



HAL
open science

Area anomaly and generalized drift of iterated sums for hidden Markov walks

Olga Lopusanschi, Damien Simon

► **To cite this version:**

Olga Lopusanschi, Damien Simon. Area anomaly and generalized drift of iterated sums for hidden Markov walks. 2017. hal-01586794

HAL Id: hal-01586794

<https://hal.science/hal-01586794>

Preprint submitted on 13 Sep 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

Area anomaly and generalized drift of iterated sums for hidden Markov walks

Olga Lopusanschi & Damien Simon

September 13, 2017

Abstract

Following our result from [13], we study the convergence in rough path topology of a certain class of discrete processes, the hidden Markov walks, to a Brownian motion with an area anomaly. This area anomaly, which is a new object keeps track of the time-correlation of the discrete models and brings into light the question of embeddings of discrete processes into continuous time. We also identify an underlying combinatorial structure in the hidden Markov walks, which turns out to be a generalization of the occupation time from the classical ergodic theorem in the spirit of rough paths. Finally, we construct a rough path out of iterated sums of a hidden Markov walk, compare it to the classical canonical lift of the piecewise linear interpolation of the process and analyse its convergence to a non-geometric rough path.

Contents

1	Introduction	2
1.1	On the importance of discrete models in the rough paths theory.	2
1.2	Structure of the article and notations.	5
2	Main results and map of the paper	6
2.1	Rough paths models from microscopic models.	6
2.1.1	Main theorems: rough paths from hidden Markov walks. .	6
2.1.2	An easy example: the diamond and round-point models. .	12
2.2	Combinatorial structure behind the iterated sums of HMW. . . .	14
2.2.1	Iterated sums and non-geometric rough paths.	14
2.2.2	Iterated occupation times of HMW as underlying combinatorial structures of rough paths.	16
3	From Hidden Markov Walks to rough paths	20
3.1	Theory of pseudo-excursions for hidden Markov walks.	20
3.2	Elements of rough paths theory	22
3.3	Embeddings	24
3.4	Getting back to theorem 2.3	28

4	Iterated structures behind discrete time and discrete space	
	Markov chains.	35
4.1	Shuffle and quasi-shuffle products.	35
4.2	From geometric to non-geometric rough paths through hidden	
	Markov walks	37
4.2.1	Geometric rough paths and shuffle products.	37
4.2.2	A discrete construction for non-geometric rough paths. . .	38
4.3	Getting back to proposition 2.1.	38
5	Open questions	43

1 Introduction

1.1 On the importance of discrete models in the rough paths theory.

From continuous to discrete setting. Rough paths theory was introduced by T. Lyons in 1997 (see, for example, [15]) in order to provide a deterministic setting to stochastic differential equations (SDEs) of the type

$$dy_s = f(y_s)[dx_s] \tag{1}$$

where $(y_s)_s$ is a path in $\mathbb{R}^{d'}$, $(x_s)_s$ is a path in of \mathbb{R}^d of Hölder regularity $\alpha < 1$ (which is often the case for stochastic processes) and $f : \mathbb{R}^{d'} \rightarrow \text{End}(\mathbb{R}^d, \mathbb{R}^{d'})$. Whenever the classical Young integration [16] fails (which is the case for $\alpha < 1/2$), paths may be lifted (in a non-unique way) to a larger, more abstract, space, *the space of rough paths*, for which existence and uniqueness of a solution map become easier to prove.

The idea behind this theory relies on the Taylor expansion: if we want to suitably approach (1), the trajectory level is not enough to register all the relevant information when passing to the limit. Thus, we have first to make an "expansion" of the path $(x_t)_t$ by constructing, on the grounds of certain rules, the associated *rough path* $(\mathbf{x}_t)_t$, which is a continuous path in $V \oplus \dots \oplus V^{[1/\alpha]}$ and then to suitably approach it.

A particularity of the rough paths theory is that it has been developed in a continuous setting (we approximate continuous processes by other continuous processes that are smoother). Since the main goal of the rough paths theory is giving a general method for solving SDEs, the exploration of the discrete setting may seem irrelevant other than for SDE approximation (see [3]). And yet, new directions of research can be found in this domain as it appears, for example, in [12], where the authors study discrete rough integrals. We believe that developing a theory of discrete models in the rough path setting can enrich the classical stochastic calculus with new tools and results. The present paper concentrates on two issues related to this field: the ways of constructing a rough path out of a discrete model and the explicit construction of new objects arising

at the limit in rough path topology, as well as the role they may play in classical stochastic calculus.

Analysing rough paths issues through discrete models. A finite-variation path $\gamma : [0, T] \rightarrow V$ can be canonically lifted to a rough path using the sequence of iterated integrals:

$$S_{\gamma,n}(t) = \int_{0 < s_1 < s_2 < \dots < s_n < t} d\gamma(s_1) \otimes d\gamma(s_2) \otimes \dots \otimes d\gamma(s_n) \in V^{\otimes n}, \quad t \in [0, T] \quad (2)$$

For any $N \geq 2$, the corresponding canonical lift coincides with *the step- N signature* of γ given by

$$S_N(\gamma)_{0,t} = (1, S_{\gamma,1}(t), \dots, S_{\gamma,N}(t)) \in T_1^{(N)}(V), \quad t \in [0, T] \quad (3)$$

where $T_1^{(N)}(V) = \{(a_0, \dots, a_N \in \bigoplus_{k=0}^N V^{\otimes k} | a_0 = 1\}$ with the convention $V^{\otimes 0} = \mathbb{R}$.

For α -Hölder paths with $\alpha < 1$, such integrals may not be well-defined and some of the values of the vectors $S_{\gamma,n}(t)$ may be *postulated* provided that they satisfy certain algebraic relations. The absence of a canonical lift of Hölder paths to a larger space leads to some natural questions at the basis of the present paper.

The first one, answered since the very beginning of rough paths theory, deals with how classical stochastic integrals (Itô and Stratonovitch) fit within this theory. The set of rough paths is separated in two categories, one for each type of integration:

- *geometric rough paths*, whose components satisfy the chain rule, for Stratonovitch integration;
- *non-geometric rough paths*, whose components do not satisfy it, for Itô integration.

Based on discrete sequences we give explicit constructions for both types of integrals as limit processes and discuss the link with geometric and non-geometric rough paths. In particular, we will see how discrete constructions can give rise to non-geometric rough paths as limit processes.

The second issue consists in understanding the nature of the non-trivial objects in the space of rough paths; for example, two-dimensional area bubbles [11] obtained as the limit of the paths:

$$\gamma^{(n)}(t) = \left(\frac{1}{\sqrt{n}} \cos(nt), \frac{1}{\sqrt{n}} \sin(nt) \right)$$

have a well-defined rough path limit different from the constant path $\gamma(t) = (0, 0)$ (which would be the limit in the uniform topology). In this case, $S_{\gamma,1}(t) = 0$ but $S_{\gamma,2}(t) = \frac{1}{2}te_1 \wedge e_2$, with e_1 and e_2 vectors of the canonical basis of \mathbb{R}^2 .

The third (and central) question is to ask whether such non-trivial limits can coexist with the classical structure of stochastic integrals. We can give a positive answer to this question by studying well-chosen discrete models with suitable time correlations. In [13], we have constructed a class of processes on a finite-dimensional vector space V (Markov chains on periodic graphs) that converge, under certain conditions, in rough path topology (see section 3.2 for details) to

$$(S_{B,1}(t), S_{B,2}(t) + \Gamma t)_{t \in [0,1]} \quad (4)$$

where $(B_t)_t$ is a standard Brownian motion on the vector space V and Γ a deterministic antisymmetric matrix called *the area anomaly*. While $S_{B,1}(t)$ and $S_{B,2}(t)$ are obtained with classical stochastic calculus, Γ is a new object.

The present paper provides a much more general construction of such limits. The main idea relies on the following: if one wants to be able to control Γ in a limit of the type (4), one has to modify the underlying discrete models in the following way:

- one may try to create area-bubble-like corrections to the trajectories;
- these corrections have to remain small so that $S_{B,1}$ does not change;
- the corrected trajectories have to satisfy the Markov property.

Our solution consists in

- introducing time-correlations in the discrete processes, tuned so that they affect only $S_{B,2}$ after renormalization, through the framework of *the hidden Markov walks*, of which the Markov chains on periodic graphs are an example, and
- showing how the area anomaly depends on the way we *embed* the discrete paths in the set of continuous paths.

This last point brings into light the difference played by embeddings in the uniform and rough paths topologies. In the first case, it is merely a way to make sense of the convergence of a discrete path and one uses usually linear interpolation by default. In the second case, it is a mean to influence and even change the nature of the limit through the non-trivial object which is the area anomaly.

We also show how the convergence of any embedding can be studied through the convergence of a discrete sequence in $T_1^{(2)}(V)$ (more precisely, in one of its sub-groups: $G^2(V)$). This allows, in particular, to make a classification of the embeddings of a given discrete process based on the Γ they provide at the limit.

Another important question concerns the combinatorial structures that can be derived from rough paths built out of hidden Markov walks. While the algebraic structure of rough paths, based on *shuffle products*, provides nice combinatorial interpretations of iterated integrals, the iterated sums have more intricate multiplication rules, which can be interpreted using *the quasi-shuffle products*. This subject will be discussed in section 4.

Acknowledgements. D. S. is partially funded by the Grant ANR-14CE25-0014 (ANR GRAAL). D. S. and O. L. thank Lorenzo Zambotti for the stimulating discussions and useful suggestions.

1.2 Structure of the article and notations.

General structure of the article. The present article is divided into five main sections. The first two sections give a general presentation of our work as follows:

- in the present section 1, we describe the motivations and the main goals of our work;
- in section 2, we state our main results and give some additional comments and illustrations to them.

The next two sections are more technical:

- section 3 is dedicated to the convergence of rough paths constructed out of hidden Markov walks: we give a detailed description of the framework and a proof of our main theorem (3.4);
- in section 4, we explore the combinatorial properties of hidden Markov chains and the associated rough paths through iterated occupation times and their link with shuffle and quasi-shuffle products.

The last section 5 is dedicated to further research perspectives and open questions.

Notations. Throughout this paper, the following objects will often appear.

- $\mathcal{C}^{1-var}([0, T], B)$ is the space of functions valued in a Banach space B with bounded variations.
- V denotes a finite-dimensional vector space.
- E denotes a finite state space (for Markov chains).
- $T_1^{(k)}(V)$ is the graded vector space:

$$T_1^{(k)}(V) = V \oplus \dots \oplus V^{\otimes k}$$

and $G^2(V)$ a subgroup of $T_1^{(2)}(V)$ defined by:

$$\forall (a, b) \in G^2(V), \text{Sym}(b) = \frac{1}{2}(a \otimes a) \quad (5)$$

(the symmetrical part of the level $V \otimes V$ depends entirely on the level V of the object).

- Instead of $(1, a_1, \dots, a_k) \in T_1^{(k)}(V)$, we will write $(a_1, \dots, a_k) \in \bigoplus_{l=1}^k V^{\otimes l}$.

- The operator " \otimes " can denote:
 - a tensor product of two elements in V (in the formula $V \otimes V$, for example);
 - an operation on $T_1^{(2)}(V)$: $(a, b) \otimes (a', b') = (a + a', b + b' + a \otimes a')$, where the operation $a \otimes a'$ is in the sense of a tensor product.
- The operator " \wedge " is an antisymmetric law on $G^2(V)$: $(a, b) \wedge (a', b') = (a + a', b + b' + \frac{1}{2}(a \otimes a' - a' \otimes a))$;
- δ_ϵ : the standard dilatation operator on $G^2(V)$: $\delta_\epsilon(a, b) = (\epsilon a, \epsilon^2 b)$.

2 Main results and map of the paper

2.1 Rough paths models from microscopic models.

2.1.1 Main theorems: rough paths from hidden Markov walks.

Definition of hidden Markov walks. The Donsker theorem, generalized to rough paths theory in [2], shows that random walks do not leave any room for area anomaly: area bubbles are killed by the independence between random variables at each time step. Following [13], we introduce the framework of *hidden Markov walks* (HMW) whose time-correlations, as it has already been mentioned, allow additional area accumulation on the second level of the corresponding rough paths.

We first give a definition of *hidden Markov chains* (HMC), which are the "bricks" from which hidden Markov walks are constructed:

Definition 2.1 (the hidden Markov chain). *Let E be a finite state space and V a finite-dimensional vector space. A hidden Markov chain is a discrete process $(R_n, F_n)_n$ on $E \times V$ such that $R = (R_n)_n$ is a Markov chain on E and under the conditional law $\mathbb{P}(\cdot | \sigma(R))$ the variables F_n are independent and*

$$\forall k \geq 1, \forall u \in E, \mathbb{P}(F_k | R_k = u) = \mathbb{P}(F_1 | R_1 = u)$$

Hidden Markov walks are constructed out of hidden Markov chains just as simple random walks can be constructed as sums of i.i.d. variables:

Definition 2.2 (the hidden Markov walk). *Let E be a finite state space and V a finite-dimensional vector space and $(R_n, F_n)_n$ a hidden Markov chain on $E \times V$. For all $n \geq 1$, set $X_n = \sum_{k=1}^n F_k$. The discrete process $(R_n, X_n)_n$ on $E \times V$ is called a hidden Markov walk.*

Remarks.

- Throughout the paper, the transition matrix of the process (R_n) will be written Q and the conditional law of $F_k = X_k - X_{k-1}$ knowing $R_k = u \in E$ will be written $\nu(\bullet | u)$ and expectations under this law $\mathbb{E}_\nu[\bullet | u]$.

- The process $(X_n)_n$ may take values in a more general space than V (and we will encounter such examples throughout the paper). The only thing that changes in this case is the definition of the increments of the process (see property 3.1).

A Donsker-type result. The following theorem is a result on convergence of rough paths constructed by piecewise linear interpolation (like in the classical Donsker theorem) out of HMW. It is a generalization of our result from [13] to a larger class of processes.

We define the excursion times of the process $(R_k)_k$ as:

$$T_0 = 0 \tag{6}$$

$$\forall n \geq 1, T_n = \inf\{k > T_{n-1} : R_k = R_0\} \tag{7}$$

Theorem 2.1. *Let $(R_n, X_n)_n$ be a hidden Markov walk on $E \times V$ such that $(R_n)_n$ is irreducible and the increments of $(X_n)_n$ are a.s. uniformly bounded, i.e.*

$$\exists K > 0, \forall j \in \mathbb{N}, |F_j|_V \leq K \text{ a.s.} \tag{8}$$

Set $\beta = \mathbb{E}[T_1]$. Under the conditions $\mathbb{E}[X_{T_1}] = 0$ and $\mathbb{E}[X_{T_1}^{\otimes 2}] = CI_d$ for a certain $C > 0$, the rough path of the piecewise linear interpolation of $(X_k)_{k \leq N}$, renormalized by $\delta_{(\beta^{-1}NC)^{-1/2}}$, converges in the rough path topology $\mathcal{C}^\alpha([0, 1], G^2(V))$ for $\alpha < 1/2$ to a rough path whose levels are given by:

$$\begin{aligned} S_{B,1}(t) &= B_t \\ S_{B,2}(t) &= \int_{0 < s_1 < s_2 < t} \circ dB_{s_1} \otimes \circ dB_{s_2} + \Gamma t \end{aligned}$$

where $(B_t)_t$ is a standard d -dimensional Brownian motion and Γ is a deterministic antisymmetric matrix, the area anomaly, given by

$$\Gamma = C^{-1} \frac{1}{2} \mathbb{E} \left[\sum_{1 \leq k_1 < k_2 \leq T_1} F_{k_1} \otimes F_{k_2} - F_{k_2} \otimes F_{k_1} \right] \tag{9}$$

Remarks.

- The condition (8) requires a.s. uniform bound for F_j s, whereas the classical Donsker theorem (and the its rough path variant from [2]) requires a uniform bound for the moments. On one hand, most of the concrete models we can apply theorem 2.1 to satisfy (8); on the other hand, as we will see further on, the variables whose role is equivalent to the ones of i.i.d.r.v. from the classical Donsker theorem are, in our case, the pseudo-excursions.
- The definition of β implies that it may depend on the initial law of $(R_n)_n$, whereas the limit does not. The reason for this is simple: β registered the slowing down in time of the limit process which, as we will see, depends on the length of an excursion of $(R_n)_n$.

- We can actually always suppose the covariance matrix of $\mathbb{X}_{T_1}^{(1)}$ to be equal to CI_d for a certain $C > 0$ in theorem 2.3 (idem for the covariance matrix of X_{T_1} in theorems 2.1 and 2.2). The reasons for this are similar to those explained in the remark that follows immediately corollary 2.1 from [13]: we can embed the process $\mathbb{X}_{T_1}^{(1)}$ in a smaller space than V , and afterwards use a linear transformation to get the identity matrix modulo a constant.
- We can still get a result similar to theorem 2.1 if the covariance matrix of X_{T_1} is not diagonal. The difference is that $(B_t)_t$ will have a covariance matrix depending on X_{T_1} and will not be a standard Brownian motion anymore (see proposition 2.1).
- The second-level drift is actually the stochastic area of an excursion modulo a multiplicative constant, i.e. $\Gamma = C^{-1}\mathbb{E}[A_{T_1}(X)]$.
- $S_{B,2}(t)$ and can be decomposed into a symmetric part, given by $1/2(B_t \otimes B_t)$, and an antisymmetric part, given by $\mathcal{A}_t + \Gamma t$, where $(\mathcal{A}_t)_t$ is the Lévy area of $(B_t)_t$, i.e.

$$\mathcal{A}_t = \frac{1}{2} \int_{0 < s_1 < s_2 < t} dB_{s_1} \otimes dB_{s_2} - dB_{s_2} \otimes dB_{s_1}$$

This decomposition highlights the fact that $(S_{B,1}(t), S_{B,2}(t))_t$ is a geometric rough path. Note that the Lévy area does not depend on the choice of integration (Itô or Stratonovich). If we endow the space of rough paths with the antisymmetric tensor product \wedge instead of the ordinary one \otimes , the $S_{B,2}(t)$ will only keep track of the antisymmetric part of the limit: this is the approach we have adopted in our previous article [13].

Examples to which theorem 2.1 applies.

- In case of a simple random walk ($E = \{0\}$) with a piecewise linear embedding, we get the Donsker theorem for rough paths from [2] (in particular, $\Gamma = 0$), with slightly different moment conditions.
- A less trivial example is *the sum of rotating Bernoulli i.i.d.r.v.* which has been presented in our previous article [13]: $X_n = \sum_{k=1}^n i^k U_k$, where the U_k s are i.i.d. Bernoulli variables. In this case, $E = \{1, 2, 3, 4\}$ and $\Gamma \neq 0$.
- More generally, theorem 2.1 applies to any Markov chain on periodic graphs as the ones described in [13].

The question of embeddings. If we want to build a rough path associated to a sequence $(x_n)_n \in V^{\mathbb{N}}$, we first need to associate to $(x_n)_n \in V^{\mathbb{N}}$ a continuous path in V for which building a rough path makes sense. In other words, we need to *embed* the discrete sequence in the space of continuous, sufficiently regular paths in V .

Definition 2.3. We define an embedding of $(x_n)_n \in V^{\mathbb{N}}$ as a sequence $\rho = (\rho_N)_N$ of continuous paths in V such that

- $\forall N \geq 1, \rho_N : [0, N] \rightarrow V;$
- $\forall N \geq 1, \forall i \in \{1, \dots, N\}, \rho_N(i) = x_i;$
- $\forall k < N, \forall t \in [0, k], \rho_k(t) = \rho_N(t).$

We can construct an embedding by curve concatenation in the following way: we start by connecting, for all $n \in \mathbb{N}$, x_n and x_{n+1} by a continuous curve β_n such that:

$$\beta_n : [0, 1] \rightarrow V, \beta_n(0) = x_n, \beta_n(1) = x_{n+1} \quad (10)$$

The embedding is then given by

$$\rho_N = \beta_1 \cdot \dots \cdot \beta_N \quad (11)$$

where \cdot is the operation of path concatenation in V . In order to pass to the limit, we define $\iota_N : [0, 1] \rightarrow V$ as $\iota_N(t) = f(N)\rho_N(Nt)$, where $f : \mathbb{N} \rightarrow \mathbb{R}_+$ is the renormalization function (for details on embeddings see section 3.3).

In theorem 2.1 we have used the piecewise linear embedding, i.e. we have connected two consecutive points by linear interpolation. While this is the most common embedding, it is not unique. In order to define a more general embedding for a hidden Markov walk, we proceed as follows.

Lemma 2.1 (construction of an embedding for HMW). *Consider a Markov chain $(R_n, F_n)_n$ on $E \times V$. We construct a sequence of processes $(\rho_N)_N$ as follows:*

- for any $u \in E$, denote by $V_u \subset V$ the set of all possible realizations of F_1 under the law $\mathbb{P}(\bullet | R_1 = u)$;
- for $u \in E$, to every $y \in V_u$ we associate in a measurable way, for the usual Borel σ -algebras, a curve $f^y : [0, 1] \rightarrow V$ with bounded variation and we denote the set of these curves by \mathcal{B}_u ;
- for all $k \geq 1$, we associate to F_k a random variable β_k in the sense that, for a $u \in E$ and under the law $\mathbb{P}(\bullet | R_k = u)$, β_k gives for every realization y of F_k from V_u the corresponding element f^y from \mathcal{B}_u ;
- for all $N \geq 1$, we set $\rho_N = \beta_1 \cdot \dots \cdot \beta_N$.

Then $(\rho_N)_N$ is an embedding of $(X_n)_n$ in the sense of definition 3.4.

Remark. Lemma 2.1 relies on the following idea: we can then "enlarge" a hidden Markov chain $(R_n, F_n)_n$ in $E \times V$ to the sequence of triplets $(R_n, F_n, \beta_n)_n$ in $E \times V \times \mathcal{C}^{1-var}([0, 1], V)$, and construct an embedding $(\rho_N)_N$ out of $(\beta_k)_k$ as in (11).

This leads to the following generalization of theorem 2.1:

Theorem 2.2. *Let $(R_n, F_n)_n$ be a hidden Markov chain on $E \times V$ such that $(R_n)_n$ is irreducible and condition (8) is satisfied, and denote by $(R_n, X_n)_n$ the corresponding Markov walk. Denote by $\rho = (\rho_n)_n$ an embedding for $(X_n)_n$ with bounded variation constructed as in lemma 2.1.*

Set $\beta = \mathbb{E}[T_1]$. Under the conditions $\mathbb{E}[X_{T_1}] = 0$ and $\mathbb{E}[X_{T_1}^{\otimes 2}] = CI_d$ for a certain $C > 0$, the rough path canonically constructed out of ρ_N , renormalized through the dilation operator $\delta_{(\beta^{-1}NC)^{-1/2}}$, converges in the rough path topology $\mathcal{C}^{\alpha-\text{Hö}}([0, 1], G^2(V))$ with $\alpha < 1/2$ to a rough path given by:

$$\begin{aligned} S_{B,1}(t) &= B_t \\ S_{B,2}(t) &= \int_{0 < s_1 < s_2 < t} \circ dB_{s_1} \otimes \circ dB_{s_2} + \Gamma_\rho t \end{aligned}$$

where $(B_t)_t$ is a standard d -dimensional Brownian motion and Γ_ρ is an antisymmetric deterministic matrix depending on ρ as follows:

$$\Gamma_\rho = \mathbb{E} \left[\frac{1}{2} \int_{0 < s_1 < s_2 < T_1} d\rho_{T_1}(s_1) \otimes d\rho_{T_1}(s_2) - d\rho_{T_1}(s_2) \otimes d\rho_{T_1}(s_1) \right] \quad (12)$$

Remarks.

- Γ_ρ is the stochastic area of the curve ρ_{T_1} between the times 0 and T_1 .
- For the ρ -embedding, we have chosen curves of finite variation. This is a sufficient but not a necessary condition for the result of theorem 2.1 to be true. We maintain it for commodity reason, a generalisation to a weaker condition on the embedding (for example, curves of finite p -variation with a suitable p) being a question of some additional computations.

Examples to which theorem 2.2 applies: All the models to which theorem 2.1 applies and for which we connect consecutive points by curves of finite variation: the round-point model (see figure 2), models for traffic with road intersections, random walks on deformed networks and deformed periodic graphs etc.

The general theorem. In both theorems 2.1 and 2.2, we have lifted paths in V to geometric paths in $G^2(V)$. As we will see in section 3.3 (proposition 3.3), any sufficiently regular embedding ρ of a sequence in V can be encoded in a rough path in $G^2(V)$, so that we can start directly with a sequence in $G^2(V)$ and associate to it a rough path, i.e. a sufficiently regular embedding in $G^2(V)$.

The embeddings for sequences in $G^2(V)$ can be defined by analogy with the ones on V (see the remark after definition 2.3). In particular, since $G^2(V)$ is a geodesic space (see the comment after definition 34), we can define for sequences valued in $G^2(V)$ the geodesic embedding, i.e. the embedding that connects two consecutive points by a curve whose length equals to the distance between them.

In figure 1, examples of geodesic curves corresponding to different elements of $G^2(\mathbb{R}^2)$ are presented. In the case of \mathbb{R}^d , the geodesic embedding corresponds to the linear interpolation.

Remark. The geodesic embedding of a sequence in $G^2(V)$ can be represented using the increments of the sequence, just like the linear interpolation of a sequence in \mathbb{R}^d .

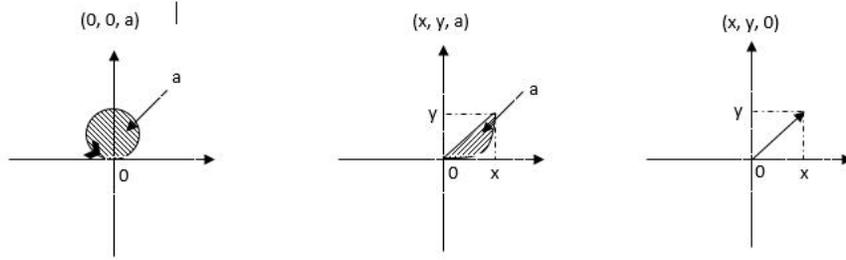


Figure 1: Examples of geodesic curves corresponding to elements of $G^2(V)$.

Thus, theorem 2.2 becomes a corollary of the following, more abstract, result:

Theorem 2.3. Let $(R_n)_n$ be an irreducible Markov chain on E and $(R_n, \mathbb{X}_n)_n$ a hidden Markov walk on $E \times G^2(V)$ such that

$$\exists K > 0, \forall j \in \mathbb{N}, \|\mathbb{X}_j^{-1} \otimes \mathbb{X}_{j+1}\| < K \text{ a.s.} \quad (13)$$

We denote by $\mathbb{X}^{(1)}$ the first component of \mathbb{X} and set $\beta = \mathbb{E}[T_1]$. Let $\rho = (\rho_N)_N$ be an embedding that can be encoded by the sequence $(\mathbb{X}_n)_n$ (as in proposition 3.3).

Under the conditions $\mathbb{E}[\mathbb{X}_{T_1}^{(1)}] = 0$ and $\mathbb{E}[(\mathbb{X}_{T_1}^{(1)})^{\otimes 2}] = CI_d$, the geodesic embedding of $(\mathbb{X}_k)_{k \leq N}$, renormalized through the dilation $\delta_{(\beta^{-1}NC)^{-1/2}}$, converges in the rough path topology $\mathcal{C}^{\alpha-\text{H\"{o}l}}([0,1], G^2(V))$ with $\alpha < 1/2$ to the rough path given by:

$$\begin{aligned} S_{B,1}(t) &= B_t \\ S_{B,2}(t) &= \int_{0 < s_1 < s_2 < t} \circ dB_{s_1} \otimes \circ dB_{s_2} + \Gamma_\rho t \end{aligned}$$

where $(B_t)_t$ is a d -dimensional Brownian motion and Γ_ρ is a deterministic antisymmetric matrix defined as in (12).

Proof. See section 3.4. Since theorems 2.1 and 2.2 are particular cases of theorem 2.3, we shall not prove them separately. \square

Remarks.

- Setting

$$\mathbb{F}_j = \left(F_j, \iint_{0 < s_1 < s_2 < 1} d\beta_j(s_1) \otimes d\beta_j(s_2) \right)$$

where the β_i s the curves that define ρ as in (11), we have $\mathbb{X}_n = \bigotimes_{j=1}^n \mathbb{F}_j$, and condition (13) becomes similar to (8):

$$\exists K > 0, \forall j \in \mathbb{N}, \|\mathbb{F}_j\| < K \text{ a.s.}$$

- Γ_ρ can be decomposed in two parts: the first one is the Γ that emerges from applying theorem 2.1 to the HMW $(R_n, \mathbb{X}_n^{(1)})_n$, the second one comes from applying the law of large numbers to $(\mathbb{X}_n^{(2)})_n$ (see section 3.4).
- If we endow the space of rough paths with the antisymmetric tensor product \wedge , we can state an antisymmetric version of theorem 2.3.

2.1.2 An easy example: the diamond and round-point models.

Let us now see an example that shows what kind of processes correspond to each of the three theorems.

Consider the irreducible Markov chain $(R_n)_n$ on $E = \{1, 2, \dots, 8\}$ as represented in figure 2. The stochastic matrix Q corresponding to it is given by

$$Q = \begin{pmatrix} 0 & p & 0 & 1-p & 0 & 0 & 0 & 0 \\ p & 0 & 0 & 0 & 0 & 1-p & 0 & 0 \\ 0 & p & 0 & 1-p & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1-p & 0 & p & 0 \\ p & 0 & 0 & 0 & 0 & 1-p & 0 & 0 \\ 0 & 0 & 1-p & 0 & 0 & 0 & 0 & p \\ 0 & 0 & 1-p & 0 & 0 & 0 & 0 & p \\ 0 & 0 & 0 & 0 & 1-p & 0 & p & 0 \end{pmatrix} \quad (14)$$

By associating to each state of E a vector of \mathbb{R}^2 as in the same figure, we define a HMC $(R_n, F_n)_n$, out of which we can eventually construct a HMW $(R_n, X_n)_n$.

If we apply the piecewise linear embedding to $(X_n)_n$, we get the diamond model which is in the left lower corner of figure 2. It is obvious that theorem 2.1 applies to this model.

We can obtain a round-point model by applying to $(X_n)_n$ an embedding of round-arched openings (right lower corner of figure 2). This model can thus be derived from the diamond model by a *change of embedding*.

We will now rewrite both models in the rough path setting, to which theorem 2.3 applies. Since in both cases we construct piecewise smooth embeddings, the rough path resulting from the canonical lift is a geometric rough path, and thus taking values in $G^2(V)$. For commodity reasons, we choose to endow the set

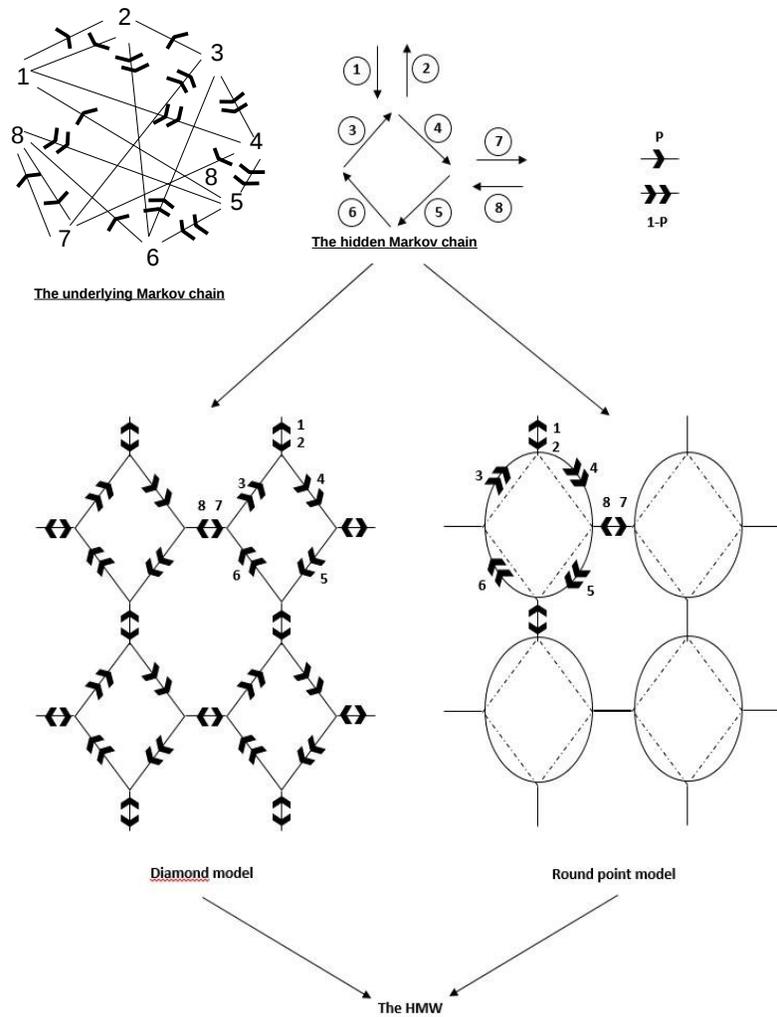


Figure 2: Construction of diamond model and round-point model out of the same HMC $(R_n, F_n)_n$.

$G^2(V)$ with the antisymmetric law \wedge , so that we keep only the antisymmetric part (the signed area) of the second-level component. In the case of the diamond model, the use of the piecewise linear (geodesic) embedding together with the

condition (5) for $G^2(V)$ elements imply that $(\mathbb{X}_n)_n$ is composed as

$$\mathbb{X}_N = \bigwedge_{k=1}^N (F_k, 0)$$

In the case of the round-point model, for $k \geq 1$, each round-arched opening can be encoded in an element F_k and a certain $a_k \in \mathbb{R}$, where the a_k s are the signed areas added by the round-arched openings when passing from the piecewise linear embedding of the diamond model to the new embedding. In particular, the a_k s are the antisymmetric part of the second-level component. The sequence $(\mathbb{X}_n)_n$ is then defined as

$$\mathbb{X}_N = \bigwedge_{k=1}^N (F_k, a_k)$$

$(R_n, (F_n, a_n))_n$ is a hidden Markov chain on $E \times G^2(V)$, with the a_n s given by the signed areas of the corresponding arches of circle.

In order to take the limit, we then construct a rough path by applying the geodesic embedding to $(\mathbb{X}_n)_n$.

2.2 Combinatorial structure behind the iterated sums of HMW.

2.2.1 Iterated sums and non-geometric rough paths.

From iterated integrals to iterated sums. We have already see that iterated integrals of type $S_{\gamma,k}$ from (2) (or algebraic objects satisfying their properties) are what rough paths are composed of.

Let us now pass from continuous objects to discrete ones by analogy. For a sequence $x = (x_n)_n$ in V , we define *the iterated sums* as

$$\tilde{S}_{x,l}(N) = \sum_{1 \leq j_1 < \dots < j_l \leq N} \Delta x_{j_1} \otimes \dots \otimes \Delta x_{j_l} \quad (15)$$

where $\Delta x_i = x_i - x_{i-1}$, and these objects can be interpreted as a discrete analogue of the iterated integrals. We will compare some algebraic properties of iterated sums $\tilde{S}_{x,l}$ with those of iterated integrals $S_{\rho,l}$, where $\rho = (\rho_N)_N$ is an embedding of x , and see how the discrete setting of iterated sums allows to isolate a particular combinatorial structure, *the iterated occupation times*.

Moreover, using the iterated sums $\tilde{S}_{x,l}(N)$, we can construct for the sequence x a discrete analogue of the step- l signature $S_l(\rho_N)$ of ρ_N constructed as in (3) using iterated integrals.

Definition 2.4. *For a sequence $(x_k)_k$ in V , we define the step- l discrete signature as*

$$\tilde{S}_l(x)_{1,N} = (\tilde{S}_{x,1}(N), \dots, \tilde{S}_{x,l}(N)) \in T_1^{(l)}(V) \quad (16)$$

One may wonder what the relation between $(S_l(\rho_N))_N$ and $(\tilde{S}_l(x)_{1,N})_N$ is and what a study on its convergence may reveal. We will partially try to answer these questions hereunder in the setting of HMW.

Non-geometric paths as limits of iterated sums of HMW. Consider a HMW $(R_n, X_n)_n$ on $E \times V$ and set

$$\mathbb{Y}_n = \tilde{S}_2(X)_{1,N} = \left(X_n, \sum_{1 \leq k_1 < k_2 \leq n} F_{k_1} \otimes F_{k_2} \right) \in T_1^{(2)}(V) \quad (17)$$

The following proposition shows that $(\mathbb{Y}_n)_n$ converges in rough path topology to a non-geometric rough path with a second-level deterministic drift.

Proposition 2.1. *Let $(R_n, X_n)_n$ be a HMW on $E \times V$ such that the corresponding F_k s satisfy (8). Choose $(X_t^{(N)})_t$ to be the piecewise linear embedding of $(X_n)_n$, i.e.*

$$\forall N \in \mathbb{N}^*, \forall t \in [0, 1], X_t^{(N)} = \sum_{k=1}^{\lfloor Nt \rfloor} X_k + (Nt - \lfloor Nt \rfloor) F_{\lfloor Nt \rfloor + 1}$$

and suppose $X_0 = 0$ a.s. Consider the geodesic embedding of $Z_k = \sum_{j=1}^k F_j^{\otimes 2}$ given by

$$\forall N \in \mathbb{N}^*, \forall t \in [0, 1], Z_t^{(N)} = \sum_{k=1}^{\lfloor Nt \rfloor} F_k^{\otimes 2} + (Nt - \lfloor Nt \rfloor) F_{\lfloor Nt \rfloor + 1}^{\otimes 2}$$

and set

$$\forall N \in \mathbb{N}^*, \forall t \in [0, 1], \mathbb{Y}_t^{(N)} = (0, Z_t^{(N)})^{-1} \otimes S_2(X^{(N)})_{0,t}$$

Suppose in addition that $\mathbb{E}[X_{T_1}] = 0$, where T_1 is as in (6), and that the covariance matrix $\mathbb{E}[X_{T_1}^{\otimes 2}]$ is equal to CI_d for some $C > 0$. We also set $\beta = \mathbb{E}[T_1]$.

Then $(\mathbb{Y}_t^{(N)})_t$ is a well-defined embedding for $(\mathbb{Y}_n)_n$ and we have the following convergence in the $\mathcal{C}^{\alpha-\text{Hö}}([0, 1], T_1^{(2)}(V))$ topology for $\alpha < 1/2$:

$$\left(\delta_{(NC\beta^{-1})^{-1/2}} \mathbb{Y}_t^{(N)} \right)_{t \in [0,1]} \xrightarrow{N \rightarrow \infty} (\mathbb{B}_t^{\text{It}\hat{o}} \otimes (0, Mt))_{t \in [0,1]}$$

where $(\mathbb{B}_t^{\text{It}\hat{o}})_t$ is the standard Brownian motion on V enhanced with iterated integrals in the Itô sense and M is a deterministic matrix given by

$$M = C^{-1} \mathbb{E} \left[\sum_{1 \leq k_1 < k_2 \leq T_1} F_{k_1} \otimes F_{k_2} \right] \quad (18)$$

In particular, $(\mathbb{B}_t^{\text{It}\hat{o}} \otimes (0, Mt))_t$ is a non-geometric rough path.

Proof. The proof can be found in section 4.2. \square

2.2.2 Iterated occupation times of HMW as underlying combinatorial structures of rough paths.

Definition and combinatorial nature of iterated occupation times. In the case of a HMW $(R_n, X_n)_n$ on $E \times V$ ($(R_n)_n$ is irreducible as before), the conditional measure $\nu(\bullet|u)$ is important for determining the trajectory on V . However, when it comes to algebraic properties, it actually plays only a small role; most of the algebraic properties of rough paths are already present at the level of the discrete Markov chain $(R_n)_n$. They are concentrated in combinatorial objects contained in the HMW, *the iterated occupation times*.

The definition of a HMW implies that

$$\mathbb{E}[X_n|\sigma(R)] = \sum_{u \in E} f(u) \sum_{k=1}^n \mathbf{1}_{R_k=u} \quad (19)$$

where $R = (R_n)_n$ and $f(u) = \mathbb{E}_\nu[X_1 - X_0|u]$. The occupation time of u by $(R_n)_n$ up to time n

$$L_{u;n}(r) = \sum_{k=1}^n \mathbf{1}_{r_k=u} \quad (20)$$

is a natural r.v. at the heart of ergodic-type theorems. We can generalize the notion of occupation time (20) as follows.

Definition 2.5 (iterated occupation times). *Let $(R_n)_n$ be an irreducible Markov chain on E . For any $k \in \mathbb{N}^*$ and any elements $(u_1, \dots, u_k) \in E^k$, the iterated occupation time of (u_1, \dots, u_k) by $(R_n)_n$ at time $n \in \mathbb{N}$ is defined as:*

$$L_{u_1, \dots, u_k; n}(R) = \text{card} \{ (n_1, \dots, n_k) \in \Delta_k(n); R_{n_1} = u_1, \dots, R_{n_k} = u_k \} \quad (21)$$

where $\Delta_k(n)$ is the set

$$\Delta_k(n) = \{ (n_1, \dots, n_k) \in \mathbb{N}^k; 0 \leq n_1 < n_2 < \dots < n_k \leq n \} \quad (22)$$

This cardinal can be written as an iterated sum of products of indicator functions

$$L_{u_1, \dots, u_k; N}(R) = \sum_{1 \leq l_1 < \dots < l_k \leq N} \mathbf{1}_{R_{l_1}=u_1} \dots \mathbf{1}_{R_{l_k}=u_k} \quad (23)$$

Further on, we can easily generalize relation (19) through the following decomposition.

Property 2.1.

$$\begin{aligned} \mathbb{E}[\tilde{S}_{X,l}(N)|\sigma(R)] &= \mathbb{E} \left[\sum_{1 \leq j_1 < \dots < j_l \leq N} \Delta X_{j_1} \otimes \dots \otimes \Delta X_{j_l} \middle| \sigma(R) \right] \\ &= \sum_{(u_1, \dots, u_l) \in E^l} (f(u_1) \otimes \dots \otimes f(u_l)) L_{u_1, \dots, u_l; N}(R) \end{aligned} \quad (24)$$

where $R = (R_n)_n$, $X = (X_n)_n$ and $f(u) = \mathbb{E}_\nu[X_1 - X_0|u]$ as before.

Remark. Formula 24 tells us that, considering under the law $\mathbb{P}(\bullet|\sigma(R))$, the expectations of iterated sums of a HMW are linear combinations of occupation times of the underlying Markov chain. Conversely, an iterated sum of the type

$$\sum_{(u_1, \dots, u_l) \in E^l} v L_{u_1, \dots, u_l; N}(R)$$

with $v \in V^{\otimes l}$ corresponds (under the same probability law) to the expectation of an iterated sum of order l of a certain HMW depending on $(R_n)_n$.

From shuffle to quasi-shuffle products. A particularity of the Stratonovich iterated integrals is that they satisfy *the chain rule*; for example, the product of two components of the first levels satisfies:

$$S_{\gamma,1}(t)_i \otimes S_{\gamma,1}(t)_j = S_{\gamma,2}(t)_{ij} + S_{\gamma,2}(t)_{ji} \quad (25)$$

where x_i is the i -th coordinate of a vector $x \in V$. Formula 25 is actually an illustration of the fact that the multiplication of Stratonovich iterated integrals is a *shuffle product* introduced in [4]. This product can be identified with a particular set of permutations giving all the ways of interlacing two ordered sets while preserving the original order of components of each of them.

In the case of Itô integrals and iterated sums. We have seen that the combinatorial properties of the latter concentrate in the iterated occupation times, which may themselves be understood, through formula (23), as basis vectors of \mathbb{R}^E . This formal identification establishes a relation between iterated integrals/sums of rough paths with the present iterated occupation time. In particular, one checks easily for any elements $u \neq v$

$$L_u(n)L_v(n) = L_{u,v}(n) + L_{v,u}(n) \quad (26)$$

which is very reminiscent of (25). However, for $u = v$, the formula is modified to:

$$L_u(n)L_u(n) = 2L_{u,u}(n) + L_u(n) \quad (27)$$

The emergence of the last term is due to the discrete nature of the sums (in the context of stochastic iterated integrals, it is due to the extra-drift of the Itô integral). This extra-term implies that we need a more general product than the shuffle product to characterize the multiplication of combinatorial structures behind HMW.

Property 2.2. *The product of two iterated sums/occupation times of a HMW can be identified with a quasi-shuffle product, where the quasi-shuffle product is as in definition 4.1 from section (27).*

The quasi-shuffle products, introduced in [8] and extended, for example, in [9], are a generalization of the shuffle product: while the latter supposes that components from two different sets can never coincide when interlacing the sets, the former allows for sets to overlap (following certain rules).

Remark: Given a HMW on $E \times V$ and an embedding for it, the structure given by the shuffle product to the space of the corresponding (Stratonovich) iterated integrals, on one hand, and the one given to the space of the corresponding iterated sums/occupation times by the quasi-shuffle product on the other gives rise to an interesting illustration of the formal algebraic definitions given in [7] for the geometric and non-geometric rough paths.

Asymptotics of the combinatorial structure of HMW. Whereas the convergence of iterated sums of a HMW $(R_n, X_n)_n$ on $E \times V$ can not be directly expressed through the convergence of their combinatorial structure, there is an interesting link between both of them.

Following theorem 2.1 and the classical ergodic theorem, we want to study the large time values and dynamics of $L_{u_1, \dots, u_k; n}(R)$ using the results presented above. We provide two types of estimates: almost sure limits of renormalized iterated occupation times and convergence in law to anomalous Brownian motion in the rough path topology for corrections to the almost sure limits.

First, the iterated occupation times satisfy the following scaling limit.

Proposition 2.2 (almost sure convergence). *Let $(R_n)_{n \in \mathbb{N}}$ be an irreducible Markov chain on a finite set E and $L_{u_1, \dots, u_k; n}(R)$ as in definition 2.5. Then, for any $k \in \mathbb{N}$, the following convergence holds almost surely:*

$$\frac{L_{u_1, \dots, u_k; \lfloor Nt \rfloor}(R)}{N^k} \xrightarrow{N \rightarrow \infty} (1/k!) t^k \pi(u_1) \dots \pi(u_k) \quad (28)$$

where π is the invariant measure for $(R_n)_n$.

Proof. We give only the main idea of the proof: using the theory of pseudo-excursions, we prove that $L_{u_1, \dots, u_k; \lfloor Nt \rfloor}(R)$ can be approximated by a polynomial of degree N^k , and show that the coefficient corresponding to this degree is given by the following a.s. convergence obtained by applying the classical ergodic theorem:

$$N^{-k} \binom{N}{k} L_{u_1}(R) \dots L_{u_k}(R) \xrightarrow{N \rightarrow \infty} (1/k!) \pi(u_1) \dots \pi(u_k)$$

□

Remarks.

- Proposition 2.2 can be seen as a generalized version of the classical ergodic theorem which is also proved using a decomposition into excursions (see, for example, [10]).
- The stochastic matrix Q of $(R_n)_n$ appears here through the invariant measure π .

We now present the results on the convergence in law. We set $L_n(r) = (L_{u;n}(r))_{u \in E}$. In this case, we get the integer-valued vector $L_n(r) \in \mathbb{N}^{\text{Card}(E)}$. We then have the following result which brings into light a universality class of convergence for the HMW:

Proposition 2.3. *Let $(R_n)_n$ be an irreducible Markov chain on the finite state space E . For $n \geq 1$, set $\tilde{L}_n(R) = (L_{u;n}(R) - n\pi(u))_{u \in E} \in \mathbb{R}^{\text{Card}(E)}$, where π is the invariant probability measure of $(R_n)_n$. Then, modulo the appropriate renormalization and a multiplicative constant, the piecewise linear embedding from proposition 4.3 of $(\tilde{L}_n(R))_n$ and the associated iterated integrals $S_{\tilde{L}_n(R),k}(t)$ with $k \geq 2$ converge in rough path topology of $\mathcal{C}^{\alpha-\text{Hö}}([0, 1], T_1^{(2)}(V))$ with $\alpha < 1/2$ towards*

$$S_{B,1}(t) = B_t,$$

$$S_{B,2}(t) = \int_{0 < s_1 < s_2 < t} dB_{s_1} \otimes dB_{s_2} + Mt$$

where $(B_t)_t$ is a $\text{Card}(E)$ -dimensional Brownian motion and M is a deterministic matrix defined as in (18).

Proof. Follows directly from proposition 2.1. □

Remark. Notice that the Brownian motion is not standard here. This comes from the fact that the covariance matrix of an excursion here is not diagonal.

We have the following immediate corollary, which can be viewed as a kind of central limit theorem for iterated sums:

Corollary 2.1. *Let $(R_n)_n$ be an irreducible Markov chain on the finite state space E . For any $u, v \in E$, set*

$$\tilde{L}_{u,v;N}(R) = \sum_{1 \leq l_1 < l_2 \leq N} (\mathbf{1}_{R_{l_1}=u} - \pi(u))(\mathbf{1}_{R_{l_2}=v} - \pi(v))$$

where for $w \in E$ c_w is as in proposition 2.3. Then for any $u, v \in E$, we have the following convergence in law in the uniform topology:

$$N^{-1} \tilde{L}_{u,v;N}(R) \xrightarrow{N \rightarrow \infty} \int_{0 < s_1 < s_2 < 1} dB_{s_1}^{(u)} dB_{s_2}^{(v)} + M_{u,v}$$

where B is a $\text{Card}(E)$ -dimensional Brownian motion and the matrix $M_{u,v}$ is the (u, v) entrance of the $\text{Card}(E) \times \text{Card}(E)$ matrix M from proposition 2.3.

Remark. We can generalize the result of corollary 2.1 for iteration times of higher order by setting, for any $u_1, \dots, u_k \in E$,

$$\tilde{L}_{u_1, \dots, u_k; N}(R) = \sum_{1 \leq l_1 < \dots < l_k \leq N} (\mathbf{1}_{R_{l_1}=u_1} - \pi(u_1)) \dots (\mathbf{1}_{R_{l_k}=u_k} - \pi(u_k))$$

However, in this case the additional terms at the limit (all the terms aside the iterated integral of Brownian motion) are more difficult to express explicitly, even if they depend on M and $(B_t)_t$.

3 From Hidden Markov Walks to rough paths

3.1 Theory of pseudo-excursions for hidden Markov walks.

Another definition of HMW. Definitions 2.1 and 2.2 can be summed up in the following definition-property, which includes also the HMW taking values in $E \times G^2(V)$, where $G^2(V)$ is as defined in notations of section 1.2 or in definition 3.3.

Property 3.1. *A process $(R_n, X_n)_n$ is a hidden Markov walk on $E \times V$ (resp. on $E \times G^2(V)$) in the sense of definition 2.2 if and only if there exists a sequence of r.v. $(F_n)_n$ with values in V (resp. $G^2(V)$) such that*

- under $\mathbb{P}(\bullet | \sigma(R))$, with $R = (R_n)_n$, $(F_n)_n$ is a sequence of independent r.v.;
- the distribution of F_n knowing $R_n = u$ is $\nu(\bullet | u)$ for all $n \in \mathbb{N}^*$;
- $\forall n \in \mathbb{N}^*$, $X_n = \sum_{k=1}^n F_k$ (resp. $X_n = \bigotimes_{k=1}^n F_k$ in $G^2(V)$).

In particular, we have $F_k = X_k - X_{k-1}$ (resp. $F_k = X_{k-1}^{-1} \otimes X_k$ in $G^2(V)$).

Theory of pseudo-excursions. As it has already been mentioned in the introduction, HMW are a way of generalizing simple random walks: instead of being i.i.d. variables, the increments are variables depending on a Markov chain on a finite state space E : to each state $u \in E$ we associate a supplementary object (a vector, a curve etc.) which gives the increment. However, we can derive a sum of i.i.d.r.v. from a HMW using *the theory of pseudo-excursions*. For commodity reasons, we state it for HMW on $E \times V$, but it is equally valid for those on $E \times G^2(V)$.

If $(R_n)_n$ is an irreducible Markov chain on a finite state space E , we can apply to it all the excursion theory we have for Markov chains. Thus, even if the hidden Markov walk $(R_n, X_n)_n$ may not even be a Markov chain, we can construct a theory of pseudo-excursions for it based on the results we have for $(R_n)_n$.

We start by defining a sequence of stopping times for the Markov chain $(R_n)_n$ as in (6). We then have the following definition, which is a property at the same time.

Definition 3.1 (pseudo-excursions). *Let $(R_n, X_n)_n$ be a HMW on $E \times V$ and the sequence $(T_k)_k$ as defined in (6). We call $(X_{T_k+1}, \dots, X_{T_{k+1}})$ the k -th pseudo-excursion of $(X_n)_n$ and $X_{T_{k+1}} - X_{T_k}$ the k -th pseudo-excursion increment.*

Property 3.2. *We have the following basic properties for pseudo-excursions:*

- The variables $X_{T_{k+1}} - X_{T_k}$ are i.i.d. random variables.
- For $i \neq j$, the trajectories $(X_{T_i+1}, \dots, X_{T_{i+1}})$ and $(X_{T_j+1}, \dots, X_{T_{j+1}})$ are independent and have the same law.

Proof. Let $k \neq m$. The pseudo-excursions $X_{T_{k+1}} - X_{T_k}$ and $X_{T_{m+1}} - X_{T_m}$ are independent knowing $(R_{T_{k+1}}, \dots, R_{T_{k+1}})$ and $(R_{T_{m+1}}, \dots, R_{T_{m+1}})$ respectively (this follows from property 3.1). Since $k \neq m$, the excursions $(R_{T_{k+1}}, \dots, R_{T_{k+1}})$ and $(R_{T_{m+1}}, \dots, R_{T_{m+1}})$ are i.i.d., and thus $X_{T_{k+1}} - X_{T_k}$ and $X_{T_{m+1}} - X_{T_m}$ are independent and have the same law. \square

Remark. Let us briefly justify the choice of the term "pseudo-excursions". The excursion part comes from the fact that $X_{T_{k+1}} - X_{T_k}$ depends on the excursion $(R_{T_{k+1}}, \dots, R_{T_{k+1}})$. Why "pseudo"? Because an ordinary excursion is the trajectory of a process before returning to its starting point, and the increments $X_{T_{k+1}} - X_{T_k}$ are non-trivial, i.e. we do not return to our starting point.

The pseudo-excursions are particularly interesting when it comes to convergence of a hidden Markov walk, as they tell us that it can be studied almost as the convergence of a sum of well-chosen i.i.d.r.v.

Proposition 3.1. *Let $(R_n, X_n)_n$ be a hidden Markov walk as in definition 2.2. Suppose that there exists $M > 0$ such that $|X_{n+1} - X_n| \leq M$ a.s. There exists a sequence $(V_n)_n$ of i.i.d.r.v. such that, setting, for all $n \geq 1$,*

$$S_n = \sum_{k=1}^{\lfloor n\mathbb{E}[T_1]^{-1} \rfloor} V_k$$

we have

$$\forall \epsilon > 0, \mathbb{P}(|X_n - S_n| > \epsilon) \rightarrow 0$$

Proof. Let $\kappa(n)$ be an integer such that $T_{\kappa(n)} \leq n < T_{\kappa(n)+1}$. We decompose

$$X_n = \sum_{i=1}^{\kappa(n)} (X_{T_i} - X_{T_{i-1}}) + \sum_{i=T_{\kappa(n)+1}}^n (X_i - X_{i-1})$$

The first sum contains $\kappa(n)$ i.i.d. variables, and moreover $\kappa(n)/n \xrightarrow[n \rightarrow \infty]{} \mathbb{E}[T_1]^{-1}$ a.s. by the ergodic theorem.

We will now prove that the second sum converges to zero in probability. For $\epsilon > 0$, we have:

$$\begin{aligned} \mathbb{P} \left(\left| \sum_{i=T_{\kappa(n)+1}}^n X_i - X_{i-1} \right| > \epsilon \right) &\leq \mathbb{P} \left(\sum_{i=T_{\kappa(n)+1}}^{T_{\kappa(n)+1}} |X_i - X_{i-1}| > \epsilon \right) \\ &\leq \mathbb{P} (M(T_{\kappa(n)+1} - T_{\kappa(n)}) > \epsilon) \\ &\leq M\epsilon^{-1} \mathbb{E}[T_1] \rightarrow 0 \end{aligned}$$

We then obtain the desired result by applying Slutsky's theorem (for reference, see [1], theorem 3.1). \square

Proposition 3.1 is very useful in the sense that it allows to adopt the convergence theorems we have for sums of i.i.d. random variables to the class of hidden Markov walks.

3.2 Elements of rough paths theory

The reader who is well familiar with rough paths may skip this part, and those who would like a more detailed study may refer, for example, to [6] or to [5].

Some definitions. In general, a rough path is defined as a continuous path on $T_1^{(N)}(V)$ for $N \geq 1$. When we want to associate a rough path to a V -valued path γ of regularity $\alpha < 1$, we choose $N = \lfloor 1/\alpha \rfloor$ and fill each level following some rules.

Remark. The way of lifting a path in V to a rough path is not unique.

In particular, the regularity of the rough path has to be coherent with the regularity of the initial path (for details on the regularity issues, see, for example, [14] or [6]). Throughout this article, we are interested in rough paths that convergence towards the Brownian motion whose regularity is $1/2^-$, so the rough paths we operate correspond to the following definition.

Definition 3.2. Let $1/3 < \alpha < 1/2$. An α -Hölder rough path $(\mathbf{x}_t)_t$ is an element of $\mathcal{C}^{\alpha-\text{Hö}}([0, 1], T_1^{(2)}(V))$. Moreover, if $(\mathbf{x}_t)_t$ is an element of $\mathcal{C}^{\alpha-\text{Hö}}([0, 1], G^2(V))$, it is an α -Hölder geometric rough path.

A more informal way to state definition 3.2 is the following:

- a geometric rough path on $G^2(V)$ endowed with the law \otimes inherited from $T_1^{(2)}(V)$: $(a, b) \otimes (a', b') = (a + a', b + b' + a \otimes a')$;
- a geometric rough path on $G^2(V)$ endowed with the antisymmetric law \wedge , which eliminates the symmetric part of the second level (it is seen as redundant since it depends entirely on the first level): $(a, b) \wedge (a', b') = (a + a', b + b' + \frac{1}{2}a \otimes a' - a' \otimes a)$;
- a non-geometric rough path on $T_1^{(2)}(V)$ endowed with the law \otimes (we do not consider the antisymmetric law \wedge since the symmetric part of the second level contains elements that do not depend on the first level).

An important property of rough paths is that they satisfy *Chen's relation*, i.e.

$$\forall 0 \leq s < u < t \leq 1, \mathbf{x}_{s,u} \otimes \mathbf{x}_{u,t} = \mathbf{x}_{s,t} \quad (29)$$

or, alternatively, in the antisymmetric setting (for $G^2(V)$),

$$\forall 0 \leq s < u < t \leq 1, \mathbf{x}_{s,u} \wedge \mathbf{x}_{u,t} = \mathbf{x}_{s,t} \quad (30)$$

where $\forall 0 \leq s < u$, $\mathbf{x}_{s,u} = \mathbf{x}_u^{-1} \otimes \mathbf{x}_s$. This reflects, in particular, the fact that \mathbf{x}_t can be decomposed into a "sum" of increments of \mathbf{x} .

The Brownian motion rough path is the most important practical example of rough path in the probabilistic setting. Since stochastic calculus allows us to

define integrals with respect to the Brownian motion (Itô or Stratonovich), we can directly construct the rough path as:

$$\mathbb{B}_{s,t} = (B_t - B_s, \int_{s < u < v < t} \circ dB_u \otimes \circ dB_v) \quad (31)$$

In the antisymmetric setting of $G^2(V)$, we only keep the antisymmetric part of (31) and we obtain *the enhanced Brownian motion*:

$$\mathbb{B}_{s,t} = (B_t - B_s, \mathcal{A}_{s,t}) \quad (32)$$

where \mathcal{A} is the stochastic signed area of B , called *the Lévy area*. The enhanced Brownian motion is a Brownian motion on $G^2(V)$.

The group $G^2(V)$: alternative definition and topology. Since our main theorems of convergence (2.1, 2.2, 2.3) are dealing with geometric rough paths, we give more details about the group $G^2(V)$.

In the introduction, we have stated that $G^2(V)$ is a subgroup of $T_1^{(2)}(V)$ whose elements satisfy condition (5) (i.e. the symmetrical part of the second level depends entirely on the first level). We will now give an alternative, more analytical definition.

Definition 3.3. *The free nilpotent group $G^2(V)$ is defined as:*

$$G^2(V) = \{S_2(\gamma)_{0,1} : \gamma \in \mathcal{C}^{1-var}([0,1], V)\}$$

where $S_2(\gamma)_{0,1}$ is as in (3).

The topology of $G^2(V)$ is induced by the *Carnot-Caratheodory norm*, which gives the length of the shortest path corresponding to a given signature:

$$\forall g \in G^2(V), \|g\| := \inf \left\{ \int_0^1 |dx| : x \in \mathcal{C}^{1-var}([0,1], V) \text{ and } S_2(\gamma)_{0,1} = g \right\} \quad (33)$$

where $|\cdot|_V$ is a restriction to V of the Euclidean norm. The norm thus defined is homogeneous ($\|\delta_\lambda g\| = |\lambda| \|g\|$ for $\lambda \in \mathbb{R}$), symmetric ($\|g\| = \|g^{-1}\|$) and sub-additive ($\|g \otimes h\| \leq \|g\| + \|h\|$), it induces a continuous metric \mathbf{d} on $G^2(V)$ through the application

$$\begin{aligned} \mathbf{d} : G^2(V) \times G^2(V) &\rightarrow \mathbb{R}_+ \\ (g, h) &\mapsto \|g^{-1} \otimes h\| \end{aligned} \quad (34)$$

In this case, $(G^2(V), \mathbf{d})$ is a geodesic space (in the sense of definition 5.19 from [6]) and a Polish space.

Another useful norm on $G^2(V)$, homogeneous but neither symmetric nor sub-additive, is given by:

$$\forall (a, b) \in G^2(V), \|(a, b)\| = \max\{|a|_V, |b|_{V \otimes V}^{1/2}\}$$

Since all homogeneous norms are equivalent on $G^2(V)$, this new norm gives us a rather easy way to get an estimate of the Carnot-Caratheodory norm:

Property 3.3. *There exist two positive constants c_1 and c_2 such that*

$$\forall (a, b) \in G^2(V), c_1 \| (a, b) \| \leq \| (a, b) \| \leq c_2 \| (a, b) \| \quad (35)$$

A convergence criterion for rough paths. We usually prove the convergence of a sequence of rough paths using pointwise convergence plus a tightness criterion. However, for some cases, we can also use the following result, deduced from exercise 2.9 in [5]:

Proposition 3.2. *Consider a sequence of rough paths $(\mathbb{X}^{(N)})_N$ which takes values in $\mathcal{C}^{\alpha-\text{Hö}}([0, 1], G^2(V))$ with $\alpha < 1/2$ such that we have the uniform bound*

$$\sup_N \mathbb{E} \left[\sup_{s \neq t} \frac{\| \mathbb{X}_{s,t}^{(N)} \|}{|t - s|^{1/2}} \right] < \infty$$

and the pointwise convergence (in probability)

$$\forall t \in [0, 1], \mathbb{X}_{0,t}^{(N)} \xrightarrow[N \rightarrow \infty]{} \mathbb{X}_{0,t}$$

for a certain rough path \mathbb{X} in $\mathcal{C}^{\alpha-\text{Hö}}([0, 1], G^2(V))$.

Then $\mathbb{X}^{(N)}$ converges in probability to \mathbb{X} in the $\mathcal{C}^{\alpha-\text{Hö}}([0, 1], G^2(V))$ topology with $\alpha < 1/2$.

3.3 Embeddings

Definition and equivalence classes in the case of a finite-dimensional vector space. In order to study the limit of a discrete process, we need to properly define this convergence in the continuous space, where the limit process lives. This is when embeddings come on the scene.

In the introduction, we have given a general definition 2.3 of an embedding and a way of constructing an embedding for a given sequence of points (11). The following property highlights the fact that the method is consistent with the definition:

Definition 3.4. *Let $(x_n)_n$ be a sequence with values in a finite-dimensional vector space V . Consider a sequence $(\beta_n)_n$ of continuous paths $\beta_n : [0, 1] \rightarrow V$ and such that $\beta_n(0) = x_n$ and $\beta_n(1) = x_{n+1}$. Let \cdot be the operator of path concatenation given by*

$$\begin{aligned} \forall \gamma : [s, t] \rightarrow V, \forall \gamma' : [s', t'] \rightarrow V, \forall u \in [s, t + t' - s'], \\ (\gamma \cdot \gamma')(u) = \begin{cases} \gamma(u) & \text{if } u \in [s, t] \\ \gamma'(u - t + s') & \text{if } u \in [t, t + t' - s'] \end{cases} \quad (36) \end{aligned}$$

Then the sequence $\rho_N = \beta_1 \cdot \dots \cdot \beta_N$ is an embedding of $(x_n)_n$.

The most common embedding is *the geodesic embedding*, i.e. the one that connects two consecutive points (in a geodesic space) by a curve minimizing the distance between them (in \mathbb{R}^d , it consists in simply connecting consecutive points of $(x_n)_n$ by straight lines). The advantage of this term is that it can be applied to embeddings in V as well as in $G^2(V)$.

In some cases, we can study the convergence of the embedding thus obtained by shrinking the space scale of the paths, i.e. by *rescaling the embedding*. We then have the following definition:

Definition 3.5. *Let V be a finite-dimensional vector space. Consider a V -valued embedding $\rho_N = \beta_1 \cdot \dots \cdot \beta_N$ as defined in 3.4. Let $f : \mathbb{N} \rightarrow \mathbb{R}_+$ be an increasing function. The sequence $\iota_N = f(N)\rho_N$ defines a rescaled embedding.*

Even if, for a given discrete process, we have a large choice of embeddings, we should be aware of the fact that the limit we get depends of the embedding we adopt (some embeddings may not even admit a limit). For a sequence $(x_n)_n$ with values in a vector space V , we can define a way of dividing the embeddings into equivalence classes by *building rough paths out of our embeddings* (when possible) and considering the convergence in the rough path topology:

Definition 3.6. *Let $(x_n)_n$ be a sequence with values in a (finite-dimensional) vector space V . We say that two embeddings $\rho = (\rho_N)_N$ and $\rho' = (\rho'_N)_N$ of $(x_n)_n$ are equivalent (in the rough path sense) if there exists a function $f : \mathbb{N} \rightarrow \mathbb{R}_+$ such that the rough paths $\mathbf{\iota}_N$ and $\mathbf{\iota}'_N$ corresponding to $\iota_N = f(N)\rho_N$ and $\iota'_N = f(N)\rho'_N$ respectively converge in distribution to the same limit \mathbf{l} in the rough path topology.*

We shall now see how this definition is important for generalizing the convergence of HMW in rough path topology.

From sequences in $G^2(V)$ to rough paths. For $n \in \mathbb{N}$, we consider a sequence $(g_k)_k = (g_k^{(1)}, g_k^{(2)})_k$ in $G^2(V)$. As in the case of a sequence in V , we can pass to the continuous framework by associating a set of embeddings to $(g_k)_k$ as follows:

Definition 3.7. *Consider a sequence $(g_k)_k \in G^2(V)^\mathbb{N}$. Consider a sequence of rough paths $(\rho_N)_N$ such that $\rho_N : [0, N] \rightarrow G^2(V)$ which satisfy*

$$\forall 1 \leq k \leq N, \quad \rho_N(k-1)^{-1} \otimes \rho_N(k) = g_k$$

We say that $(\rho_N)_N$ is an embedding for $(g_k)_k$.

Just like in the case of a vector space V , a particularly important embedding is the geodesic embedding, which, in this case, consists in connecting the elements of a sequence by geodesic curves minimizing the distance between them:

Definition 3.8 (the geodesic embedding in $G^2(V)$). *Consider a sequence $(g_k)_k \in G^2(V)^\mathbb{N}$. We call geodesic embedding of $(g_k)_k$ the sequence of rough paths*

$$\begin{aligned} \rho_N : [0, N] &\rightarrow G^2(V) \\ t &\mapsto \bigotimes_{k=1}^{\lfloor t \rfloor} g_k \otimes \delta_{t-\lfloor t \rfloor}(g_{\lfloor t \rfloor+1}) \end{aligned} \tag{37}$$

The universality of the geodesic embedding in $G^2(V)$ for HMW. We can now explain what we mean by the universality of the geodesic embedding. The geodesic embedding on $G^2(V)$ allows us to describe further beyond the equivalence classes the convergence of different embeddings of a hidden Markov walk. More precisely, the limit of all the embeddings that are equivalent in the rough path sense is the same as the limit of the geodesic embedding of a well-chosen sequence in $G^2(V)$:

Proposition 3.3. *Consider a process $(X_n)_n$ on V and suppose that there exists $(\rho_N)_N$ an embedding of $(X_n)_n$ with bounded variation (constructed, for example, as in lemma 2.1) and such that $(N^{-1/2}\rho_N(Nt))_{t \in [0,1]}$ converges in law in $\mathcal{C}^{\alpha-\text{Hö}}([0,1], G^2(V))$ topology for $\alpha < 1/2$. Then there exists a sequence $g_n = (g_n^{(1)}, g_n^{(2)})_n \in G^2(V)$ such that its geodesic embedding rescaled by the operator $\delta_{N^{-1/2}}$ converges to the same limit in $\mathcal{C}^{\alpha-\text{Hö}}([0,1], G^2(V))$ topology as any embedding equivalent to ρ_N in the sense of definition 3.6.*

Proof. Denote by $\boldsymbol{\rho}_N$ the geometric rough path corresponding to ρ_N . Set

$$\forall k \geq 1, g_k = \boldsymbol{\rho}_k(k-1)^{-1} \otimes \boldsymbol{\rho}_k(k)$$

By definition of the embedding 3.4, this means that

$$\forall k \geq 1, \forall N \geq k, g_k = \boldsymbol{\rho}_N(k-1)^{-1} \otimes \boldsymbol{\rho}_N(k)$$

The geodesic embedding for the sequence $(g_k)_k$ is given by

$$\begin{aligned} \forall t \in [0,1], \boldsymbol{\varrho}_N(Nt) &= \bigotimes_{k=1}^{\lfloor Nt \rfloor} g_k \otimes \delta_{Nt - \lfloor Nt \rfloor}(g_{\lfloor Nt \rfloor + 1}) \\ &= \boldsymbol{\rho}_N(\lfloor Nt \rfloor) \otimes \delta_{Nt - \lfloor Nt \rfloor}(\boldsymbol{\rho}_N(\lfloor Nt \rfloor)^{-1} \otimes \boldsymbol{\rho}_N(\lfloor Nt \rfloor + 1)) \end{aligned}$$

it is now left to prove that $(\delta_{N^{-1/2}}\boldsymbol{\varrho}_N(Nt))_{t \in [0,1]}$ and $(\delta_{N^{-1/2}}\boldsymbol{\rho}_N(Nt))_{t \in [0,1]}$ converge towards the same limit in rough path topology.

As before, we denote by $\|\cdot\|$ the Carnot-Caratheodory norm defined in (33). Since ρ_N is α -Hölder, the equivalence of norms (35) tells us that there exists $c > 0$ such that

$$\forall n \geq 1, \forall 0 \leq u < v \leq N, \|\boldsymbol{\rho}_N(u)^{-1} \otimes \boldsymbol{\rho}_N(v)\| \leq c|v - u|^\alpha$$

Consequently, we have

$$\begin{aligned} &\sup_{t \in [0,1]} \|\delta_{N^{-1/2}}\boldsymbol{\rho}_N(Nt)^{-1} \otimes \delta_{N^{-1/2}}\boldsymbol{\varrho}_N(Nt)\| \\ &\leq N^{-1/2} \left(\sup_{t \in [0,1]} \|\boldsymbol{\rho}_N(Nt)^{-1} \otimes \boldsymbol{\rho}_N(\lfloor Nt \rfloor)\| \right. \\ &\quad \left. + \sup_{t \in [0,1]} \|\delta_{Nt - \lfloor Nt \rfloor}(\boldsymbol{\rho}_N(\lfloor Nt \rfloor)^{-1} \otimes \boldsymbol{\rho}_N(\lfloor Nt \rfloor + 1))\| \right) \\ &\leq N^{-1/2} \left(\sup_{t \in [0,1]} (Nt - \lfloor Nt \rfloor)^\alpha + \sup_{t \in [0,1]} (Nt - \lfloor Nt \rfloor) \right) \xrightarrow{N \rightarrow \infty} 0 \end{aligned}$$

which achieves the proof. \square

Proposition 3.3 shows the interest of theorem 2.3: rather than operating with concatenation of continuous paths in each particular case, we can manipulate a more general object in $G^2(V)$. The advantage is a general result for an equivalence class and a more convenient setting for computations.

Embeddings for differential and difference equations. As it has already been mentioned, given a differential equation $dY_t = f(Y_t)dX_t$, its solution map $\xi(Y_0, X) = Y$ is not always continuous in uniform topology, which means that we have to choose an appropriate sequence of continuous paths $(X^{(n)})_n$ approaching X in order for $\xi(Y_0, X^{(n)}) = Y^{(n)}$ to suitably approach Y . The rough paths theory gives us a universal recipe for making this choice: a suitable sequence $(X^{(n)})_n$ is such that the corresponding sequence of rough paths $(\mathbb{X}^{(n)})_n$ converges to the rough path \mathbb{Y} of Y in the rough path topology.

Even if the rough paths theory deals only with continuous objects, the above statement has surprising and important consequences in the discrete setting. A crucial difference is that, in the continuous case, the limit is given, and we have to find a suitable sequence that convergence towards it. On the contrary, if we have a sequence of difference equations, the limit is not known a priori and thus will be function of the embedding. Otherwise said, *in the first case we have a given limit and several possible approximations, whereas in the second one we have a given discretization and several possible limits.*

Let us now see how this fact applies to the analysis of the discrete analogue of differential equations, *the difference equations*. While solving these equations directly can have useful applications, it is also of great interest to consider their convergence after renormalization, i.e. the convergence of

$$\Delta Y_n = \epsilon \sum_{i=1}^d f^{(i)}(Y_{n-1})(\Delta X_n)^{(i)} \quad (38)$$

such that $\epsilon \rightarrow 0$. This is where that the embedding comes on the scene.

Property 3.4. *Consider a difference equation as in (38) and suppose that $(X_n)_n$ allows two different embeddings of finite variation ρ_N and ρ'_N . Then, if ρ_N and ρ'_N are equivalent in the rough path sense, they define the same limit equation in the rough path topology (if it exists).*

In particular, it means that, in the case of rough paths in $G^2(V)$, the difference between the corresponding limit equations will depend on the second-level component of the rough path, which contains *the area anomaly and the square bracket of the process*. Let us now see how this translates in the setting of the classical stochastic calculus in the case where $(X_n)_n$ is a hidden Markov walk.

In our article [13], we have discussed the convergence of a difference equation driven by the sum of the Bernoulli "turning" variables. In that example, we have implicitly supposed a piecewise linear, geodesic embedding. What happens if

we change it? The additional term we obtain at the limit by choosing a hidden Markov walk as driving process is a drift consisting of two components. The first one, featuring the constant K , depends on the square bracket of the process, i.e. is a common term in stochastic calculus. The second one, containing the area anomaly Γ , is a new term that can only be brought up by means of rough path analysis. However, *both of them depend on the embedding*: the first one due to the construction of the square bracket as a limit, the second one due to the fact that different embeddings may generate different stochastic area anomalies at the limit, as it has been mentioned earlier. Thus, the choice of an embedding is a problem that has consequences on the classical stochastic calculus even independently of the rough paths theory.

3.4 Getting back to theorem 2.3

Notations: For $g \in G^2(V)$, we denote by $g^{(1)}$ its first level component and by $g^{(2)}$ the second-level one. We denote by δ is the standard dilatation operator on $G^2(V)$ (i.e. $\delta_\epsilon(g^{(1)}, g^{(2)}) = (\epsilon g^{(1)}, \epsilon^2 g^{(2)})$).

We start with two preliminary lemmas.

Lemma 3.1. *Let $(\xi_n)_n$ be a sequence of i.i.d. $G^2(V)$ -valued centred random variables with bounded moments of all orders, i.e.*

$$\forall p \geq 1, \mathbb{E} \left[|\xi_n^{(1)}|^p \right] < \infty$$

Furthermore, let $(k_n)_n$ be a sequence of \mathbb{N} -valued r.v. such that

- $k_0 = 0$ a.s.
- $\mathbb{P}(\forall n \geq 0, k_{n+1} \in \{k_n, k_n + 1\}) = 1$
- $\frac{k_n}{n} \xrightarrow{n \rightarrow \infty} a \in \mathbb{R}_+^*$ a.s.

Set $\Xi_n = \bigotimes_{k=1}^{k_n} \xi_k$ and for $t \in [0, 1]$, $\Xi_t^n = \Xi_{[nt]} \otimes \delta_{nt - [nt]}(\xi_{k_{[nt]+1}})$. We then have the following convergence:

$$\left(\delta_{(n\sigma^2)^{-1/2}}(\Xi_t^n) \right)_{t \in [0, 1]} \xrightarrow[n \rightarrow \infty]{(d)} (\mathbb{B}_t)_{t \in [0, 1]}$$

where $(\mathbb{B}_t)_{t \in [0, 1]}$ is the enhanced Brownian motion, and $\sigma^2 = \mathbb{E} \left[|\xi_0^{(1)}|^2 \right]$.

Remark. The difference with the Donsker-type theorem from [2] is that the sums of i.i.d. r.v. are a function of k_n and not simply of n .

Proof. We will proceed by the classical method, proving first the convergence of the finite-dimensional marginals of the process and then its tightness.

For the first part, we use the result from in [2], which states that a (renormalized) sum of i.i.d.r.v. in $G^2(V)$ converges in law to the Brownian rough path.

Since k_n/n converges a.s. and that, moreover, $(k_n)_n$ is increasing a.s., a sum of k_n i.i.d.r.v. yields the same type of convergence, but with a change in time depending on a . Otherwise said, for $t \in [0, 1]$, we have

$$\delta_{(\sigma^2 n)^{-1/2}} \Xi_t^n = \delta_{(k_n/n)^{1/2}} \delta_{(\sigma^2 k_n)^{-1/2}} \Xi_t^n \xrightarrow[n \rightarrow \infty]{(d)} \sqrt{a} \mathbb{B}_t$$

and we can then generalize this convergence to any finite-dimensional marginals using the independence of the variables and Slutsky's theorem (theorem 3.1 in [1]) as in the classical Donsker theorem.

In order to prove the tightness of the process in $\mathcal{C}^\alpha([0, 1], G^2(V))$ for $\alpha < 1/2$, we can use the Kolmogorov's criterion (as already exposed in the proof of theorem 1.1 from [13]), i.e. we have to prove that, for any $p > 1$, there exists $c > 0$ such that

$$\mathbb{E} \left[n^{-2p} \mathbf{d}(\Xi_t^n, \Xi_s^n)^{4p} \right] \leq c |t - s|^{2p}$$

for any $s, t \in [0, 1]$. The definition of $\mathbf{d}(\cdot, \cdot)$ implies that

$$\mathbf{d}(\Xi_t^n, \Xi_s^n) = \left\| \delta_{ns - \lfloor ns \rfloor} \xi_{k_{\lfloor ns \rfloor + 1}}^{-1} \otimes \Xi_{\lfloor ns \rfloor}^{-1} \otimes \Xi_{\lfloor nt \rfloor} \otimes \delta_{nt - \lfloor nt \rfloor} \xi_{k_{\lfloor nt \rfloor + 1}} \right\| \quad (39)$$

If $s, t \in [k/n, (k+1)/n[$, then this expression becomes

$$\mathbf{d}(\Xi_t^n, \Xi_s^n) = n(t - s) \left\| \xi_{k_{\lfloor ns \rfloor + 1}} \right\|$$

which further on gives us

$$\mathbb{E} \left[n^{-2p} \mathbf{d}(\Xi_t^n, \Xi_s^n)^{4p} \right] \leq M_p n^{2p} |t - s|^{4p} \leq M_p |t - s|^{2p}$$

where M_p is such that $\mathbb{E} \left[\|\xi_0\|^{4p} \right] \leq M_p$.

If $s \in [i/n, (i+1)/n[$ and $t \in [(i+1)/n, (i+2)/n[$, the expression (39) becomes:

$$\begin{aligned} \mathbf{d}(\Xi_t^n, \Xi_s^n) &= \left\| \delta_{ns - \lfloor ns \rfloor} \xi_{k_{\lfloor ns \rfloor + 1}}^{-1} \otimes \Xi_{\lfloor ns \rfloor + 1} \otimes \delta_{nt - \lfloor nt \rfloor} \xi_{k_{\lfloor nt \rfloor + 1}} \right\| \\ &= \left\| \delta_{1 - (ns - \lfloor ns \rfloor)} \xi_{k_{\lfloor ns \rfloor + 1}} \otimes \delta_{nt - \lfloor nt \rfloor} \xi_{k_{\lfloor nt \rfloor + 1}} \right\| \\ &= (\lfloor nt \rfloor - ns) \left\| \xi_{k_{\lfloor ns \rfloor + 1}} \right\| + (nt - \lfloor nt \rfloor) \left\| \xi_{k_{\lfloor nt \rfloor + 1}} \right\| \\ &\leq 2n(t - s) \left(\left\| \xi_{k_{\lfloor ns \rfloor + 1}} \right\| \right) \end{aligned}$$

and we can conclude as in the first case.

Finally, for $s \in [i/n, (i+1)/n[$ and $t \in [(i+l)/n, (i+l+1)/n[$ with $l \geq 2$, we use the properties of $\mathbf{d}(\cdot, \cdot)$ to get

$$\mathbf{d}(\Xi_t^n, \Xi_s^n)^{4p} \leq 2^{4p} \left(\left\| \delta_{1 - (ns - \lfloor ns \rfloor)} \xi_{k_{\lfloor ns \rfloor + 1}} \right\|^{4p} + \left\| \bigotimes_{j=k_{\lfloor ns \rfloor + 2}}^{k_{\lfloor nt \rfloor}} \xi_j \right\|^{4p} + \left\| \delta_{nt - \lfloor nt \rfloor} \xi_{k_{\lfloor nt \rfloor + 1}} \right\|^{4p} \right)$$

and consequently

$$\mathbb{E} \left[n^{-2p} \mathbf{d}(\Xi_t^n, \Xi_s^n)^{4p} \right] \leq 2^{4p} \left(2\mathbb{E} \left[n^{-2p} \|\xi_0\|^{4p} \right] + \mathbb{E} \left[\left\| n^{-2p} \bigotimes_{j=k_{[ns]}+2}^{k_{[nt]}} \xi_j \right\|^{4p} \right] \right)$$

The first part of the right-hand side inequality can be bounded using the fact that there exists $M_p > 0$ such that $\mathbb{E} \left[n^{-2p} \|\xi_0\|^{4p} \right] \leq M_p n^{-2p} \leq M_p |t-s|^{2p}$, as $|t-s| > 1/n$. We transform the second one using the independence of variables, and then what is left to prove is that there exists $c' > 0$ such that

$$\begin{aligned} \mathbb{E} \left[n^{-2p} \left\| \bigotimes_{j=k_{[ns]}+2}^{k_{[nt]}} \xi_j \right\|^{4p} \right] &= \mathbb{E} \left[n^{-2p} \left\| \bigotimes_{j=1}^{k_{[nt]}-k_{[ns]}-1} \xi_j \right\|^{4p} \right] \\ &\leq c' |t-s|^{2p} \end{aligned}$$

We use the result $\mathbb{E} \left[\left\| \bigotimes_{k=1}^m \xi_k \right\|^{4p} \right] = O(m^{2p})$ proven in [2] and the inequality $k_{j+l} - k_j \leq l$ to get:

$$\begin{aligned} \mathbb{E} \left[\left\| \bigotimes_{j=1}^{k_{[nt]}-k_{[ns]}-1} \xi_j \right\|^{4p} \right] &= \mathbb{E} \left[\left\| \bigotimes_{j=1}^{k_{i+l}-k_i-1} \xi_j \right\|^{4p} \right] \\ &= O((k_{i+l} - k_i)^{2p}) = O((l-1)^{2p}) = O((n|t-s|)^{2p}) \end{aligned}$$

We thus get a uniform bound, which achieves the proof. \square

Lemma 3.2. Consider $(k_n)_n$ a sequence of r.v. as the one in lemma 3.1. Let $((0, C_n))_n$ be a sequence of uniformly bounded r.v. taking values in the centre of the group $G^2(V)$ and such that we have the following a.s. convergence for any $t \in [0, 1]$:

$$\delta_{n^{-1/2}} \bigotimes_{i=1}^{k_{[nt]}} (0, C_i) \xrightarrow[n \rightarrow \infty]{} (0, atM) \quad (40)$$

where M is a deterministic matrix and $a \in \mathbb{R}_+^*$ is as in lemma 3.1. Then we have the following convergence in probability

$$\left(\delta_{n^{-1/2}} \bigotimes_{i=1}^{k_{[nt]}} (0, C_i) \right)_{t \in [0,1]} \xrightarrow[n \rightarrow \infty]{} (0, atM)_{t \in [0,1]}$$

in the rough path topology $\mathcal{C}^{\alpha-\text{Hö}}([0, 1], G^2(V))$ for $\alpha < 1/2$.

Proof. We will use here proposition 3.2. The pointwise convergence is here a hypothesis of the lemma, so what is left to prove is the uniform bound for the

Carnot-Caratheodory norm. On one hand, the upper bound for the Carnot-Caratheodory norm deduced from the norm equivalence (35) implies that

$$\forall (0, y) \in G^2(V), \|(0, y)\| \leq |y|_{V \otimes V}^{1/2}$$

Using the fact that the sequence $(k_n)_n$ is such that, for $m \leq n$, $k_n - k_m \leq n - m$, we obtain

$$\left\| \bigotimes_{i=k_{\lfloor ns \rfloor}+1}^{k_{\lfloor nt \rfloor}} (0, C_i) \right\|^2 = \left\| \left(0, \sum_{i=k_{\lfloor ns \rfloor}+1}^{k_{\lfloor nt \rfloor}} C_i \right) \right\|^2 \leq \left| \sum_{i=k_{\lfloor ns \rfloor}+1}^{k_{\lfloor nt \rfloor}} C_i \right|_{V \otimes V} \leq (\lfloor nt \rfloor - \lfloor ns \rfloor) \xi$$

where ξ is such that for all $i \in \mathbb{N}$, $|C_i|_{V \otimes V} \leq \xi$. We then have:

$$\forall n \geq 1, \mathbb{E} \left[\sup_{s \neq t} \frac{\left\| \delta_{n^{-1/2}} \bigotimes_{i=k_{\lfloor ns \rfloor}+1}^{k_{\lfloor nt \rfloor}} (0, C_i) \right\|}{|t - s|^{1/2}} \right] \leq \mathbb{E} \left[\sup_{s \neq t} \frac{(\lfloor nt \rfloor - \lfloor ns \rfloor)^{1/2} \xi^{1/2}}{(n|t - s|)^{1/2}} \right]$$

which is uniformly bounded as the $(0, C_i)$ s are supposed to be uniformly bounded. \square

We will now give the proof of theorem 2.3. Before doing so, we will give a slightly different formulation of it and explain why we do so. As before, E denotes a finite state space and V a finite-dimensional vector space.

Theorem (theorem 2.3 for non-centred variables). *Let $(R_n, \mathbb{X}_n)_n$ be a hidden Markov walk on $E \times G^2(V)$ that satisfies condition (13). As in property 3.1, we can decompose*

$$\forall n \geq 1, \mathbb{X}_n = \bigotimes_{k=1}^n F_k$$

where $F_k = \mathbb{X}_{k-1}^{-1} \otimes \mathbb{X}_k$.

Furthermore, set $\beta = \mathbb{E}[T_1]$, $v = \delta_{\beta^{-1}} \left(\mathbb{E} \left[\mathbb{X}_{T_1}^{(1)} \right], 0 \right)$, $\tilde{F}_k = v^{-1} \otimes F_k$ and $\tilde{\mathbb{X}}_n = \bigotimes_{k=1}^n \tilde{F}_k$.

In this case, the geodesic embedding of $(\tilde{\mathbb{X}}_n)_n$ in $G^2(V)$ will be

$$\forall t \in [0, 1], \tilde{\mathbb{X}}_t^n = \bigotimes_{k=1}^{\lfloor nt \rfloor} \tilde{F}_k \otimes \delta_{nt - \lfloor nt \rfloor} \tilde{F}_{\lfloor nt \rfloor + 1}$$

Let $\rho = \rho_{N_N}$ be an embedding encoded by $(\tilde{\mathbb{X}}_n)_n$ as in proposition 3.3. If $\mathbb{X}_{T_1}^{(1)}$ is non-degenerate, we can suppose the covariance matrix of $\mathbb{X}_{T_1}^{(1)}$ is CI_d without loss of generality and we have the following convergence in $\mathcal{C}^\alpha([0, 1], G^2(V))$ for $\alpha < 1/2$:

$$\left(\delta_{(nC\beta^{-1})^{-1/2}} \tilde{\mathbb{X}}_t^n \right)_{t \in [0, 1]} \xrightarrow[n \rightarrow \infty]{(d)} \left(\mathbb{B}_t^{Strat} \otimes (0, t\Gamma_\rho) \right)_{t \in [0, 1]}$$

where \mathbb{B}^{Strat} is the Brownian motion enhanced with second-level Stratonovich integrals, i.e. as the second-level limit from theorem 2.3, and Γ_ρ is the deterministic antisymmetric matrix from theorem 2.3.

Alternatively, Γ_ρ can be represented as

$$\Gamma_\rho = \mathbb{E} \left[\frac{1}{2} \sum_{1 \leq p < m \leq T_1} \tilde{F}_p^{(1)} \otimes \tilde{F}_m^{(1)} - \tilde{F}_m^{(1)} \otimes \tilde{F}_p^{(1)} \right] + \mathbb{E} \left[\sum_{p=1}^{T_1} a_p \right] \quad (41)$$

Remarks.

- $(R_n, \tilde{X}_n)_n$ is a HMW in the sense of definition 2.2, and it satisfies the conditions from theorem 2.3: in particular, $\mathbb{E} \left[\tilde{X}_{T_1}^{(1)} \right] = 0$. Thus, the interest of the present version of theorem 2.3 resides in the fact that it allows to treat the case of more general HMW on $G^2(V)$ by recentring their excursions.
- If we compare the statement of the theorem 2.2 given in the introduction to the present one, a fundamental difference is that the first one was more analytic, whereas the present one is more algebraic. We have avoided the "heavy" rough path formulation in the introduction as the rough path setting appears only further in the paper (section 3.2).
- In particular, instead of presenting the first and second level as random processes in the uniform topology $(S_{B,i}(t))$, the limit is presented here as the rough path process $(\mathbb{B}_t^{Strat} \otimes (0, t\Gamma_\rho))_{t \in [0,1]}$. By doing so, we stress the area anomaly obtained at the limit and given by $(0, t\Gamma_\rho)$.

Proof. Since the F_k s are in $G^2(V)$, the particular form of the elements from this group stated in (5) gives the decomposition:

$$\forall k \geq 1, F_k^{(2)} = \frac{1}{2} F_k^{(1)} \otimes F_k^{(1)} + a_k \quad (42)$$

where a_k is the antisymmetric part of $F_k^{(2)}$, $Anti(F_k^{(2)})$.

Thus, using (42), we have the following decomposition for the \tilde{F}_k s:

$$\forall k \geq 1, \tilde{F}_k = (\tilde{F}_k^{(1)}, \frac{1}{2} \tilde{F}_k^{(1)} \otimes \tilde{F}_k^{(1)}) \otimes (0, a_k) \quad (43)$$

where $(\tilde{F}_k^{(1)}, \frac{1}{2} \tilde{F}_k^{(1)} \otimes \tilde{F}_k^{(1)}) \in G^2(V)$ is an element with a symmetric second component and $(0, a_k) \in G^2(V)$ is a "pure area" element (i.e. such that the first component equal to zero and the second component antisymmetric) and is in the centre of the group.

We denote by $\kappa(n)$ the rank of the excursion to which n belongs, i.e. the unique integer such that $T_{\kappa(n)} \leq n < T_{\kappa(n)+1}$, where the T_i s are as defined in (6). Based on the fact that the elements of type $(0, a) \in G^2(V)$ are in the

centre of the group and thus commute with all the others, we have the following decomposition of $\tilde{\mathbb{X}}_t^n$:

$$\begin{aligned}
\tilde{\mathbb{X}}_t^n &= \bigotimes_{k=1}^{\lfloor nt \rfloor} \left((\tilde{F}_k^{(1)}, \frac{1}{2} \tilde{F}_k^{(1)} \otimes \tilde{F}_k^{(1)}) \otimes (0, a_k) \right) \bigotimes \delta_{nt - \lfloor nt \rfloor} \tilde{F}_{\lfloor nt \rfloor + 1} \\
&= \bigotimes_{k=1}^{\kappa(\lfloor nt \rfloor)} \left(\sum_{p=T_{k-1}+1}^{T_k} \tilde{F}_p^{(1)}, \frac{1}{2} \sum_{p=T_{k-1}+1}^{T_k} \tilde{F}_p^{(1)} \otimes \tilde{F}_p^{(1)} + \sum_{T_{k-1}+1 \leq p < m \leq T_k} \tilde{F}_p^{(1)} \otimes \tilde{F}_m^{(1)} \right) \\
&\quad \bigotimes \left(\bigotimes_{k=1}^{\kappa(\lfloor nt \rfloor)} (0, \sum_{p=T_{k-1}+1}^{T_k} a_p) \right) \bigotimes \left(\bigotimes_{k=T_{\kappa(\lfloor nt \rfloor)}+1}^{\lfloor nt \rfloor} \tilde{F}_k \right) \bigotimes \delta_{nt - \lfloor nt \rfloor} \tilde{F}_{\lfloor nt \rfloor + 1} \\
&= \bigotimes_{k=1}^{\kappa(\lfloor nt \rfloor)} \left(\sum_{p=T_{k-1}+1}^{T_k} \tilde{F}_p^{(1)}, \frac{1}{2} \left(\sum_{p=T_{k-1}+1}^{T_k} \tilde{F}_p^{(1)} \right)^{\otimes 2} \right) \\
&\quad \bigotimes \left(\bigotimes_{k=1}^{\kappa(\lfloor nt \rfloor)} (0, \frac{1}{2} \sum_{T_{k-1}+1 \leq p < m \leq T_k} \tilde{F}_p^{(1)} \otimes \tilde{F}_m^{(1)} - \tilde{F}_m^{(1)} \otimes \tilde{F}_p^{(1)} + \sum_{p=T_{k-1}+1}^{T_k} a_p) \right) \\
&\quad \bigotimes \left(\bigotimes_{k=T_{\kappa(\lfloor nt \rfloor)}+1}^{\lfloor nt \rfloor} \tilde{F}_k \right) \bigotimes \delta_{nt - \lfloor nt \rfloor} \tilde{F}_{\lfloor nt \rfloor + 1} \\
&= \mathcal{P}_t^n \otimes \mathcal{A}_t^n \otimes \mathcal{R}_t^n
\end{aligned}$$

where

$$\mathcal{P}_t^n = \bigotimes_{k=1}^{\kappa(\lfloor nt \rfloor)} \left(\sum_{p=T_{k-1}+1}^{T_k} \tilde{F}_p^{(1)}, \frac{1}{2} \left(\sum_{p=T_{k-1}+1}^{T_k} \tilde{F}_p^{(1)} \right)^{\otimes 2} \right)$$

is the term that concatenates the excursions,

$$\mathcal{A}_t^n = \bigotimes_{k=1}^{\kappa(\lfloor nt \rfloor)} \left(\bigotimes_{k=1}^{\kappa(\lfloor nt \rfloor)} (0, \frac{1}{2} \sum_{T_{k-1}+1 \leq p < m \leq T_k} \tilde{F}_p^{(1)} \otimes \tilde{F}_m^{(1)} - \tilde{F}_m^{(1)} \otimes \tilde{F}_p^{(1)} + \sum_{p=T_{k-1}+1}^{T_k} a_p) \right)$$

is a "pure area" process that takes into consideration the antisymmetric part of the $F_k^{(2)}$ s, as well as the stochastic areas of the pseudo-excursions, and

$$\mathcal{R}_t^n = \left(\bigotimes_{k=T_{\kappa(\lfloor nt \rfloor)}+1}^{\lfloor nt \rfloor} \tilde{F}_k \right) \bigotimes \delta_{nt - \lfloor nt \rfloor} \tilde{F}_{\lfloor nt \rfloor + 1}$$

is the rest left from the geodesic embedding.

We will now compute separately the limit of each of the three terms. Let us first consider the term $(\delta_{n^{-1/2}} \mathcal{P}_t^n)_{t \in [0,1]}$. By construction, the variables

$$\left(\sum_{p=T_{k-1}}^{T_k} \tilde{F}_p^{(1)}, \frac{1}{2} \left(\sum_{p=T_{k-1}+1}^{T_k} \tilde{F}_p^{(1)} \right)^{\otimes 2} \right)$$

are i.i.d. and centred. Moreover, $\kappa(\lfloor nt \rfloor)/n \xrightarrow[n \rightarrow \infty]{} t\beta^{-1}$ a.s., since $\kappa(n)$ is the number of full excursions accomplished until time n . Thus, since the function κ satisfies the conditions of lemma 3.1, we can deduce that:

$$\left(\delta_{(nC\beta^{-1})^{-1/2}} \bigotimes_{k=1}^{\kappa(\lfloor nt \rfloor)} \left(\sum_{p=T_{k-1}}^{T_k} \tilde{F}_p^{(1)}, \frac{1}{2} \left(\sum_{p=T_{k-1}+1}^{T_k} \tilde{F}_p^{(1)} \right)^{\otimes 2} \right) \right)_{t \in [0,1]} \xrightarrow[n \rightarrow \infty]{(d)} (\mathbb{B}_t^{Strat})_{t \in [0,1]}$$

in the rough path topology.

We now have to study the convergence of $(\delta_{n^{-1/2}} \mathcal{A}_t^n)_{t \in [0,1]}$. We notice that the two tensor products containing $\kappa(\lfloor nt \rfloor)$ terms are "sums" of i.i.d.r.v., since each term depends entirely on a different excursion. The law of large numbers thus applies to both of them, and we get that, a.s., for $t \in [0, 1]$ fixed:

$$\delta_{(n\beta^{-1})^{-1/2}} \bigotimes_{k=1}^{\kappa(\lfloor nt \rfloor)} \left(0, \frac{1}{2} \sum_{T_{k-1}+1 \leq p < m \leq T_k} \tilde{F}_p^{(1)} \otimes \tilde{F}_m^{(1)} - \tilde{F}_m^{(1)} \otimes \tilde{F}_p^{(1)} \right) \xrightarrow[n \rightarrow \infty]{} (0, t\Gamma)$$

where

$$\Gamma = \mathbb{E} \left[\frac{1}{2} \sum_{1 \leq p < m \leq T_1} \tilde{F}_p^{(1)} \otimes \tilde{F}_m^{(1)} - \tilde{F}_m^{(1)} \otimes \tilde{F}_p^{(1)} \right]$$

is the area anomaly we recover.

For the second part of the sum, we have, for $t \in [0, 1]$ fixed:

$$\delta_{(n\beta^{-1})^{-1/2}} \bigotimes_{k=1}^{\kappa(\lfloor nt \rfloor)} \left(0, \sum_{p=T_{k-1}+1}^{T_k} a_p \right) \xrightarrow[n \rightarrow \infty]{} (0, t\Gamma_0)$$

where

$$\Gamma_0 = \mathbb{E} \left[\sum_{p=1}^{T_1} a_p \right]$$

is generated by the antisymmetric part of the second-level components of the F_k s.

Since κ is as the function from lemma 3.2, the functions \tilde{F}_p are a.s. uniformly bounded under condition (13) and T_1 has finite moments of all order, the conditions of lemma 3.2 are satisfied, and we deduce the following convergence in probability in rough path topology:

$$(\delta_{n^{-1/2}} \mathcal{A}_t^n)_{t \in [0,1]} \xrightarrow[n \rightarrow \infty]{} (0, t\beta^{-1}\Gamma_\rho)_{t \in [0,1]} \quad (44)$$

where $\Gamma_\rho = \Gamma + \Gamma_0$.

We have the residue $(\delta_{n^{-1/2}} \mathcal{R}_t^n)_{t \in [0,1]}$ left to deal with. Its first part contains $\lfloor nt \rfloor - T_{\kappa(\lfloor nt \rfloor)} \leq (T_{\kappa(\lfloor nt \rfloor)+1} - 1) - T_{\kappa(\lfloor nt \rfloor)}$ terms, and $\|\delta_{nt - \lfloor nt \rfloor} \tilde{F}_{\lfloor nt \rfloor + 1}\| \leq$

$\|\tilde{F}_{\lfloor nt \rfloor + 1}\|$ adds one more term. Moreover, since by (13) the F_k s are uniformly bounded, there exists $K > 0$ such that $\sup_k \|\tilde{F}_k\| < K$. Therefore, we get:

$$\mathbb{E} \left[\sup_{s \neq t} \frac{\|\delta_{n-1/2}((\mathcal{R}_s^n)^{-1} \otimes \mathcal{R}_t^n)\|}{|t-s|^{1/2}} \right] \leq K \mathbb{E} \left[\sup_{s \neq t} \frac{T_{\kappa(\lfloor nt \rfloor)} - T_{\kappa(\lfloor ns \rfloor)} + \lfloor nt \rfloor - \lfloor ns \rfloor}{(n|t-s|)^{1/2}} \right]$$

This quantity is bounded since $\lfloor nt \rfloor = \lfloor ns \rfloor$ for $s, t \in [i/n, (i+1)/n]$ for any $i = 1, \dots, n-1$. We also have the convergence in probability

$$\forall t \in [0, 1], \delta_{n-1/2}(\mathcal{R}_t^n) \xrightarrow[n \rightarrow \infty]{} 0$$

We can thus use once again proposition 3.2 to conclude to the following convergence in probability in rough path topology:

$$(\delta_{n-1/2}(\mathcal{R}_t^n))_{t \in [0,1]} \xrightarrow[n \rightarrow \infty]{} 0$$

Finally, putting altogether the convergences of $(\delta_{n-1/2} \mathcal{P}_t^n)_{t \in [0,1]}$, of $(\delta_{n-1/2} \mathcal{A}_t^n)_{t \in [0,1]}$ and of $(\delta_{n-1/2} \mathcal{R}_t^n)_{t \in [0,1]}$, our result follows from Slutsky's theorem:

$$(\delta_{(nC\beta-1)-1/2} \mathcal{P}_t^n \otimes \mathcal{A}_t^n \otimes \mathcal{R}_t^n)_{t \in [0,1]} \xrightarrow[n \rightarrow \infty]{(d)} (\mathbb{B}_t^{Strat} \otimes (0, t\Gamma_\rho))_{t \in [0,1]}$$

□

4 Iterated structures behind discrete time and discrete space Markov chains.

In this section we study more thoroughly the algebraic properties of HMW through the corresponding iterated occupation times.

4.1 Shuffle and quasi-shuffle products.

The definitions and properties from this section are mainly inspired from [9].

Definitions.

Definition 4.1 (the quasi-shuffle product). *Let (A, \cdot) be an algebra, $a, b \in A$, and x, y obtained by concatenation of a finite number of elements from A . Then the quasi-shuffle product $*$ is defined recursively by:*

$$ax * by = [a, b](x * y) + a(x * by) + b(ax * y)$$

where $[\cdot, \cdot]$ is a commutative operation on A .

The shuffle product can be viewed as a particular case of the quasi-shuffle product. More specifically, it is the case when $a \cdot b = 0$ for any $a, b \in R$. Contrary to the shuffle product, the quasi-shuffle product is not always commutative.

Definition 4.2 (shuffle product). *Let (A, \cdot) be an algebra, $a, b \in A$, and x, y obtained by concatenation of a finite number of elements from A . Then the shuffle product \sqcup is defined recursively by:*

$$ax \sqcup by = a(x \sqcup by) + b(ax \sqcup y)$$

A more informal way of putting it is as follows:

Definition 4.3 (and corollary of 4.2). *Consider the set of permutation $\Sigma_{m,n} = \{\sigma \in \Sigma_{m+n} : \sigma(1) < \dots < \sigma(m), \sigma(m+1) < \dots < \sigma(m+n)\}$. For $x = x_1 \dots x_m$ and $y = y_1 \dots y_n$ with $x_i, y_j \in A$, the shuffle product of x and y is given by the sum of images of the concatenated word ab under all the permutations of $\Sigma_{m,n}$: $x \sqcup y = \sum_{\sigma \in \Sigma_{m,n}} \sigma(xy)$ (for example, $ab \sqcup c = abc + acb + cab$).*

Quasi-shuffle structure of iterated sums. One of the main example of sets we can endow with a shuffle product is the set of iterated integrals on a vector space V . Denote by $\mathbf{e}_n^{i_1, \dots, i_n} = e_{i_1} \otimes \dots \otimes e_{i_n}$ and $\mathbf{e}_m^{j_1, \dots, j_m} = e_{j_1} \otimes \dots \otimes e_{j_m}$ where the e_k s are vectors from the canonical basis of V . We then define, for an integrable path y in V ,

$$\begin{aligned} S_{y,n;i_1, \dots, i_n}(t) &= \langle \mathbf{e}_n^{i_1, \dots, i_n}, S_{y,n}(t) \rangle \\ &= \int_{0 < t_1 < \dots < t_n < t} dy_{t_1}^{i_1} \dots dy_{t_n}^{i_n} \end{aligned}$$

and idem for $\mathbf{e}_m^{j_1, \dots, j_m}$, with $\langle \cdot, \cdot \rangle$ the induced scalar product on $V^{\otimes l}$. The shuffle product of two iterated integrals depending both on the same path y is then given by:

$$S_{y,n;i_1, \dots, i_n}(t) S_{y,m;j_1, \dots, j_m}(t) = S_{y,n+m;(i_1, \dots, i_n) \sqcup (j_1, \dots, j_m)}(t) \quad (45)$$

showing that the shuffle product of two iterated integrals is a linear combination of iterated integrals.

If we now consider a sequence $x = (x_n)_n$ in V , the components of the iterated sum $\tilde{S}_{x,n}(N)$ from (24) given by

$$\begin{aligned} \tilde{S}_{x,n;i_1, \dots, i_n}(N) &= \langle \mathbf{e}_n^{i_1, \dots, i_n}, \tilde{S}_{x,n}(N) \rangle \\ &= \sum_{1 \leq k_1 < \dots < k_n \leq N} x_{k_1}^{i_1} \dots x_{k_n}^{i_n} \end{aligned} \quad (46)$$

are discrete versions of (45).

However, the product of two elements of this kind is not a shuffle product anymore, as shown below.

$$\begin{aligned} \tilde{S}_{x,n;i_1, \dots, i_n}(N) \tilde{S}_{x,m;j_1, \dots, j_m}(N) &= \tilde{S}_{x,n+m;(i_1, \dots, i_n) \sqcup (j_1, \dots, j_m)}(N) \\ &\quad + r_{x,n,m;i_1, \dots, i_n; j_1, \dots, j_m}(N) \end{aligned} \quad (47)$$

where $r_{x,n,m;i_1, \dots, i_n; j_1, \dots, j_m}(N)$ is a rest that comes from the fact that, when multiplying two iterated sums, we get sets of indices that are not necessarily

strictly ordered. Moreover, the result of the multiplication can not be in the form of a function \tilde{S} so we can not apply the new product exclusively to $\mathbf{e}_n^{i_1, \dots, i_n}$ and $\mathbf{e}_m^{j_1, \dots, j_m}$. We thus need a more general product, that would better keep track of this rest. The product of two iterated sums of the type $S_{x,n;i_1, \dots, i_n}(N)$ is a quasi-shuffle product, as already stated in property 2.2.

4.2 From geometric to non-geometric rough paths through hidden Markov walks

4.2.1 Geometric rough paths and shuffle products.

We know that a piecewise linear embedding of $(X_n)_n$ is canonically represented by a (geometric) rough path as follows:

Property 4.1. *Denote by $(X_t^{(N)})_t$ a smooth embedding of $(X_n)_n$. The canonical rough path in $G^2(V)$ corresponding to this embedding is given by the 2-step signature*

$$\forall t \in [0, 1], S_2(X^{(N)})_{0,t} = (X_t^{(N)}, \int_{0 < s_1 < s_2 < t} dX_{s_1}^{(N)} \otimes dX_{s_2}^{(N)})$$

The path $t \mapsto S_2(X^{(N)})_{0,t}$ is in $\mathcal{C}^{1-var}([0, 1], G^2(V))$.

We thus obtain a sequence of geometric rough paths and can study the convergence in law of $(\delta_{N^{-1/2}} S_2(X^{(N)})_{0,t})_t$ in the topology of $\mathcal{C}^{\alpha-H\ddot{o}l}([0, 1], G^2(V))$ for $\alpha < 1/2$. The limit (if it exists) will be a *geometric α -Hölder rough path*, i.e. an element of $\mathcal{C}^{\alpha-H\ddot{o}l}([0, 1], G^2(V))$, for $\alpha < 1/2$.

Since $G^2(V)$ is the space in which geometric rough paths take values, an element $(g^{(1)}, g^{(2)}) \in G^2(V)$ needs to satisfy the shuffle product relation in the sense of the following lemma.

Property 4.2. *An element $(g^{(1)}, g^{(2)}) \in T_1^{(2)}(V)$ is in $G^2(V)$ if and only if it satisfies the relation*

$$\langle g^{(1)}, e_i \rangle \langle g^{(1)}, e_j \rangle = \langle g^{(2)}, e_i \sqcup e_j \rangle \quad (48)$$

where e_i, e_j are in the canonical basis of V and $e_i \sqcup e_j = e_i \otimes e_j + e_j \otimes e_i$.

Proof. We can decompose the second component into a symmetric and an antisymmetric part :

$$g^{(2)} = \text{Sym}(g^{(2)}) + \text{Antisym}(g^{(2)})$$

On one hand, we have

$$\begin{aligned} \langle \text{Sym}(g^{(2)}), e_i \sqcup e_j \rangle &= \frac{1}{2} \langle g^{(1)} \otimes g^{(1)}, e_i \otimes e_j + e_j \otimes e_i \rangle \\ &= \langle g^{(1)}, e_i \rangle \langle g^{(1)}, e_j \rangle \end{aligned}$$

On the other hand, for the antisymmetric term we get

$$\langle \text{Antisym}(g^{(2)}), e_i \sqcup e_j \rangle = \langle \text{Antisym}(g^{(2)}), e_i \otimes e_j \rangle - \langle \text{Antisym}(g^{(2)}), e_j \otimes e_i \rangle = 0$$

which achieves the proof. \square

Remark. The condition (48) is a version of the relation which is part of the abstract definition of geometric rough paths given in [7].

4.2.2 A discrete construction for non-geometric rough paths.

We now present a way of constructing non-geometric rough paths out of a hidden Markov walk $(R_n, X_n)_n$. Instead of first constructing a continuous path $(X_t^{(N)})_t$ in V and then associate a rough path to it, we will first construct a sequence $(\mathbb{Y}_n)_n$ in $T_1^{(2)}(V)$ and then associate an embedding to it.

Let us consider the sequence $(\mathbb{Y})_n$ as in (17), i.e.

$$\mathbb{Y}_n = \left(X_n, \sum_{1 \leq k_1 < k_2 \leq n} F_{k_1} \otimes F_{k_2} \right) \in T_1^{(2)}(V)$$

and explain how connect the non-geometric nature of the corresponding rough path to the quasi-shuffle product.

In the case of the sequence $(\mathbb{Y}_k)_k$, we have:

$$\begin{aligned} \langle X_n, e_i \rangle \langle X_n, e_j \rangle &= \sum_{k_1, k_2=1}^n F_{k_1}^{(i)} F_{k_2}^{(j)} \\ &= \sum_{1 \leq k_1 < k_2 \leq n} F_{k_1}^{(i)} F_{k_2}^{(j)} + \sum_{1 \leq k_2 < k_1 \leq n} F_{k_1}^{(i)} F_{k_2}^{(j)} + \sum_{k=1}^n F_k^{(i)} F_k^{(j)} \end{aligned} \quad (49)$$

$$\begin{aligned} &= \langle \sum_{1 \leq k_1 < k_2 \leq n} F_{k_1} \otimes F_{k_2}, e_i \sqcup e_j \rangle + \langle \sum_{k=1}^n F_k^{\otimes 2}, e_i \otimes e_j \rangle \end{aligned} \quad (50)$$

$$= \langle \sum_{1 \leq k_1 < k_2 \leq n} F_{k_1} \otimes F_{k_2} + \frac{1}{2} \sum_{k=1}^n F_k^{\otimes 2}, e_i \sqcup e_j \rangle$$

Once again, we see that (50) does not coincide with (48). Moreover, if we identify e_i as corresponding to the index k_1 and e_j as corresponding to the index k_2 , the line (49) in the computation above can be identified with *the quasi-shuffle product*: the sums over $\{k_1 < k_2\} \cup \{k_2 < k_1\}$ give the shuffle product part, whereas the last sum over $\{k_1 = k_2\}$ gives the quasi-shuffle rest.

From all the above, we can draw the following conclusion:

Property 4.3. *The sequence $(\mathbb{Y}_n)_n$ takes values in $T_1^{(2)}(V) \setminus G^2(V)$. In particular, any embedding $(\mathbb{Y}_t^{(N)})_t$ of $(\mathbb{Y}_n)_n$, no matter how smooth, will be a non-geometric rough path.*

4.3 Getting back to proposition 2.1.

We want to study the convergence of (17) and the convergence of the rough paths $(\mathbb{Y}_t^{(N)})_{t \in [0,1]}$ which are obtained from the sequence $(\mathbb{Y}_n)_n$ by a continuous

embedding in the sense of definition 3.4.

The first thing we need to notice is that the sequence $(\mathbb{Y}_n)_n$ is not in $G^2(V)$. In the present case, we have

$$\begin{aligned} \text{Sym} \left(\sum_{1 \leq k_1 < k_2 \leq n} F_{k_1} \otimes F_{k_2} \right) &= \frac{1}{2} \left(\sum_{1 \leq k_1 < k_2 \leq n} F_{k_1} \otimes F_{k_2} + \sum_{1 \leq k_1 < k_2 \leq n} F_{k_2} \otimes F_{k_1} \right) \\ &= \frac{1}{2} X_n \otimes X_n - \frac{1}{2} \sum_{k=1}^n F_k^{\otimes 2} \end{aligned}$$

and thus we have an extra term that does not depend on the first level X_n .

In order to study the convergence of the non-geometric rough paths $(\mathbb{Y}_t^{(N)})_t$ associated to $(\mathbb{Y}_n)_n$ in property 4.3, we need to answer the following questions:

- What embedding do we choose for $(\mathbb{Y}_n)_n$?
- What is the topology to consider when studying the convergence?
- What will be the nature of the limit rough path (if it exists)?

We start with a few preliminary results. First, we want to show that property 4.1 can be used to construct a suitable embedding for $(\mathbb{X}_n)_n$. This is a consequence of the following lemma:

Lemma 4.1. *Choose $(X_t^{(N)})_t$ to be the geodesic embedding of $(X_n)_n$ and suppose $X_0 = 0$ a.s. For all $N \geq 1$ and $1 \leq i \leq n$, we have:*

$$S_2(X^{(N)})_{0,i/N} = \mathbb{Y}_i \otimes \left(0, \sum_{k=1}^i F_k^{\otimes 2} \right)$$

Proof. For any $N \geq 1$ and $t \in [0, 1]$, we can decompose the second component of $S_2(X^{(N)})_{0,t}$ into a symmetric and an antisymmetric part as follows:

$$\begin{aligned} \int_{0 < s_1 < s_2 < t} dX_{s_1}^{(N)} \otimes dX_{s_2}^{(N)} &= \frac{1}{2} \int_{0 < s_1 < s_2 < t} dX_{s_1}^{(N)} \otimes dX_{s_2}^{(N)} + dX_{s_2}^{(N)} \otimes dX_{s_1}^{(N)} \\ &\quad + \frac{1}{2} \int_{0 < s_1 < s_2 < t} dX_{s_1}^{(N)} \otimes dX_{s_2}^{(N)} - dX_{s_2}^{(N)} \otimes dX_{s_1}^{(N)} \end{aligned}$$

Since $(X_t^{(N)})_t$ is piecewise linear, the symmetric part becomes

$$\frac{1}{2} \int_{0 < s_1 < s_2 < t} dX_{s_1}^{(N)} \otimes dX_{s_2}^{(N)} + dX_{s_2}^{(N)} \otimes dX_{s_1}^{(N)} = (X_t^{(N)})^{\otimes 2}$$

The antisymmetric part corresponds to the stochastic area of the process $(X_t^{(N)})_t$, $A_{i/N}(X^{(N)})$, which, for a piecewise linear process, can be expressed by second

iterated sums, i.e.

$$\begin{aligned}
\forall i \in \{1, \dots, N\}, \quad A_i(X) &:= A_{i/N}(X^{(N)}) \\
&= \frac{1}{2} \int_{0 < s_1 < s_2 < i/N} dX_{s_1}^{(N)} \otimes dX_{s_2}^{(N)} - dX_{s_2}^{(N)} \otimes dX_{s_1}^{(N)} \\
&= \frac{1}{2} \sum_{1 \leq k_1 < k_2 \leq i} F_{k_1} \otimes F_{k_2} - F_{k_2} \otimes F_{k_1}
\end{aligned}$$

Since we have

$$\forall i \in \{1, \dots, N\}, \quad \sum_{1 \leq k_1 < k_2 \leq i} F_{k_1} \otimes F_{k_2} + F_{k_2} \otimes F_{k_1} = (X_{i/N}^{(N)})^{\otimes 2} - \sum_{k=1}^i F_k^{\otimes 2}$$

it is straightforward to show

$$\sum_{1 \leq k_1 < k_2 \leq i} F_{k_1} \otimes F_{k_2} = \int_{0 < s_1 < s_2 < i/N} dX_{s_1}^{(N)} \otimes dX_{s_2}^{(N)} - \sum_{k=1}^i F_k^{\otimes 2} \quad (51)$$

We thus conclude that:

$$\forall N \geq 1, \forall 1 \leq i \leq N, \quad S_2(X^{(N)})_{0, i/N} = \left(X_i, \sum_{1 \leq k_1 < k_2 \leq i} F_{k_1} \otimes F_{k_2} \right) \otimes \left(0, \sum_{k=1}^i F_k^{\otimes 2} \right)$$

□

We can now give an answer to the two questions concerning the choice of the embedding and of the topology:

Lemma 4.2. *Choose $(X_t^{(N)})_t$ to be the geodesic embedding of $(X_n)_n$ and suppose $X_0 = 0$ a.s. Consider the geodesic embedding of $Z_k = \sum_{j=1}^k F_j^{\otimes 2}$ given by*

$$\forall N \in \mathbb{N}^*, \forall t \in [0, 1], \quad Z_t^{(N)} = \sum_{k=1}^{\lfloor Nt \rfloor} F_k^{\otimes 2} + (Nt - \lfloor Nt \rfloor) F_{\lfloor Nt \rfloor + 1}^{\otimes 2}$$

and set

$$\forall N \in \mathbb{N}^*, \forall t \in [0, 1], \quad \mathbb{Y}_t^{(N)} = (0, Z_t^{(N)})^{-1} \otimes S_2(X^{(N)})_{0, t}$$

Then $t \mapsto \mathbb{Y}_t^{(N)}$ is an embedding for $(\mathbb{Y}_n)_n$. Furthermore, $t \mapsto \mathbb{Y}_t^{(N)}$ is in $\mathcal{C}^{1-var}([0, 1], T_1^{(2)}(V))$ and is a non-geometric rough path.

Proof. The path $t \mapsto \mathbb{Y}_t^{(N)}$ is constructed by concatenation of two rough paths, $(S_2(X^{(N)})_{0, t})_t$ being a smooth rough path and $(0, Z_t^{(N)})_t$ a smooth path in $T_1^{(2)}(V)$, which implies that $t \mapsto \mathbb{Y}_t^{(N)}$ is an element of $\mathcal{C}^{1-var}([0, 1], T_1^{(2)}(V))$. Moreover, property 4.3 implies that $t \mapsto \mathbb{Y}_t^{(N)}$ is a non-geometric rough path.

The fact that $t \mapsto \mathbb{Y}_t^{(N)}$ is an embedding for $(\mathbb{X}_n)_n$ follows from lemma 4.1, which tells us that:

$$\forall N \geq 1, \forall 1 \leq i \leq N, \mathbb{Y}_{i/N}^{(N)} = (0, Y_{i/N}^{(N)})^{-1} \otimes S_2(X^{(N)})_{0, i/N}$$

□

Remark. Another way of expressing $t \mapsto \mathbb{Y}_t^{(N)}$ is through *the rough path bracket* which concentrates the symmetric part of the second-level component (see definition 5.5 from [5]), and implies:

$$\forall t \in [0, 1], \mathbb{Y}_t^{(N)} = (X_t^{(N)}, A_t(X^{(N)})) \otimes (0, \frac{1}{2}[\mathbb{Y}^{(N)}]_t)$$

where

$$\forall t \in [0, 1], [\mathbb{Y}^{(N)}]_t = X_t^{(N)} \otimes X_t^{(N)} - Z_t^{(N)}$$

and $t \mapsto A_t(X^{(N)})$ is, as before, the stochastic area of $t \mapsto X_t^{(N)}$.

We can now proceed to the main proof of proposition 2.1.

Proof. Theorem 2.1 implies the following convergence in law in the space $\mathcal{C}^{\alpha-\text{Hö}}([0, 1], G^2(V))$ for $\alpha < 1/2$:

$$\left(\delta_{(NC\beta^{-1})^{-1/2}} S_2(X^{(N)})_{0,t} \right)_{t \in [0,1]} \xrightarrow{N \rightarrow \infty} (\mathbb{B}_t^{\text{Strat}} \otimes (0, \Gamma t))_{t \in [0,1]} \quad (52)$$

where Γ is the area anomaly from 2.1.

Next, we apply the decomposition in pseudo-excursions to Z_n i.e.

$$\forall n \geq 1, Z_n = \sum_{j=0}^{\kappa(n)-1} \sum_{i=T_j+1}^{T_{j+1}} F_i^{\otimes 2} + \sum_{i=T_{\kappa(n)}+1}^n F_i^{\otimes 2}$$

and, by the law of large numbers, we deduce the convergence of the finite-dimensional marginals

$$\forall t \in [0, 1], \delta_{(NC\beta^{-1})^{-1/2}}(0, Z_t^{(N)}) \xrightarrow{N \rightarrow \infty} \left(0, tC^{-1} \mathbb{E} \left[\sum_{i=1}^{T_1} F_i^{\otimes 2} \right] \right)$$

We can now prove the tightness of the sequence of processes $(0, Z_t^{(N)})_t$ in $\mathcal{C}^{\alpha-\text{Hö}}([0, 1], T_1^{(2)}(V))$ for $\alpha < 1/2$, by using a version of the Kolmogorov criterion. Theorem 3.10 from [5] implies that it is enough to prove

$$\exists K > 0, \forall s, t \in [0, 1], \forall N \geq 1, N^{-1} \mathbb{E} \left[|Z_{s,t}^{(N)}|_{V \otimes V} \right] \leq K|t - s|$$

We use the fact that the F_k s are uniformly bounded:

$$\begin{aligned} N^{-1} \mathbb{E} \left[|Z_{s,t}^{(N)}|_{V \otimes V} \right] &= N^{-1} \mathbb{E} \left[\left| \sum_{i=\lfloor Ns \rfloor + 1}^{\lfloor Nt \rfloor} F_i^{\otimes 2} \right|_{V \otimes V} \right] \\ &\leq K|t - s| \end{aligned}$$

where K comes from the condition (8).

We thus conclude to the following convergence in probability in the $\mathcal{C}^{\alpha-\text{H\"{o}l}}([0, 1], T_1^{(2)}(V))$ topology for $\alpha < 1/2$:

$$\left(\delta_{(NC\beta^{-1})^{-1/2}}(0, Z_t^{(N)})\right)_{t \in [0,1]} \xrightarrow{N \rightarrow \infty} \left(0, tC^{-1}\mathbb{E}\left[\sum_{i=1}^{T_1} F_i^{\otimes 2}\right]\right)_{t \in [0,1]} \quad (53)$$

Using the Slutsky's theorem (theorem 3.1 from [1]), and combining (52) and (53), we get

$$\left(\delta_{(NC\beta^{-1})^{-1/2}}\mathbb{Y}_t^{(N)}\right)_{t \in [0,1]} \xrightarrow{N \rightarrow \infty} \left(\mathbb{B}_t^{\text{Strat}} \otimes (0, t(\Gamma - C^{-1}\mathbb{E}\left[\sum_{i=1}^{T_1} F_i^{\otimes 2}\right]))\right)_{t \in [0,1]} \quad (54)$$

where $\mathbb{B}^{\text{Strat}}$ is the standard Brownian motion $(B_t)_t$ on V enhanced with its iterated integrals in the Stratonovich sense.

At the same time, we can compute

$$\Gamma - C^{-1}\mathbb{E}\left[\sum_{i=1}^{T_1} F_i^{\otimes 2}\right] = C^{-1}\mathbb{E}\left[\frac{1}{2}\sum_{1 \leq k_1 < k_2 \leq T_1} F_{k_1} \otimes F_{k_2} - F_{k_2} \otimes F_{k_1}\right] - C^{-1}\mathbb{E}\left[\frac{1}{2}\sum_{i=1}^{T_1} F_i^{\otimes 2}\right] \quad (55)$$

$$\begin{aligned} &= C^{-1}\mathbb{E}\left[\sum_{1 \leq k_1 < k_2 \leq T_1} F_{k_1} \otimes F_{k_2}\right] - \frac{1}{2}C^{-1}\mathbb{E}[X_{T_1}^{\otimes 2}] \\ &= C^{-1}\mathbb{E}\left[\sum_{1 \leq k_1 < k_2 \leq T_1} F_{k_1} \otimes F_{k_2}\right] - \frac{1}{2}Id = M - \frac{1}{2}Id \end{aligned} \quad (56)$$

where $M = C^{-1}\mathbb{E}\left[\sum_{1 \leq k_1 < k_2 \leq T_1} F_{k_1} \otimes F_{k_2}\right]$ (as stated in (18)) and $\mathbb{E}[X_{T_1}^{\otimes 2}]$ is non other than the covariance matrix of X_{T_1} since the excursions are centred.

The expression (55) shows that $(0, Mt)$ is a non-geometric rough path: while the first term of the sum is antisymmetric, the second one is symmetric and can not be expressed through the first component of the rough path (which is zero). This shows that the limit of the $(\mathbb{Y}_t^{(N)})_t$ is a non-geometric rough path.

The expression (56) shows that we can rewrite the limit using Itô integration, as

$$\forall 0 \leq s < t \leq 1, \mathbb{B}_{t,s}^{\text{Strat}} = \mathbb{B}_{t,s}^{\text{It}\hat{o}} \otimes (0, \frac{1}{2}(t-s)Id)$$

We thus obtain the convergence:

$$\left(\delta_{(NC\beta^{-1})^{-1/2}}\mathbb{Y}_t^{(N)}\right)_{t \in [0,1]} \xrightarrow{N \rightarrow \infty} \left(\mathbb{B}_t^{\text{It}\hat{o}} \otimes (0, tM)\right)_{t \in [0,1]}$$

which achieves the proof. \square

Remark. We find in the expression of the limit (54) the decomposition of a non geometric rough path in a geometric rough path plus an element in the centre of $T_1^{(2)}(V)$, highlighted in particular in [5] and expressed by

$$\forall 1/3 < \alpha < 1/2, \mathcal{C}^\alpha([0, 1], T_1^{(2)}(V)) \simeq \mathcal{C}^\alpha([0, 1], G^2(V)) \oplus \mathcal{C}^{2\alpha}([0, 1], V \otimes V)$$

5 Open questions

We have seen that hidden Markov walks are a natural generalization of the class of Markov chains on periodic graphs when one wants to study the area anomaly using techniques related to the theory of excursions, in particular allowing to analyse the convergence of a HMW as that of a sum of i.i.d. variables. It would be therefore interesting to know whether we can obtain results on area anomaly on other processes which present only weak time-correlations that could be ignored when passing to the limit in uniform convergence topology (for example, the α -mixing processes described in [1]).

A further study of the iterated occupation times $L_{u_1, \dots, u_k; N}(R)$ 2.5 of a Markov chain $(R_n)_n$ could also prove fruitful. Several directions can be considered: a generalization of some results of the ergodic theory (in particular, the connection with the invariant measure), their study as random combinatorial objects (using their representation as cardinals), the construction and analysis of an abstract vector space generated by this objects (which would in particular allow an abstract representation of a HMW depending on $(R_n)_n$), etc.

Furthermore, using the example of HMW, we would like to see if there is a connection between the iterated occupation times as combinatorial structures and the Hopf algebras which describe the combinatorics of (abstract) rough paths (as in [7]).

We have seen that, in the case of HMW, iterated sums allow, for example, to get a decomposition of rough paths that isolates the area anomaly or to propose an interesting construction of non-geometric rough paths. It can prove useful to continue studying these objects (possibly for more general processes) as they may contribute on one hand to developing the discrete setting in the rough paths theory and on the other hand to construct some concrete examples which allow explicit computations, which is often difficult when it comes to rough paths.

It could be interesting to use the framework we present in this paper to study in a new way finite difference equations and in particular their asymptotics and continuous limits; or, in the reverse way, if new discrete models could be imagined to discretize efficiently classical (stochastic) differential equations.

Overall, we hope that the present paper will provide the reader with good arguments to be interested in the role of the area anomaly in the classical stochastic calculus and that of discrete processes in the classical rough paths setting.

References

- [1] P. Billingsley. *Convergence of Probability Measures*. Wiley, 1999.
- [2] E. Breuillard, P. Friz, and M. Huesmann. From random walks to rough paths. *Proc. Amer. Math. Soc.*, 137(10):3487–3496, 2009.
- [3] A. Davie. Differential equations driven by rough paths: an approach via discrete approximation. *Appl. Math. Res. Express*, pages 009–40, 2007.
- [4] S. Eilenberg and S. Mac Lane. On the groups $H(\Pi, n)$, I. *Annals of Mathematics*, 58(1):55–106, 1953.
- [5] P. Friz and M. Hairer. *A Course on Rough Paths - With an introduction to regularity structures*. Springer, 2014.
- [6] P. Friz and N. Victoir. *Multidimensional Stochastic Processes as Rough Paths: Theory and Applications*. Cambridge University Press, 2010.
- [7] M. Hairer and D. Kelly. Geometric versus non-geometric rough paths. *Annals of Institute Henri Poincaré*, 51(1):207–251, 2015.
- [8] M. Hoffman. Quasi-shuffle products. *Journal of Algebraic Combinatorics*, 11:49–68, 2000.
- [9] M. Hoffman and K. Ihara. Quasi-shuffle products revisited. *Journal of Algebra*, 481:293–326, 2017.
- [10] J. Norris. *Markov Chains*. Cambridge University Press, 1997.
- [11] A. Lejay and T. Lyons. On the importance of the Lévy area for studying the limits of functions of converging stochastic processes. application to homogenization. *Current Trends in Potential Theory*, page 63–84, 2003.
- [12] Y. Liu and S. Tindel. Discrete rough paths and limit theorems. arXiv:1707.01783, 2017.
- [13] O. Lopusanschi and D. Simon. Lévy area with a drift as a renormalization limit of Markov chains on periodic graphs. *Accepted for publication in Stochastic Processes and Applications*, 2017.
- [14] T. Lyons, M. Caruana, and Th. Lévy. *Differential Equations Driven by Rough Paths (Ecole d’Eté de Probabilités de Saint-Flour XXXIV)*. Springer, 2004.
- [15] T. Lyons and Z. Qian. Flow equations on spaces of rough paths. *Journal of Functional Analysis*, 149:135–159, 1997.
- [16] L. Young. *The Theory of Integration*. The University Press, 1927.