| \leq G$ and $\left|h(\mathbf{X}_{i_{l+1}}, \ldots, \mathbf{X}_{i_n})\right| \leq H$ for two positive constants G and H. Then, there exists a constant $\lambda > 0$, not depending on G and H*[5]*, such that*

$$\begin{aligned} &\left|E\left[g(\mathbf{X}_{i_1}, \ldots, \mathbf{X}_{i_l}) \; h(\mathbf{X}_{i_{l+1}}, \ldots, \mathbf{X}_{i_n})\right] - E\left[g(\mathbf{X}_{i_1}, \ldots, \mathbf{X}_{i_l})\right] E\left[h(\mathbf{X}_{i_{l+1}}, \ldots, \mathbf{X}_{i_n})\right]\right| \\ &\leq 4GH \exp(-\lambda(i_{l+1} - i_l)). \qquad (8.8) \end{aligned}$$

In particular, for $1 \leq i_1 \leq i_2$ and $k, m \in \{1, 2\}$, we have

$$\left|E\left[\mathbf{X}_{i_1}^{1;k} \; \mathbf{X}_{i_2}^{1;m}\right]\right| = \left|E\left[\xi_{i_1}^{1;k} \; \xi_{i_2}^{1;m}\right]\right| \leq 4M^2 \exp(-\lambda(i_2 - i_1)). \qquad (8.9)$$

[5] λ is the same constant as in Proposition 8.4.

Proof. Let $n \in \mathbb{N}$ and $l \in \{1, \ldots, n-1\}$ be fixed. The first part of the proof is the same as for the real-valued Markov chain (proof of Proposition 6.6). We have

$$\begin{aligned}
&\left|E\left[g(\mathbf{X}_{i_1}, \ldots, \mathbf{X}_{i_l})\; h(\mathbf{X}_{i_{l+1}}, \ldots, \mathbf{X}_{i_n})\right] - E\left[g(\mathbf{X}_{i_1}, \ldots, \mathbf{X}_{i_l})\right] E\left[h(\mathbf{X}_{i_{l+1}}, \ldots, \mathbf{X}_{i_n})\right]\right| \\
&\leq 4GH\alpha\left(\sigma\left(g(\mathbf{X}_{i_1}, \ldots, \mathbf{X}_{i_l})\right), \sigma\left(h(\mathbf{X}_{i_{l+1}}, \ldots, \mathbf{X}_{i_n})\right)\right) \qquad (8.10) \\
&\leq 4GH\alpha(\sigma(\mathbf{X}_1), \sigma(\mathbf{X}_{i_{l+1}-i_l+1})), \qquad (8.11)
\end{aligned}$$

where in (8.10), we have applied the covariance bound for two bounded random variables (Lemma 6.7) and (8.11) follows from the Markov property and stationarity. We set $\mathcal{G}_i := \sigma(\mathbf{X}_i)$ for $i \in \mathbb{N}$ and $m = i_{l+1} - i_l + 1$. Then,

$$\begin{aligned}
\alpha(\mathcal{G}_1, \mathcal{G}_m) &= \sup_{A \in \mathcal{G}_1, B \in \mathcal{G}_m} |P(A \cap B) - P(A)P(B)| \\
&= \sup_{A, B \in \mathcal{B}_{G^2(\mathbb{R}^2)}} |P(\mathbf{X}_1 \in A \wedge \mathbf{X}_m \in B) - P(\mathbf{X}_1 \in A)P(\mathbf{X}_m \in B)| \\
&= \sup_{A, B \in \mathcal{B}_{G^2(\mathbb{R}^2)}} \left|\int_A P^{m-1}(x, B)\; \nu(dx) - \nu(A)\; \nu(B)\right| \\
&= \sup_{A, B \in \mathcal{B}_{G^2(\mathbb{R}^2)}} \left|\int_A (P^{m-1}(x, B) - \nu(B))\; \nu(dx)\right| \\
&\leq \sup_{A, B \in \mathcal{B}_{G^2(\mathbb{R}^2)}} \int_A \left|P^{m-1}(x, B) - \nu(B)\right| \nu(dx) \\
&\leq \sup_{x \in G^2(\mathbb{R}^2)} \left\|P^{m-1}(x, .) - \nu\right\|_{\mathrm{tv}} \\
&\leq \exp(-\lambda(m-1)) = \exp(-\lambda(i_{l+1} - i_l)),
\end{aligned}$$

where[6] the last inequality is due to Proposition 8.4. Combining the above results gives (8.8). □

As in the case of independent random variables, we will need moment bounds for partial sums of the real-valued random variables $\xi_i^{1;m}$, $i \in \mathbb{N}$, $m \in \{1, 2\}$, where $\xi_i = \log(\mathbf{X}_i)$. The above covariance estimate enables us to derive such bounds for the first and second moments. For the third-order moments, we will use Yokoyama's moment bound again.

Lemma 8.6. *There exists a constant $K > 0$ such that for all $n \in \mathbb{N}$, for $k, \overline{k} \in \{1, \ldots, n\}$, $k \leq \overline{k}$, and $l, m \in \{1, 2\}$, we have*

$$\left|E\left[\frac{1}{n}\left(\sum_{i=1}^{k} \xi_i^{1;l}\right)\left(\sum_{j=1}^{\overline{k}} \xi_j^{1;m}\right)\right]\right| \leq K\; \frac{k}{n} \leq K.$$

[6]Some steps in the above sequence of equalities and inequalities are taken from [9].

Proof. We will use stationarity of the Markov chain $(\mathbf{X}_i)_{i\geq 1}$ several times in this proof. We have

$$\left| E\left[\left(\sum_{i=1}^{k} \xi_i^{1;l}\right)\left(\sum_{j=1}^{\bar{k}} \xi_j^{1;m}\right)\right]\right| \leq\ k\left|E\left[\xi_1^{1;l}\xi_1^{1;m}\right]\right| + \sum_{i=1}^{k}\sum_{\substack{j=1\\ j\neq i}}^{\bar{k}}\left|E\left[\xi_i^{1;l}\xi_j^{1;m}\right]\right|.$$

Let $i \in \{1,\dots,k\}$. Using the covariance estimate (8.9), we get

$$\sum_{\substack{j=1\\ j\neq i}}^{\bar{k}}\left|E\left[\xi_i^{1;l}\xi_j^{1;m}\right]\right| \leq \sum_{j=2}^{\infty}\left|E\left[\xi_1^{1;l}\xi_j^{1;m}\right]\right| \leq \sum_{j=1}^{\infty} 4M^2\exp(-\lambda j) =: \overline{K} < \infty.$$

Hence, for $\beta := \max\left\{E\left[\left(\xi_1^{1;1}\right)^2\right], E\left[\left(\xi_1^{1;2}\right)^2\right]\right\}$, we have

$$\left| E\left[\left(\sum_{i=1}^{k} \xi_i^{1;l}\right)\left(\sum_{j=1}^{\bar{k}} \xi_j^{1;m}\right)\right]\right| \leq k(\beta + \overline{K}). \tag{8.12}$$

Setting $K := \beta + \overline{K}$ gives the claim. $\square$

By applying Schwarz's inequality and then the inequality (8.12), one obtains

Lemma 8.7. *There exists a constant $\widehat{K} > 0$ such that for all $n \in \mathbb{N}$, for $k \in \{1,\dots,n\}$ and $m \in \{1,2\}$, we have*

$$E\left[\frac{1}{\sqrt{n}}\left|\sum_{i=1}^{k}\xi_i^{1;m}\right|\right] \leq \widehat{K}\sqrt{\frac{k}{n}} \leq \widehat{K}.$$

Lemma 8.8. *There exists a constant $\overline{C} > 0$ such that for all $n \in \mathbb{N}$, for $k \in \{1,\dots,n\}$ and for $m_1, m_2 \in \mathbb{N}_0$ satisfying $m_1 + m_2 = 3$, we have*

$$E\left[\frac{1}{n^{3/2}}\left|\sum_{i=1}^{k}\xi_i^{1;1}\right|^{m_1}\left|\sum_{i=1}^{k}\xi_i^{1;2}\right|^{m_2}\right] \leq \overline{C}\left(\frac{k}{n}\right)^{3/2} \leq \overline{C}.$$

Proof. The proof is the same as in the case of independent random variables except for the argumentation why Yokoyama's moment bound (Theorem 5.3) can be applied. For this, we consider the strong mixing coefficients of the stationary sequence $(\xi_i^{1;j})_{i\geq 1}$, $j \in \{1,2\}$: We have

$$\alpha_i = \alpha\left(\xi_1^{1;j}, \xi_{i+1}^{1;j}\right) \leq \alpha\left(\sigma(\mathbf{X}_1), \sigma(\mathbf{X}_{i+1})\right) \leq \exp(-\lambda i),$$

where $\lambda > 0$ is the constant introduced in Proposition 8.4 and the last inequality was established in the proof of the exponential covariance bound (Proposition 8.5). Consequently, $\lim_{i\to\infty} \alpha_i = 0$ i.e. the sequence $(\xi_i^{1;j})_{i\geq 1}$ is strong mixing. In addition, we have

$$\sum_{i=1}^{\infty}(i+1)^{1/2}\alpha_i^{1/4} < \infty.$$

Thus, we can choose $\delta = 1$ in Theorem 5.3. It is clear that the sequence $(\xi_i^{1;j})_{i\geq 1}$ satisfies the remaining conditions of this theorem. □

8.3 Proof

8.3.1 Step 1: Tightness

Claim: The sequence $(\mathbf{W}^{(n)})_{n\geq 1}$ is tight in $C^{0,\alpha-Höl}\left([0,1], G^2\left(\mathbb{R}^2\right)\right)$ for any $\alpha \in (0, 1/2)$.

Proof of the claim:

By the same arguments as in the proof of the BFH theorem[7] (Theorem 4.2), it suffices to show that for all $p \in \mathbb{N}$

$$E\left[\sum_{m=1}^{2} |\pi_{1;m}(\log(\mathbf{X}_1 \otimes \cdots \otimes \mathbf{X}_k))|^{4p} + |\pi_{2;12}(\log(\mathbf{X}_1 \otimes \cdots \otimes \mathbf{X}_k))|^{2p}\right] = O\left(k^{2p}\right).$$

Note that up to this point in the argumentation, we have not used independence of the $\mathbf{X}_i$ in the proof of the BFH theorem.

As seen before, it follows from the Campbell-Baker-Hausdorff formula that

$$\pi_{1;m}(\log(\mathbf{X}_1 \otimes \cdots \otimes \mathbf{X}_k)) = \sum_{i=1}^{k} \xi_i^{1;m} \quad \text{for } m \in \{1,2\} \quad \text{and}$$

$$\pi_{2;12}(\log(\mathbf{X}_1 \otimes \cdots \otimes \mathbf{X}_k)) = \sum_{i=1}^{k} \xi_i^{2;12} + \frac{1}{2}\sum_{1\leq i<j\leq k}\left(\xi_i^{1;1}\xi_j^{1;2} - \xi_i^{1;2}\xi_j^{1;1}\right).$$

Hence, it is enough to prove the following three bounds:

$$E\left[\left(\sum_{i=1}^{k}\xi_i^{1;1}\right)^{4p}\right] = O\left(k^{2p}\right); \tag{8.13}$$

$$E\left[\left(\sum_{i=1}^{k}\xi_i^{2;12}\right)^{2p}\right] = O\left(k^{2p}\right); \tag{8.14}$$

$$E\left[\left(\sum_{1\leq i<j\leq k}\xi_i^{1;1}\xi_j^{1;2} - \sum_{1\leq i<j\leq k}\xi_i^{1;2}\xi_j^{1;1}\right)^{2p}\right] = O\left(k^{2p}\right). \tag{8.15}$$

[7] see (4.8) on p. 44.

(8.14) is easy to see: It follows from the boundedness assumption (8.1) that for $i \in \mathbb{N}$

$$\left|\xi_i^{2:12}\right| < M + M^2,$$

which implies (8.14).
In order to establish (8.13) and (8.15), it suffices to show that

Proposition 8.9. *For $\eta_i \in \left\{\xi_i^{1;1}, \xi_i^{1;2}, -\xi_i^{1;2}\right\}$, we have*

$$\sum_{1\leq i_1\leq\cdots\leq i_{4p}\leq k} \left| E\left[\prod_{m=1}^{4p} \eta_{i_m}\right]\right| = O\left(k^{2p}\right).$$

Proof. By the covariance bound (8.8), we have:
For $n \in \mathbb{N}$, $1 \leq i_1 \leq \cdots \leq i_n$ and $l \in \{1, \ldots, n-1\}$

$$\left| E\left[\prod_{m=1}^{n} \eta_{i_m}\right] - E\left[\prod_{m=1}^{l} \eta_{i_m}\right] E\left[\prod_{m=l+1}^{n} \eta_{i_m}\right]\right| \leq 4M^n \exp(-\lambda(i_{l+1} - i_l)). \tag{8.16}$$

In view of the following argumentation, it is important to note that η_i is centered and $|\eta_i| \leq M$ for all $i \in \mathbb{N}$.
We distinguish several cases for the spacing of the indices i_m: Either each index is close to its neighbors, or there is (at least) one index which is far away from its neighbors - we say this index is "isolated"[8] – or each index has at least one neighbor that is close and there is at least one large gap between two neighbors.
More formally, we define $I := \{i = (i_1, \ldots i_{4p}) : 1 \leq i_1 \leq \cdots \leq i_{4p} \leq k\}$ and $k_p := \left\lfloor k^{1/(4p)} \right\rfloor$ and consider the following subsets of I:

$$\begin{aligned}
I_1 &:= \left\{ i \in I : \max_{m=2,\ldots,4p} i_m - i_{m-1} \leq k_p \right\}; \\
I_2 &:= \{ i \in I : i_2 - i_1 > k_p \vee i_{4p} - i_{4p-1} > k_p \}; \\
I_3 &:= \{ i \in I : \exists j \in \{3, \ldots, 4p-2\} : \min\{i_j - i_{j-1}, i_{j+1} - i_j\} > k_p \}; \\
I_4 &:= \{ i \in I : \forall m \in \{2, \ldots, 4p-1\} : \min\{i_m - i_{m-1}, i_{m+1} - i_m\} \leq k_p \\
&\qquad \wedge \exists j \in \{3, \ldots, 4p-1\} : i_j - i_{j-1} > k_p \}.
\end{aligned}$$

Note that $I_1 \cup I_2 \cup I_3 \cup I_4 = I$, so that the claim will follow once we have proven that

$$\sum_{I_n} \left| E\left[\prod_{m=1}^{4p} \eta_{i_m}\right]\right| = O\left(k^{2p}\right) \tag{8.17}$$

for $n = 1, 2, 3, 4$. In the rest of the proof, we will establish these four upper bounds.

[8] For technical reasons, we will distinguish the case that the smallest or the largest index is isolated from the case that an index with two neighbors is isolated.

Case n $= 1$: $\max_{m=2,\dots,4p}\ i_m - i_{m-1} \le k_p$.

In the sum (8.17), there are at most $k(k_p)^{4p-1} < k^2 \le k^{2p}$ summands. Hence, $M^{4p}k^{2p}$ is an upper bound for this sum.

We will use the bound (8.16) to find upper bounds for the sums over I_2, I_3 and I_4. We set $C := 4M^{4p}$.

Case n $= 2$: We assume $i_2 - i_1 > k_p$. (The case $i_{4p} - i_{4p-1} > k_p$ is analogous.)

Then, we have

$$\left| E\left[\prod_{m=1}^{4p} \eta_{i_m}\right]\right| \le |E[\eta_{i_1}]| \left| E\left[\prod_{m=2}^{4p} \eta_{i_m}\right]\right| + C\exp(-\lambda k_p) = C\exp(-\lambda k_p).$$

As there are less than k^{4p} summands in the sum (8.17),

$$k^{4p} C\exp(-\lambda k_p)$$

is an upper bound for this sum and thus, it is of order $O(1)$.

Case n $= 3$: $\exists j \in \{3,\dots,4p-2\} : \min\{i_j - i_{j-1}, i_{j+1} - i_j\} > k_p$.

By applying (8.16) twice, we obtain

$$\begin{aligned}
&\left| E\left[\prod_{m=1}^{4p} \eta_{i_m}\right]\right| \le \left| E\left[\prod_{m=1}^{j-1} \eta_{i_m}\right]\right| \left| E\left[\prod_{m=j}^{4p} \eta_{i_m}\right]\right| + C\exp(-\lambda k_p)\\
\le\ & \left| E\left[\prod_{m=1}^{j-1} \eta_{i_m}\right]\right| \left(\left|E[\eta_{i_j}]\right| \left| E\left[\prod_{m=j+1}^{4p} \eta_{i_m}\right]\right| + C\exp(-\lambda k_p)\right)\\
&+ C\exp(-\lambda k_p)\\
=\ & \left| E\left[\prod_{m=1}^{j-1} \eta_{i_m}\right]\right| C\exp(-\lambda k_p) + C\exp(-\lambda k_p)\\
\le\ & (M^{j-1}+1) C\exp(-\lambda k_p).
\end{aligned}$$

By the same argumentation as in the Case $n = 2$, the sum (8.17) over I_3 is of order $O(1)$, too.

Case n $= 4$: $\forall m \in \{2,\dots,4p-1\} : \min\{i_m - i_{m-1}, i_{m+1} - i_m\} \le k_p$
$\wedge\ \exists j \in \{3,\dots,4p-1\} : i_j - i_{j-1} > k_p$.

We will iteratively apply the estimate (8.16) for each index i_j satisfying $i_{j+1} - i_j > k_p$. There are at most $2p-1$ gaps (i.e. distances $> k_p$ between neighboring indices) in this case. We will distinguish the case where the number of gaps is maximal from the other cases. If the number of gaps is $\le 2p-2$ (Case (ii)), we can proceed similarly as in the above cases and use a simple upper bound for the number of different involved products of η_{i_m}. But if the number of gaps is maximal (Case (i)), then such a simple upper bound will not yield the desired result. In this case, we will apply the covariance bound (8.9) to find an upper bound for the sum over all involved products.

Case (i): $I_{4,1} := \{i \in I : i_{2m+1} - i_{2m} > k_p$ for $m = 1, 2, \dots, 2p-1$
$\wedge\ i_{2m} - i_{2m-1} \leq k_p$ for $m = 1, 2, \dots, 2p\}$.
If (8.16) is applied $2p-1$ times, one obtains

$$\left| E\left[\prod_{m=1}^{4p} \eta_{i_m}\right]\right| \leq \prod_{m=1}^{2p} \left|E[\eta_{i_{2m-1}}\eta_{i_{2m}}]\right| + (2p-1)M^{4p-2}C\exp(-\lambda k_p).$$

Now, for $m \in \{1, \dots, 2p\}$ and a fixed index $i_{2m-1} \in \{1, \dots, k\}$, we use the covariance bound (8.9) to get

$$\sum_{\substack{i_{2m}\in\{1,\dots,k\}:\\ i_{2m}-i_{2m-1}\leq k_p}} \left|E[\eta_{i_{2m-1}}\eta_{i_{2m}}]\right| \leq \sum_{l=0}^{k_p} 4M^2\exp(-\lambda l) \leq \underbrace{\sum_{l=0}^{\infty} 4M^2\exp(-\lambda l)}_{=:\overline{C}} < \infty.$$

Consequently,

$$\begin{aligned}
&\sum_{I_{4,1}} \prod_{m=1}^{2p} \left|E[\eta_{i_{2m-1}}\eta_{i_{2m}}]\right| \\
&\leq \sum_{i_1=1}^{k}\left(\sum_{\substack{i_2\in\{1,\dots,k\}:\\ i_2-i_1\leq k_p}} |E[\eta_{i_1}\eta_{i_2}]| \left(\sum_{i_3=1}^{k}\left(\sum_{\substack{i_4\in\{1,\dots,k\}:\\ i_4-i_3\leq k_p}} |E[\eta_{i_3}\eta_{i_4}]|\,(\dots)\right)\right)\right) \\
&\leq (\overline{C}k)^{2p} = O\left(k^{2p}\right).
\end{aligned}$$

Case (ii): $\exists q \in \{1, \dots, 2p-2\}$, $\exists m_1, \dots, m_q$, $2 \leq m_1 < m_2 < \dots < m_q \leq 4p-2$: $i_{m_j+1} - i_{m_j} > k_p$ for $j = 1, \dots, q \wedge i_{m+1} - i_m \leq k_p$ for $m \notin \{m_1, \dots, m_q\}$.
We fix $q \in \{1, \dots, 2p-2\}$. Then,

$$\begin{aligned}
&\left| E\left[\prod_{m=1}^{4p} \eta_{i_m}\right]\right| \\
\leq\ &\left|E[\eta_{i_1}\cdots\eta_{i_{m_1}}]\right| \left|E[\eta_{i_{m_1+1}}\cdots\eta_{i_{m_2}}]\right| \cdots \left|E[\eta_{i_{m_q+1}}\cdots\eta_{i_{4p}}]\right| \\
&+ qM^{4p-2}C\exp(-\lambda k_p).
\end{aligned} \tag{8.18}$$

Now, we find an upper bound for the number of different products of the form (8.18) under the conditions for the indices i_m given above. There are at most

$$\begin{aligned}
&k(k_p)^{m_1-1}k(k_p)^{m_2-m_1-1}\cdots k(k_p)^{4p-m_q-1} \\
&= k^{q+1}\,(k_p)^{4p-(q+1)} \\
&\leq k^{2p-1}\left(\left\lfloor k^{1/(4p)}\right\rfloor\right)^{4p-2} \leq k^{2p}
\end{aligned}$$

such products. Each of these products is $\leq M^{4p}$.
By summing over $q \in \{1, \dots, 2p-2\}$, we get the following upper bound for the sum of all products of the form (8.18): $2pM^{4p}k^{2p}$.

Consequently, sum (8.17) over I_4 is of order $O\left(k^{2p}\right)$.

$\square$

This completes the proof of tightness. $\square$

8.3.2 Step 2: Additional auxiliary results

To characterize the limit process and handle the dependencies, we will need similar auxiliary results as in Chapter 6 for the real-valued Markov chain.
As in the previous chapter, we consider the 3-dimensional stochastic process $\widetilde{S}^{(n)}$ as defined in (7.2) and the relations (7.3) and (7.4) hold (see p. 70f.).

The following result is the analogon of Lemma 6.10. As in Chapter 6, let $L(n) = \lfloor sn^{1/10} \rfloor$.

Lemma 8.10. *Let $h : \mathbb{R}^3 \to \mathbb{R}$ be a continuously partially differentiable function with compact support. Then, we have*

$$(i) \quad \left| \sum_{j=\lceil n\bar{s}\rceil+1}^{\lceil nt\rceil} \frac{1}{n} E\left[h\left(\widetilde{S}^{(n)}\left(\frac{j-L(n)}{n}\right)\right)\right] - \sum_{j=\lceil n\bar{s}\rceil+1}^{\lceil nt\rceil} \frac{1}{n} E\left[h\left(\widetilde{S}^{(n)}\left(\frac{j}{n}\right)\right)\right]\right| = O\left(\frac{L(n)}{n}\right).$$

$$(ii) \quad \left| \sum_{j=\lceil n\bar{s}\rceil+1}^{\lceil nt\rceil} \frac{1}{n} E\left[h\left(\widetilde{S}^{(n)}\left(\frac{j-L(n)}{n}\right)\right) Y_m^{(n)}\left(\frac{j-L(n)}{n}\right)\right] - \sum_{j=\lceil n\bar{s}\rceil+1}^{\lceil nt\rceil} \frac{1}{n} E\left[h\left(\widetilde{S}^{(n)}\left(\frac{j}{n}\right)\right) Y_m^{(n)}\left(\frac{j}{n}\right)\right]\right| = O\left(\frac{L(n)}{\sqrt{n}}\right) \quad \textit{for } m \in \{1,2\}.$$

$$(iii) \quad \left| \sum_{j=\lceil n\bar{s}\rceil+1}^{\lceil nt\rceil} \frac{1}{n} E\left[h\left(\widetilde{S}^{(n)}\left(\frac{j-L(n)}{n}\right)\right) Y_m^{(n)}\left(\frac{j-L(n)}{n}\right) \cdot Y_l^{(n)}\left(\frac{j-L(n)}{n}\right)\right] - \sum_{j=\lceil n\bar{s}\rceil+1}^{\lceil nt\rceil} \frac{1}{n} E\left[h\left(\widetilde{S}^{(n)}\left(\frac{j}{n}\right)\right) Y_m^{(n)}\left(\frac{j}{n}\right) \cdot Y_l^{(n)}\left(\frac{j}{n}\right)\right]\right| = O\left(\frac{L(n)}{\sqrt{n}}\right) \quad \textit{for } m, l \in \{1,2\}.$$

Proof. We will only prove (ii) and (iii). (i) is easy. In the sequel, let $m, l \in \{1, 2\}$. Note that $\left|\xi_i^{1;m}\right| \leq M$ implies

$$\left|Y_m^{(n)}\left(\frac{k}{n}\right)\right| = \left|\frac{1}{\sqrt{n}}\sum_{i=1}^{k}\xi_i^{1;m}\right| \leq M\sqrt{n} \text{ for } k \in \{1, \ldots, n\}. \tag{8.19}$$

Consequently, we have

$$\begin{aligned}
(ii) \qquad & \left|\sum_{j=\lceil n\bar{s}\rceil+1}^{\lceil nt\rceil} \frac{1}{n} E\left[h\left(\widetilde{S}^{(n)}\left(\frac{j-L(n)}{n}\right)\right) Y_m^{(n)}\left(\frac{j-L(n)}{n}\right)\right]\right. \\
& \left. - \sum_{j=\lceil n\bar{s}\rceil+1}^{\lceil nt\rceil} \frac{1}{n} E\left[h\left(\widetilde{S}^{(n)}\left(\frac{j}{n}\right)\right) Y_m^{(n)}\left(\frac{j}{n}\right)\right]\right| \\
& \leq \frac{\|h\|_\infty}{n}\left(\sum_{j=\lceil n\bar{s}\rceil-L(n)+1}^{\lceil n\bar{s}\rceil}\left|Y_m^{(n)}\left(\frac{j}{n}\right)\right| + \sum_{j=\lceil nt\rceil-L(n)+1}^{\lceil nt\rceil}\left|Y_m^{(n)}\left(\frac{j}{n}\right)\right|\right) \\
& \leq 2M\|h\|_\infty \frac{L(n)}{\sqrt{n}}.
\end{aligned}$$

Note that the simple deterministic bound for the expression in (iii) does not tend to 0 as $n \to \infty$. But applying the moment bound given in Lemma 8.7 to $Y_l^{(n)}(.)$ will yield the desired bound (For $Y_m^{(n)}(.)$, we can use the deterministic bound (8.19) again.):

$$\begin{aligned}
(iii) \quad & \left|\sum_{j=\lceil n\bar{s}\rceil+1}^{\lceil nt\rceil} \frac{1}{n} E\left[h\left(\widetilde{S}^{(n)}\left(\frac{j-L(n)}{n}\right)\right) Y_m^{(n)}\left(\frac{j-L(n)}{n}\right) Y_l^{(n)}\left(\frac{j-L(n)}{n}\right)\right]\right. \\
& \left. - \sum_{j=\lceil n\bar{s}\rceil+1}^{\lceil nt\rceil} \frac{1}{n} E\left[h\left(\widetilde{S}^{(n)}\left(\frac{j}{n}\right)\right) Y_m^{(n)}\left(\frac{j}{n}\right) Y_l^{(n)}\left(\frac{j}{n}\right)\right]\right| \\
& \leq \frac{M\|h\|_\infty}{\sqrt{n}}\left(\sum_{j=\lceil n\bar{s}\rceil-L(n)+1}^{\lceil n\bar{s}\rceil} E\left[\left|Y_l^{(n)}\left(\frac{j}{n}\right)\right|\right] + \sum_{j=\lceil nt\rceil-L(n)+1}^{\lceil nt\rceil} E\left[\left|Y_l^{(n)}\left(\frac{j}{n}\right)\right|\right]\right) \\
& \leq 2M\widehat{K}\|h\|_\infty \frac{L(n)}{\sqrt{n}}.
\end{aligned}$$

□

The analogon of Lemma 6.11 involves the two-sided extension $(\mathbf{X}_i)_{-\infty<i<\infty}$ of the stationary Markov chain $(\mathbf{X}_i)_{i\geq 1}$. It can be established by similar arguments as in the first part of the proof of Lemma 6.11. By using stationarity of the sequence $(\xi_i^{1;m})_{-\infty<i<\infty}$, $m \in \{1, 2\}$, and the covariance bound (8.9), one obtains

Lemma 8.11. *For $j \in \mathbb{N}$ and $l, m \in \{1,2\}$, we have*

$$\lim_{n\to\infty} \sum_{k=j-\lfloor L(n)/2\rfloor+1}^{j-1} E\left[\xi_k^{1;l}\xi_j^{1;m}\right] = \sum_{k=2}^{\infty} E\left[\xi_1^{1;l}\xi_k^{1;m}\right].$$

8.3.3 Step 3: Characterization of the limit process

We assume that $\widetilde{S} = (Y_1, Y_2, Z)$ is the limit process of a weakly convergent subsequence of $\widetilde{S}^{(n)} = \left(Y_1^{(n)}, Y_2^{(n)}, Z^{(n)}\right)$.
The aim is to establish the following
Claim: $Q := P\widetilde{S}^{-1}$ is the unique solution to the martingale problem for (a,b) starting at $(0,0,0)$, where the maps $a : \mathbb{R}^3 \to S_3$ and $b : \mathbb{R}^3 \to \mathbb{R}^3$ are defined as in (8.3) and (8.4).

Proof of the claim:
We first establish uniqueness: Note that the map $\sigma : \mathbb{R}^3 \to M_3$ defined by

$$\sigma(x,y,z) = \begin{pmatrix} \sqrt{v} & 0 & 0 \\ c/\sqrt{v} & \sqrt{vw-c^2}/\sqrt{v} & 0 \\ (cx/\sqrt{v} - \sqrt{v}y)/2 & \sqrt{vw-c^2}x/(2\sqrt{v}) & 0 \end{pmatrix}$$

is Lipschitz continuous and satisfies $\sigma(x,y,z)\sigma(x,y,z)^T = a(x,y,z)$. Thus, by Theorem 5.13, the martingale problem for (a,b) is well-posed.
In the rest of the proof, we will show that $P\widetilde{S}^{-1}$ is a solution to the martingale problem for (a,b) starting at $(0,0,0)$. $P(\widetilde{S}(0) = (0,0,0)) = 1$ is clear. The second-order differential operator corresponding to the above martingale problem is

$$\begin{aligned} L_{a,b}F(x,y,z) =& d\ \partial_3 F(x,y,z) + \frac{1}{2}v\ \partial_{11}F(x,y,z) + \frac{1}{2}w\ \partial_{22}F(x,y,z) \\ &+ \frac{1}{4}\left(\frac{1}{2}wx^2 + \frac{1}{2}vy^2 - cxy\right)\partial_{33}F(x,y,z) \\ &+ c\ \partial_{12}F(x,y,z) \\ &+ \frac{1}{2}(cx - vy)\,\partial_{13}F(x,y,z) + \frac{1}{2}(wx - cy)\,\partial_{23}F(x,y,z). \end{aligned}$$

The first part of the argumentation is similar to the one for independent random variables. We recall the notation introduced there:
$\mathcal{F}_j := \sigma(\mathbf{X}_i : 1 \le i \le j)$ for $j \in \{1,\dots,n\}$ and $\mathcal{F}_0$ is the trivial σ-algebra. Let $F \in C_0^\infty(\mathbb{R}^3)$, $s,\ t \in [0,1]$, $s < t$ and let $\phi : C([0,1],\mathbb{R}^3) \to \mathbb{R}$ be a continuous, bounded, $\mathcal{M}_s$-measurable map. We define $\phi_n := \phi\left(\widetilde{S}^{(n)}\right)$ and $\phi_S := \phi\left(\widetilde{S}\right)$.
Again, we aim at proving (7.9). It is a consequence of Lebesgue's dominated convergence theorem that it suffices to show that for $\epsilon \in (0, t-s)$

$$E\left[\left(F\left(\widetilde{S}(t)\right) - F\left(\widetilde{S}(s+\epsilon)\right) - \int_{s+\epsilon}^{t} L_{a,b}F\left(\widetilde{S}(u)\right)du\right)\phi_S\right] = 0.$$

For the rest of the proof, we fix $\epsilon \in (0, t-s)$. We will establish (7.10) with s replaced by $s+\epsilon$. It is clear that (7.11) with s replaced by $s+\epsilon$ also holds in this setting. We set $\overline{s} := s+\epsilon$. The aim is thus to prove that

$$\begin{aligned}
&\lim_{n\to\infty} E\left[\sum_{j=\lceil n\overline{s}\rceil+1}^{\lceil nt\rceil}\left(E\left[F\left(\widetilde{S}^{(n)}\left(\frac{j}{n}\right)\right) \,\Big|\, \mathcal{F}_{j-1}\right] - F\left(\widetilde{S}^{(n)}\left(\frac{j-1}{n}\right)\right)\right)\phi_n\right]\\
&\quad = E\left[\left(\int_{\overline{s}}^{t} L_{a,b}F\left(\widetilde{S}(u)\right)du\right)\phi_S\right]. \qquad (8.20)
\end{aligned}$$

Clearly, the Taylor expansion of $F\left(\widetilde{S}^{(n)}\left(\frac{j}{n}\right)\right)$, $j \in \{1,\dots,n\}$, is the same as for independent random variables $(\mathbf{X}_i)_{i\geq 1}$, so that (7.12) also holds in this setting. To see that the expectation involving the remainder term is of order $O(n^{-3/2})$ (i.e. (7.13) holds), we need the corresponding moment bounds for Markov chains (Lemma 8.6 - 8.8).

For all the summands in (7.12) which do not have expectations of order $o(1/n)$, we will do another Taylor expansion and then we will apply the covariance bound (8.8) to deal with the dependencies. For the application of this bound, we will use that the conditional expectations given $\mathcal{F}_{j-1}$ appearing in (7.12) are bounded, measurable functions of $\mathbf{X}_{j-1}$:

Let $h \in \left\{\pi_{1;1}, \pi_{1;2}, \pi_{2;12}, \pi_{1;1}^2, \pi_{1;2}^2, \pi_{1;1}\pi_{1;2}, \pi_{1;1}\pi_{2;12}, \pi_{1;2}\pi_{2;12}\right\}$. Then,

$$E[h(\mathbf{X}_j)|\mathcal{F}_{j-1}] = \int_{G^2(\mathbb{R}^2)} h(y)p(\mathbf{X}_{j-1}, y)\ \mu(dy) = g_h(\mathbf{X}_{j-1}),$$

where $g_h : G^2\left(\mathbb{R}^2\right) \to \mathbb{R}$ is a bounded, measurable function. g_h is bounded as we have

$$\left|\int_{G^2(\mathbb{R}^2)} h(y)p(\mathbf{X}_{j-1}, y)\ \mu(dy)\right| \leq M^2 \sup_{x,y\in G^2(\mathbb{R}^2)} p(x,y) \leq M^2 C.$$

In the sequel, we will illustrate how the dependencies can be handled in the summands involving $\partial_3 F$. We chose these summands as they require the most involved argumentation. Similar, but simpler arguments also apply to the other summands. First, we do Taylor expansion with expansion point $\widetilde{S}^{(n)}\left(\frac{j-L(n)}{n}\right)$:

$$\begin{aligned}
&\partial_3 F\left(\widetilde{S}^{(n)}\left(\frac{j-1}{n}\right)\right)\\
&= \partial_3 F\left(\widetilde{S}^{(n)}\left(\frac{j-L(n)}{n}\right)\right)\\
&+ \partial_{13} F\left(\widetilde{S}^{(n)}\left(\frac{j-L(n)}{n}\right)\right)\left(\frac{1}{\sqrt{n}}\sum_{k=j-L(n)+1}^{j-1}\xi_k^{1;1}\right)\\
&+ \partial_{23} F\left(\widetilde{S}^{(n)}\left(\frac{j-L(n)}{n}\right)\right)\left(\frac{1}{\sqrt{n}}\sum_{k=j-L(n)+1}^{j-1}\xi_k^{1;2}\right)
\end{aligned}$$

$$
\begin{aligned}
&+ \partial_{33}F\left(\widetilde{S}^{(n)}\left(\frac{j-L(n)}{n}\right)\right)\left(\frac{1}{\sqrt{n}}\left(\frac{1}{\sqrt{n}}\sum_{k=j-L(n)+1}^{j-1}\xi_k^{2;12}\right.\right.\\
&\qquad\left.\left.+\frac{1}{2}\sum_{k=j-L(n)+1}^{j-1}\xi_k^{1;2}Y_1^{(n)}\left(\frac{k-1}{n}\right)-\frac{1}{2}\sum_{k=j-L(n)+1}^{j-1}\xi_k^{1;1}Y_2^{(n)}\left(\frac{k-1}{n}\right)\right)\right)\\
&+ \dots .
\end{aligned}
$$

Hence, we have to consider the following expectations[9]:

$$
\begin{aligned}
&\frac{1}{\sqrt{n}}E\left[\partial_3F\left(\widetilde{S}^{(n)}\left(\frac{j-1}{n}\right)\right)\left(\frac{1}{\sqrt{n}}E\left[\xi_j^{2;12}\,\middle|\,\mathcal{F}_{j-1}\right]\right.\right.\\
&\qquad\left.\left.+\frac{1}{2}Y_1^{(n)}\left(\frac{j-1}{n}\right)E\left[\xi_j^{1;2}\,\middle|\,\mathcal{F}_{j-1}\right]-\frac{1}{2}Y_2^{(n)}\left(\frac{j-1}{n}\right)E\left[\xi_j^{1;1}\,\middle|\,\mathcal{F}_{j-1}\right]\right)\phi_n\right]\\
=\;&\frac{1}{n}E\left[\partial_3F\left(\widetilde{S}^{(n)}\left(\frac{j-L(n)}{n}\right)\right)E\left[\xi_j^{2;12}\,\middle|\,\mathcal{F}_{j-1}\right]\phi_n\right]\\
&+\frac{1}{2\sqrt{n}}E\left[\partial_3F\left(\widetilde{S}^{(n)}\left(\frac{j-L(n)}{n}\right)\right)Y_1^{(n)}\left(\frac{j-1}{n}\right)E\left[\xi_j^{1;2}\,\middle|\,\mathcal{F}_{j-1}\right]\phi_n\right] \qquad (8.21)\\
&-\frac{1}{2\sqrt{n}}E\left[\partial_3F\left(\widetilde{S}^{(n)}\left(\frac{j-L(n)}{n}\right)\right)Y_2^{(n)}\left(\frac{j-1}{n}\right)E\left[\xi_j^{1;1}\,\middle|\,\mathcal{F}_{j-1}\right]\phi_n\right]\\
&+\frac{1}{2n}E\left[\partial_{13}F\left(\widetilde{S}^{(n)}\left(\frac{j-L(n)}{n}\right)\right)\left(\sum_{k=j-L(n)+1}^{j-1}\xi_k^{1;1}\right)Y_1^{(n)}\left(\frac{j-1}{n}\right)\right.\\
&\qquad\left.E\left[\xi_j^{1;2}\,\middle|\,\mathcal{F}_{j-1}\right]\phi_n\right]\\
&-\frac{1}{2n}E\left[\partial_{13}F\left(\widetilde{S}^{(n)}\left(\frac{j-L(n)}{n}\right)\right)\left(\sum_{k=j-L(n)+1}^{j-1}\xi_k^{1;1}\right)Y_2^{(n)}\left(\frac{j-1}{n}\right)\right.\\
&\qquad\left.E\left[\xi_j^{1;1}\,\middle|\,\mathcal{F}_{j-1}\right]\phi_n\right] \qquad (8.22)\\
&+\frac{1}{2n}E\left[\partial_{23}F\left(\widetilde{S}^{(n)}\left(\frac{j-L(n)}{n}\right)\right)\left(\sum_{k=j-L(n)+1}^{j-1}\xi_k^{1;2}\right)Y_1^{(n)}\left(\frac{j-1}{n}\right)\right.\\
&\qquad\left.E\left[\xi_j^{1;2}\,\middle|\,\mathcal{F}_{j-1}\right]\phi_n\right]\\
&-\frac{1}{2n}E\left[\partial_{23}F\left(\widetilde{S}^{(n)}\left(\frac{j-L(n)}{n}\right)\right)\left(\sum_{k=j-L(n)+1}^{j-1}\xi_k^{1;2}\right)Y_2^{(n)}\left(\frac{j-1}{n}\right)\right.\\
&\qquad\left.E\left[\xi_j^{1;1}\,\middle|\,\mathcal{F}_{j-1}\right]\phi_n\right]
\end{aligned}
$$

[9]Note that we omit expressions including the factor $n^{-3/2}$, since they can all be shown to be of order $o(1/n)$. For those expressions which also include $Y_m^{(n)}\left(\frac{j-1}{n}\right)$, $m \in \{1,2\}$, as a factor, we have to apply Lemma 8.7 to prove this fact.

$$
\begin{aligned}
&+\frac{1}{4n}E\left[\partial_{33}F\left(\widetilde{S}^{(n)}\left(\frac{j-L(n)}{n}\right)\right)\left(\sum_{k=j-L(n)+1}^{j-1}\xi_k^{1;2}Y_1^{(n)}\left(\frac{k-1}{n}\right)\right)\right.\\
&\qquad\left.\cdot Y_1^{(n)}\left(\frac{j-1}{n}\right)E\left[\xi_j^{1;2}\;\middle|\;\mathcal{F}_{j-1}\right]\phi_n\right]\\
&-\frac{1}{4n}E\left[\partial_{33}F\left(\widetilde{S}^{(n)}\left(\frac{j-L(n)}{n}\right)\right)\left(\sum_{k=j-L(n)+1}^{j-1}\xi_k^{1;2}Y_1^{(n)}\left(\frac{k-1}{n}\right)\right)\right.\\
&\qquad\left.\cdot Y_2^{(n)}\left(\frac{j-1}{n}\right)E\left[\xi_j^{1;1}\;\middle|\;\mathcal{F}_{j-1}\right]\phi_n\right] &(8.23)\\
&-\frac{1}{4n}E\left[\partial_{33}F\left(\widetilde{S}^{(n)}\left(\frac{j-L(n)}{n}\right)\right)\left(\sum_{k=j-L(n)+1}^{j-1}\xi_k^{1;1}Y_2^{(n)}\left(\frac{k-1}{n}\right)\right)\right.\\
&\qquad\left.\cdot Y_1^{(n)}\left(\frac{j-1}{n}\right)E\left[\xi_j^{1;2}\;\middle|\;\mathcal{F}_{j-1}\right]\phi_n\right]\\
&+\frac{1}{4n}E\left[\partial_{33}F\left(\widetilde{S}^{(n)}\left(\frac{j-L(n)}{n}\right)\right)\left(\sum_{k=j-L(n)+1}^{j-1}\xi_k^{1;1}Y_2^{(n)}\left(\frac{k-1}{n}\right)\right)\right.\\
&\qquad\left.\cdot Y_2^{(n)}\left(\frac{j-1}{n}\right)E\left[\xi_j^{1;1}\;\middle|\;\mathcal{F}_{j-1}\right]\phi_n\right]\\
&+o\left(\frac{1}{n}\right).
\end{aligned}
$$

The next step is to apply the covariance estimate (8.8) to the above summands such that the resulting error terms tend to 0 if we apply $\lim_{n\to\infty}\sum_{j=\lceil n\bar{s}\rceil+1}^{\lceil nt\rceil}$ to them. To achieve this, we will often have to split expectations of sums of random variables into different summands and then apply (8.8) in different ways to them. We will demonstrate this procedure for the three numbered summands above. Similar arguments also apply to the other summands.

For the following three examples, we will assume that $n\in\mathbb{N}$ is so large that we have[10] $L(n)\leq\lceil n\bar{s}\rceil-\lceil ns\rceil$ and thus ϕ_n only depends on the random variables $X_1,\ldots,X_{\lceil n\bar{s}\rceil-L(n)}$.

We start with the summand (8.21). In this case, we split $Y_1^{(n)}\left(\frac{j-1}{n}\right)$ into two partial sums. In the expectation involving the lower partial sum, we can isolate $E[\xi_j^{1;2}]$, whereas in the upper partial sum, $E\left[\partial_3F\left(\widetilde{S}^{(n)}\left(\frac{j-L(n)}{n}\right)\right)\phi_n\right]$ can be isolated:

$$
\frac{1}{2\sqrt{n}}E\left[\partial_3F\left(\widetilde{S}^{(n)}\left(\frac{j-L(n)}{n}\right)\right)Y_1^{(n)}\left(\frac{j-1}{n}\right)E\left[\xi_j^{1;2}\;\middle|\;\mathcal{F}_{j-1}\right]\phi_n\right]
$$

[10] Recall that we also relied on this assumption for the covariance bound applications in Chapter 6.

$$= \frac{1}{2n} E\left[\partial_3 F\left(\widetilde{S}^{(n)}\left(\frac{j-L(n)}{n}\right)\right)\left(\sum_{k=1}^{j-\lfloor L(n)/2\rfloor} \xi_k^{1;1}\right)\phi_n\right] \underbrace{E\left[\xi_j^{1;2}\right]}_{=0}$$
$$+ O\left(\frac{1}{\sqrt{n}}\exp\left(-\lambda\left(\left\lfloor\frac{L(n)}{2}\right\rfloor - 1\right)\right)\right)$$
$$+ \frac{1}{2n} E\left[\partial_3 F\left(\widetilde{S}^{(n)}\left(\frac{j-L(n)}{n}\right)\right)\phi_n\right] E\left[\sum_{k=j-\lfloor L(n)/2\rfloor+1}^{j-1} \xi_k^{1;1}\xi_j^{1;2}\right]$$
$$+ O\left(\frac{1}{\sqrt{n}}\exp\left(-\lambda\left(L(n) - \left\lfloor\frac{L(n)}{2}\right\rfloor + 1\right)\right)\right).$$

The fact that the constant GH in the above covariance bound applications is of order $O(1/\sqrt{n})$ follows from the moment bound given in Lemma 8.7 and the fact that the functions $\partial_3 F$ and ϕ are bounded.
Since $\sqrt{n}\exp(-\lambda(\lfloor L(n)/2\rfloor - 1)) = o(1/n)$, we thus obtain

$$\sum_{j=\lceil n\bar{s}\rceil+1}^{\lceil nt\rceil} \frac{1}{2\sqrt{n}} E\left[\partial_3 F\left(\widetilde{S}^{(n)}\left(\frac{j-L(n)}{n}\right)\right) Y_1^{(n)}\left(\frac{j-1}{n}\right) E\left[\xi_j^{1;2}\;\middle|\;\mathcal{F}_{j-1}\right]\phi_n\right]$$
$$= \sum_{j=\lceil n\bar{s}\rceil+1}^{\lceil nt\rceil} \frac{1}{2n} E\left[\partial_3 F\left(\widetilde{S}^{(n)}\left(\frac{j-L(n)}{n}\right)\right)\phi_n\right] E\left[\sum_{k=j-\lfloor L(n)/2\rfloor+1}^{j-1} \xi_k^{1;1}\xi_j^{1;2}\right] + o\left(\frac{1}{n}\right).$$

To deal with expression (8.22), we have to consider four summands. First, we split the sum over k into a lower and an upper part. For each of these parts, we then split $Y_2^{(n)}\left(\frac{j-1}{n}\right)$ into two suitable partial sums. We obtain

$$\frac{1}{2n} E\left[\partial_{13} F\left(\widetilde{S}^{(n)}\left(\frac{j-L(n)}{n}\right)\right)\left(\sum_{k=j-L(n)+1}^{j-1} \xi_k^{1;1}\right) Y_2^{(n)}\left(\frac{j-1}{n}\right) E\left[\xi_j^{1;1}\;\middle|\;\mathcal{F}_{j-1}\right]\phi_n\right]$$
$$= \frac{1}{2n} E\left[\partial_{13} F\left(\widetilde{S}^{(n)}\left(\frac{j-L(n)}{n}\right)\right)\left(\sum_{k=j-L(n)+1}^{j-\lfloor L(n)/2\rfloor} \xi_k^{1;1}\right)\right.$$
$$\left.\cdot Y_2^{(n)}\left(\frac{j-\lfloor L(n)/2\rfloor}{n}\right)\phi_n\right] \underbrace{E\left[\xi_j^{1;1}\right]}_{=0} + O\left(\frac{L(n)}{n}\exp\left(-\lambda\left(\left\lfloor\frac{L(n)}{2}\right\rfloor - 1\right)\right)\right)$$
$$+ \frac{1}{2n^{3/2}} E\left[\partial_{13} F\left(\widetilde{S}^{(n)}\left(\frac{j-L(n)}{n}\right)\right)\left(\sum_{k=j-L(n)+1}^{j-\lfloor L(n)/2\rfloor} \xi_k^{1;1}\right)\right.$$
$$\left.\cdot\left(\sum_{l=j-\lfloor L(n)/2\rfloor+1}^{j-1} \xi_l^{1;2}\right) E[\xi_j^{1;1}|\mathcal{F}_{j-1}]\;\phi_n\right]$$
$$+ \frac{1}{2n} E\left[\partial_{13} F\left(\widetilde{S}^{(n)}\left(\frac{j-L(n)}{n}\right)\right) Y_2^{(n)}\left(\frac{j-L(n)}{n}\right)\phi_n\right]$$

$$\cdot E\left[\sum_{k=j-\lfloor L(n)/2\rfloor+1}^{j-1} \xi_k^{1;1}\xi_j^{1;1}\right] + O\left(\frac{L(n)}{n}\exp\left(-\lambda\left(L(n)-\left\lfloor\frac{L(n)}{2}\right\rfloor+1\right)\right)\right)$$

$$+\frac{1}{2n^{3/2}}E\left[\partial_{13}F\left(\widetilde{S}^{(n)}\left(\frac{j-L(n)}{n}\right)\right)\left(\sum_{k=j-\lfloor L(n)/2\rfloor+1}^{j-1}\xi_k^{1;1}\right)\right.$$

$$\left.\cdot\left(\sum_{l=j-L(n)+1}^{j-1}\xi_l^{1;2}\right)E[\xi_j^{1;1}|\mathcal{F}_{j-1}]\ \phi_n\right].$$

To see that the constant GH in the exponential covariance bound is of order $O(L(n)/n)$, we need Lemma 8.7 again.
As before, expressions including the factor $n^{-3/2}$ are of order $o\left(1/n\right)$. Observe that $L(n)\exp(-\lambda(\lfloor L(n)/2\rfloor-1))=o(1/n)$, so that we obtain

$$\sum_{j=\lceil n\bar{s}\rceil+1}^{\lceil nt\rceil}\frac{1}{2n}E\left[\partial_{13}F\left(\widetilde{S}^{(n)}\left(\frac{j-L(n)}{n}\right)\right)\left(\sum_{k=j-L(n)+1}^{j-1}\xi_k^{1;1}\right)Y_2^{(n)}\left(\frac{j-1}{n}\right)\right.$$

$$\left.E\left[\xi_j^{1;1}\,\middle|\,\mathcal{F}_{j-1}\right]\phi_n\right]$$

$$=\sum_{j=\lceil n\bar{s}\rceil+1}^{\lceil nt\rceil}\frac{1}{2n}E\left[\partial_{13}F\left(\widetilde{S}^{(n)}\left(\frac{j-L(n)}{n}\right)\right)Y_2^{(n)}\left(\frac{j-L(n)}{n}\right)\phi_n\right]$$

$$\cdot E\left[\sum_{k=j-\lfloor L(n)/2\rfloor+1}^{j-1}\xi_k^{1;1}\xi_j^{1;1}\right]+o\left(\frac{1}{n}\right).$$

Expression (8.23) is an example for the most involved case we have to handle. We have

$$\frac{1}{4n}E\left[\partial_{33}F\left(\widetilde{S}^{(n)}\left(\frac{j-L(n)}{n}\right)\right)\left(\sum_{k=j-L(n)+1}^{j-1}\xi_k^{1;2}Y_1^{(n)}\left(\frac{k-1}{n}\right)\right)\right.$$

$$\left.\cdot Y_2^{(n)}\left(\frac{j-1}{n}\right)E\left[\xi_j^{1;1}\,\middle|\,\mathcal{F}_{j-1}\right]\phi_n\right]$$

$$=\frac{1}{4n}E\left[\partial_{33}F\left(\widetilde{S}^{(n)}\left(\frac{j-L(n)}{n}\right)\right)\left(\sum_{k=j-L(n)+1}^{j-\lfloor L(n)/2\rfloor}\xi_k^{1;2}Y_1^{(n)}\left(\frac{k-1}{n}\right)\right)\right.$$

$$\left.\cdot Y_2^{(n)}\left(\frac{j-\lfloor L(n)/2\rfloor}{n}\right)\phi_n\right]E\left[\xi_j^{1;1}\right]+O\left(\frac{L(n)}{n}\exp\left(-\lambda\left(\left\lfloor\frac{L(n)}{2}\right\rfloor-1\right)\right)\right)$$

$$+\frac{1}{4n^{3/2}}E\left[\partial_{33}F\left(\widetilde{S}^{(n)}\left(\frac{j-L(n)}{n}\right)\right)\left(\sum_{k=j-L(n)+1}^{j-\lfloor L(n)/2\rfloor}\xi_k^{1;2}Y_1^{(n)}\left(\frac{k-1}{n}\right)\right)\right.$$

$$
\begin{aligned}
&\cdot \left(\sum_{l=j-\lfloor L(n)/2 \rfloor+1}^{j-1} \xi_l^{1;2} \right) E[\xi_j^{1;1} | \mathcal{F}_{j-1}] \ \phi_n \Bigg] \\
&+ \frac{1}{4n} E \left[\partial_{33} F \left(\widetilde{S}^{(n)} \left(\frac{j-L(n)}{n} \right) \right) Y_1^{(n)} \left(\frac{j-L(n)}{n} \right) \right. \\
&\left. \cdot Y_2^{(n)} \left(\frac{j-L(n)}{n} \right) \phi_n \right] E \left[\sum_{k=j-\lfloor L(n)/2 \rfloor+1}^{j-1} \xi_k^{1;2} \xi_j^{1;1} \right] \\
&+ O \left(\frac{L(n)}{n} \exp \left(-\lambda \left(L(n) - \left\lfloor \frac{L(n)}{2} \right\rfloor + 1 \right) \right) \right) \\
&+ \frac{1}{4n^{3/2}} E \left[\partial_{33} F \left(\widetilde{S}^{(n)} \left(\frac{j-L(n)}{n} \right) \right) \left(\sum_{k=j-\lfloor L(n)/2 \rfloor+1}^{j-1} \xi_k^{1;2} Y_1^{(n)} \left(\frac{j-L(n)}{n} \right) \right) \right. \\
&\left. \cdot \left(\sum_{l=j-L(n)+1}^{j-1} \xi_l^{1;2} \right) E[\xi_j^{1;1} | \mathcal{F}_{j-1}] \ \phi_n \right] \\
&+ \frac{1}{4n^{3/2}} E \left[\partial_{33} F \left(\widetilde{S}^{(n)} \left(\frac{j-L(n)}{n} \right) \right) \left(\sum_{k=j-\lfloor L(n)/2 \rfloor+1}^{j-1} \xi_k^{1;2} \left(\sum_{m=j-L(n)+1}^{k-1} \xi_m^{1;1} \right) \right) \right. \\
&\left. \cdot Y_2^{(n)} \left(\frac{j-1}{n} \right) E[\xi_j^{1;1} | \mathcal{F}_{j-1}] \ \phi_n \right].
\end{aligned}
$$

The fact that the constant GH is of order $O(L(n)/n)$ in the two above covariance bound applications follows from the moment bound given in Lemma 8.6.

The next step is to apply the auxiliary results given in Step 2 to the relevant summands obtained in the covariance bound applications. We will demonstrate this procedure for the summands of the form (8.22). Note that Lemma 7.6 with s replaced by $\overline{s} = s + \epsilon$ also holds in this setting. The proof of this fact is analogous to the proof of Lemma 7.6 – in place of Lemma 7.4, we have to apply Lemma 8.7.
We have

$$
\begin{aligned}
&\lim_{n\to\infty} \sum_{j=\lceil n\overline{s} \rceil+1}^{\lceil nt \rceil} -\frac{1}{2n} E \left[\partial_{13} F \left(\widetilde{S}^{(n)} \left(\frac{j-L(n)}{n} \right) \right) \left(\sum_{k=j-L(n)+1}^{j-1} \xi_k^{1;1} \right) Y_2^{(n)} \left(\frac{j-1}{n} \right) \right. \\
&\left. \cdot E \left[\xi_j^{1;1} \ \middle| \ \mathcal{F}_{j-1} \right] \phi_n \right] \\
= \ & - \lim_{n\to\infty} \sum_{j=\lceil n\overline{s} \rceil+1}^{\lceil nt \rceil} \frac{1}{2n} E \left[\partial_{13} F \left(\widetilde{S}^{(n)} \left(\frac{j-L(n)}{n} \right) \right) Y_2^{(n)} \left(\frac{j-L(n)}{n} \right) \phi_n \right] \\
&\cdot E \left[\sum_{k=j-\lfloor L(n)/2 \rfloor+1}^{j-1} \xi_k^{1;1} \xi_j^{1;1} \right] \qquad (8.24)
\end{aligned}
$$

$$= -\underbrace{\frac{1}{2}\sum_{j=2}^{\infty} E\left[\xi_1^{1;1}\xi_j^{1;1}\right]}_{=s_{11}} \lim_{n\to\infty} \sum_{j=\lceil n\bar{s}\rceil+1}^{\lceil nt\rceil} \frac{1}{n} E\left[\partial_{13}F\left(\widetilde{S}^{(n)}\left(\frac{j-L(n)}{n}\right)\right)\right.$$

$$\left.\cdot Y_2^{(n)}\left(\frac{j-L(n)}{n}\right)\phi_n\right] \tag{8.25}$$

$$= -\frac{s_{11}}{2}\lim_{n\to\infty}\sum_{j=\lceil n\bar{s}\rceil+1}^{\lceil nt\rceil}\left(\frac{1}{n}E\left[\partial_{13}F\left(\widetilde{S}^{(n)}\left(\frac{j}{n}\right)\right)Y_2^{(n)}\left(\frac{j}{n}\right)\phi_n\right]+O\left(\frac{L(n)}{\sqrt{n}}\right)\right) \tag{8.26}$$

$$= -\frac{s_{11}}{2}\lim_{n\to\infty}\left(E\left[\left(\int_{\bar{s}}^{t}\partial_{13}F\left(\widetilde{S}^{(n)}(u)\right)Y_2^{(n)}(u)\,du\right)\phi_n\right]+O\left(n^{-1/2}\right)\right) \tag{8.27}$$

$$= -\frac{s_{11}}{2}E\left[\left(\int_{\bar{s}}^{t}\partial_{13}F\left(\widetilde{S}(u)\right)Y_2(u)\,du\right)\phi_S\right]. \tag{8.28}$$

(8.24) follows from the above covariance bound applications. (8.25) is due to Lemma 8.11. In (8.26), we have applied Lemma 8.10, (*ii*). (8.27) is a consequence of Lemma 7.6 with s replaced by $\bar{s}$. Finally, (8.28) follows from weak convergence of $\widetilde{S}^{(n)}$ to $\widetilde{S}$ and the continuous mapping theorem.
Using analogous arguments for all the other summands we have to consider, we obtain (8.20):

$$\lim_{n\to\infty} E\left[\sum_{j=\lceil n\bar{s}\rceil+1}^{\lceil nt\rceil}\left(E\left[F\left(\widetilde{S}^{(n)}\left(\frac{j}{n}\right)\right)\,\Big|\,\mathcal{F}_{j-1}\right]-F\left(\widetilde{S}^{(n)}\left(\frac{j-1}{n}\right)\right)\right)\phi_n\right]$$

$$= \left(E\left[\xi_1^{2;12}\right]+\frac{1}{2}(s_{12}-s_{21})\right)E\left[\left(\int_{\bar{s}}^{t}\partial_3F\left(\widetilde{S}(u)\right)du\right)\phi_S\right]$$

$$+ \left(\frac{1}{2}\sigma_1^2+s_{11}\right)E\left[\left(\int_{\bar{s}}^{t}\partial_{11}F\left(\widetilde{S}(u)\right)du\right)\phi_S\right]$$

$$+ \left(\frac{1}{2}\sigma_2^2+s_{22}\right)E\left[\left(\int_{\bar{s}}^{t}\partial_{22}F\left(\widetilde{S}(u)\right)du\right)\phi_S\right]$$

$$+ \frac{1}{4}E\left[\left(\int_{\bar{s}}^{t}\partial_{33}F\left(\widetilde{S}(u)\right)\left(\left(\frac{1}{2}\sigma_2^2+s_{22}\right)(Y_1(u))^2+\left(\frac{1}{2}\sigma_1^2+s_{11}\right)(Y_2(u))^2\right.\right.\right.$$

$$\left.\left.\left. - \ (\gamma+s_{12}+s_{21})\,Y_1(u)Y_2(u)\right)du\right)\phi_S\right]$$

$$+ \ (\gamma+s_{12}+s_{21})\,E\left[\left(\int_{\bar{s}}^{t}\partial_{12}F\left(\widetilde{S}(u)\right)du\right)\phi_S\right]$$

$$+ \frac{1}{2}E\left[\left(\int_{\bar{s}}^{t}\partial_{13}F\left(\widetilde{S}(u)\right)\left((\gamma+s_{12}+s_{21})\,Y_1(u)-\left(\sigma_1^2+2s_{11}\right)Y_2(u)\right)du\right)\phi_S\right]$$

$$+ \frac{1}{2}E\left[\left(\int_{\bar{s}}^{t}\partial_{23}F\left(\widetilde{S}(u)\right)\left(\left(\sigma_2^2+2s_{22}\right)Y_1(u)-(\gamma+s_{12}+s_{21})\,Y_2(u)\right)du\right)\phi_S\right]$$

$$= E\left[\left(\int_{\bar{s}}^{t}L_{a,b}F(\widetilde{S}(u))du\right)\phi_S\right],$$

which is what we wanted to prove. $\square$

A Appendix

Two results related to weak convergence

Let (S, d) be a metric space. Note that if S is separable, then the product σ-algebra on $S \times S$ coincides with the Borel σ-algebra induced by the product topology on $S \times S$ (see [1], p. 225). In the sequel, we endow $S \times S$ with this σ-algebra.
If S is separable and X and Y are two S-valued random variables defined on the same probability space, then $d(X, Y)$ is a (real-valued) random variable.

Proposition A.1 ([1], Th. 4.1)**.** *Let the metric space S be separable and let $(X_n)_{n\geq 1}$ and $(Y_n)_{n\geq 1}$ be sequences of S-valued random variables defined on the same probability space $(\Omega, \mathcal{F}, P)$.*
If $X_n \overset{D}{\to} X$ and $d(X_n, Y_n) \overset{P}{\to} 0$, then $Y_n \overset{D}{\to} X$.

For two probability measures P and Q, we denote the product measure by $P \otimes Q$.

Theorem A.2 ([1], Th. 3.2)**.** *Let the metric space S be separable and let $(P_n)_{n\geq 1}$ and $(Q_n)_{n\geq 1}$ be sequences of probability measures on S. Then, we have*

$$\text{w-lim}\, P_n \otimes Q_n = P \otimes Q \Leftrightarrow \text{w-lim}\, P_n = P \textit{ and } \text{w-lim}\, Q_n = Q.$$

Stationary sequences of random variables

Definition A.3. A sequence of random variables $(X_i)_{i\geq 1}$ is called **stationary** if for each $k \in \mathbb{N}$, $(X_k, X_{k+1}, \dots)$ has the same distribution as $(X_1, X_2, \dots)$.

Some authors use the term "strictly stationary" to refer to the above property.

Transformation theorem for image measures

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, $(\Omega', \mathcal{F}')$ a measurable space and $\phi : \Omega \to \Omega'$ a $\mathcal{F}$-$\mathcal{F}'$-measurable map.

Theorem A.4. *Let $f : \Omega' \to \mathbb{R}$ be measurable. f is integrable with respect to $\mu\phi^{-1}$ if and only if $f \circ \phi : \Omega \to \mathbb{R}$ is integrable with respect to μ.*
If these two equivalent conditions are satisfied, we have

$$\int f \; d\left(\mu\phi^{-1}\right) = \int f \circ \phi \; d\mu.$$

Nilpotent groups

Let G be a group and U_1, U_2 two subgroups of G. Then, $[U_1, U_2]$ denotes the subgroup of G generated by the commutators $[x, y] := xyx^{-1}y^{-1}$, where $x \in U_1$ and $y \in U_2$.

Definition A.5 ([8])**.** Let $G^1 := [G, G]$ and $G^n := [G, G^{n-1}]$ for $n \geq 2$. Then, $(G^n)_{n\geq 1}$ is called **descending central series** of G.
We say that G is **nilpotent** if there exists $n \in \mathbb{N}$ such that $G^n = \{e\}$, where e denotes the identity element of G.

References

[1] Billingsley, P.: *Convergence of Probability Measures.* Wiley series in probability and statistics, John Wiley and sons, Inc., 1968.

[2] Bradley, R. C.: *Introduction to Strong Mixing Conditions*, vol. 1. Kendrick Press, Utah, 2007.

[3] Breuillard, E., Friz, P. K., and Huesmann, M.: *From Random Walks to Rough Paths.* Proc. Amer. Math. Soc., **137**(10) (2009), 3487-3496.

[4] Crépel, P., and Raugi, A.: *Théorème central limite sur les groupes nilpotents.* Ann. Inst. H. Poincaré, Sect. B, **14**(2) (1978), 145-164.

[5] Friz, P. K., and Victoir, N. B.: *Multidimensional Stochastic Processes as Rough Paths: Theory and Applications.* Cambridge studies in advanced mathematics, vol. 120, Cambridge University Press, 2010.

[6] Hamadouche, D.: *Invariance Principles in Hölder Spaces.* Portugaliae Mathematica, **57**(2) (2000), 127-151.

[7] Hamadouche, D. and Taleb, Y.: *Hölderian Version of Donsker-Prohorov's Invariance Principle.* IAENG Int. J. Appl. Math., **39**(1) (2009), 1-8.

[8] Hilgert, J. and Neeb, K.-H.: *Lie-Gruppen und Lie-Algebren.* Vieweg, Braunschweig/Wiesbaden, 1991.

[9] Jones, G. L.: *On the Markov chain central limit theorem.* Probability Surveys, **1** (2004), 299-320.

[10] Kirillov, A., Jr.: *An Introduction to Lie Groups and Lie Algebras.* Cambridge studies in advanced mathematics, vol. 113, Cambridge University press, Cambridge, 2008.

[11] Lamperti, J.: *On Convergence of Stochastic Processes.* Trans. Amer. Math. Soc., **104**(3) (1962), 430-435.

[12] Lejay, A.: *An Introduction to Rough Paths.* Sém. Prob. XXXVII, Lecture Notes in Mathematics, vol. 1832, 1-59, Springer, Berlin/Heidelberg, 2003.

[13] Lyons, T. J.: *Differential equations driven by rough signals.* Rev. Mat. Iberoamericana, **14**(2) (1998), 215-310.

[14] Lyons, T. J., Caruana, M., and Lévy, T.: *Differential Equations Driven by Rough Paths.* École d'Été de Probabilités de Saint-Flour XXXIV - 2004, Lecture Notes in Mathematics, vol. 1908, Springer, Berlin/Heidelberg, 2007.

[15] McKean, H. P., Jr.: *Stochastic Integrals*. Academic Press, Inc., New York, 1969.

[16] Ramasubramanian, S.: *Diffusions and the Martingale Problem of Stroock and Varadhan*. Connected at Infinity: a selection of Mathematics by Indians, Texts and Readings in Mathematics, vol. 25, 105-129, Hindustan Book Agency, New Delhi, 2003.

[17] Revuz, D., and Yor, M.: *Continuous Martingales and Brownian Motion*. Corrected Third Printing of the Third Ed., Grundl. Math. Wiss., vol. 293, Springer, Berlin/Heidelberg, 2005.

[18] Rogers, L. C. G., and Williams, D.: *Diffusions, Markov Processes, and Martingales*, vol. 2: Itô Calculus, second Ed.. Cambridge University Press, Cambridge, 2000.

[19] Stroock, D. W., and Varadhan, S.R.S.: *Limit Theorems for Random Walks on Lie Groups*. Sankhyā: The Indian Journal of Statistics, Series A, **35**(3) (1973), 277-294.

[20] Stroock, D. W., and Varadhan, S.R.S.: *Multidimensional Diffusion Processes*, Reprint of the 1997 Ed.. Grundl. Math. Wiss., vol. 233, Springer, Berlin/Heidelberg, 2006.

[21] Varadhan, S.R.S.: *Probability Theory*. Courant Lecture Notes in Mathematics, vol. 7, Amer. Math. Soc. and Courant Institute of Mathematical Sciences at New York University, 2001.

[22] Yokoyama, R.: *Moment bounds for stationary mixing sequences*. Z. Wahrscheinlichkeitstheorie verw. Gebiete, **52**(1) (1980), 45-57.