

FUNCTIONS ON UNPARAMETERISED PATH SPACE

SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY OF IMPERIAL
COLLEGE LONDON AND THE DIPLOMA OF IMPERIAL
COLLEGE LONDON BY

WILLIAM F. TURNER

IMPERIAL

IMPERIAL COLLEGE LONDON
DEPARTMENT OF MATHEMATICS

SEPTEMBER 2025

To my family.

– DECLARATION OF ORIGINALITY –

I certify that this thesis and the research to which it refers are the product of my own work, and that any ideas or quotations from the work of other people, published or otherwise, are fully acknowledged in accordance with the standard referencing practices of the discipline.

Parts of this thesis have previously been published or were written in collaboration with other authors. All co-authors are aware of the inclusion of the following material in this thesis.

1. The majority of the work presented in [Chapter 2](#) is based on [\[CT24b\]](#) and was published in *Journal of Functional Analysis* 286.4 (2024) as **Topologies on unparameterised path space** by Thomas Cass and William F. Turner. The remainder of this chapter is based on the preprint [\[CT24c\]](#), also written in collaboration with Thomas Cass.
2. Most of [Chapter 3](#) is based on the preprint [\[CT24a\]](#) which has been accepted for publication in the *Annals of Applied Probability*. This work was written in collaboration with my supervisor Thomas Cass. Part of this chapter also is based on the recent preprint [\[Cre+25\]](#) which was co-authored with Samuel Crew, Cristopher Salvi, Thomas Cass and Antoine Jacquier. Samuel Crew, Cristopher Salvi and myself contributed equally to this work.
3. The material presented in [Chapter 4](#) is based on [\[CMT23\]](#) and was published in *Electronic Journal of Probability* 28 (2023) as **Signature asymptotics, empirical processes, and optimal transport** by Thomas Cass, Remy Messadene and William F. Turner.
4. [Chapter 5](#) is based on [\[Jel+23\]](#) which has been accepted for publication in the *SIAM Journal on Mathematics of Data Science*. This work was written in collaboration with Andraž Jelinčič, Jiajie Tao, Thomas Cass James Foster and Hao Ni. Andraž Jelinčič, Jiajie Tao and myself shared equal contribution to this paper.

– COPYRIGHT STATEMENT –

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a [Creative Commons Attribution-Non Commercial 4.0 International Licence](#) (CC BY-NC). Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that: you credit the author and do not use it, or any derivative works, for a commercial purpose. When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes. Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

© 2025 William F. Turner

– ABSTRACT –

This thesis studies functions and topologies on unparameterised path space, obtained as the quotient of parameterised path space by tree-like equivalence. This space has gained prominence in recent years through its connection to the signature transform, a key tool in modern machine learning.

We introduce and analyse candidate topologies for unparameterised path space in both bounded-variation and rough-path settings, focusing on separability, metrisability, and Polishness. We also investigate how the choice of topology influences uniform approximation with linear functionals on the signature and the expected signature model.

Turning to functions on unparameterised path space, we examine two examples: scaling limits of randomised path developments and signature asymptotics. For the former, we generalise a recently established link between randomised general linear group path developments and the ordinary signature kernel to a broader class of randomisations. We also extend this framework to randomised unitary path developments, where connections with non-commutative probability emerge. For the latter we extend a result of Hambly & Lyons, who showed that the Hilbert–Schmidt norms of suitably rescaled signature terms recover a weighted curvature of the tree-reduced representative of each equivalence class. By exploiting connections between Wasserstein distances and expected fluctuations of Brownian bridges, we establish this result for a class of C^2 curves improving on the original C^3 assumption.

Finally, we propose LévyGAN: a generative modelling approach for producing approximate samples from the Lévy area of multi-dimensional Brownian motion, a quantity central to high-order SDE solvers. Our method leverages symmetries in the joint distribution of Brownian increments and Lévy area, removing the need for training data. We demonstrate its application through multilevel Monte Carlo simulation of the log-Heston model.

– ACKNOWLEDGEMENTS –

First and foremost I would like to thank my supervisor Prof. Thomas Cass, without whom this thesis could not have been possible. His time, dedication, insights and patience have shaped my mathematical journey. I am grateful to have been able to explore so many different aspects of an area that is at the forefront of both modern stochastic analysis and data science.

I am also grateful to all the fantastic co-authors that I have had the pleasure of collaborating with, including Remy Messadene, Andraž Jelinčič, Jiajie Tao, James Foster, Hao Ni, Samuel Crew, Antoine Jacquier and Cristopher Salvi. I have learnt a huge amount from every one of our discussions. There are many other (too many to list!) amazing mathematicians that I have met along the way who have helped shape this thesis.

Next, I would like to thank my friends. To those with whom I shared the PhD office, thank you for the many ideas and wonderful conversations. To those outside the office, thank you for helping me think about things beyond this thesis.

I would like to thank Tim and Stephanie for their support, especially while I have been writing this thesis over the past year. To Ashton, you bring constant joy to my life and your support has always kept me moving forward.

Finally, I would like to thank my parents for their never-ending encouragement. I would never have undertaken this journey without your unconditional support.

– PREFACE –

An unparameterised path is the object one obtains by starting from a parameterised path $\gamma : [0, T] \rightarrow \mathbb{R}^d$ and forgetting all the information that leaves terminal time solutions to controlled differential equations driven by γ unchanged. Such information includes not only the parameterisation $[0, T]$ but also any parts of γ that retrace themselves. When two parameterised paths lead to the same unparameterised path, they are called tree-like equivalent. The space of unparameterised paths has become increasingly relevant in the past decade due to the popularity of the signature transform as a tool in modern machine learning. Indeed, the seminal 2010 paper [HL10] of Hambly & Lyons proved that two paths are tree-like equivalent if and only if their signatures are equal. The signature of a path is a graded collection of tensors given by the set of all iterated integrals of the path. The first level of the signature is the increment $\gamma_T - \gamma_0$, while the second level is related to the area captured between coordinates of the path with higher order terms also having geometric interpretations. The signature transform is the analogue to paths for what the collection of monomials is for vector valued data. In particular, the usage of the signature transform in applications is motivated by the following two facts which hold under suitable assumptions:

1. Any continuous function from a compact set of unparameterised paths to the real numbers may be uniformly approximated by a linear functional on the signature – *polynomials are dense in continuous functions on compact sets of $\mathbb{R}^d$* ;
2. The law of a random variable on unparameterised path space is characterised by its expected signature [CL16; CO22] – *a vector valued random variable is characterised by its moment sequence*.

We address the question of finding a suitable topology on unparameterised path space in Chapter 2. We study a collection of candidate topologies, each motivated by privileging different aspects of the signature transform, tree-like equivalence or the underlying path space. Motivated by applications to probability theory, we concentrate on whether any of the candidate topologies are Polish. Unfortunately,

our main result (loosely) states that for topologies under which the quotient map is continuous, there exists no Polish topology on the space of unparameterised p -variation rough paths that extends (in a natural fashion) to a Polish topology on unparameterised q -variation rough paths for $q > p$. However, in the case $p = 1$ we identify a Polish topology generated by an explicit metric between specific representatives of the equivalence classes. We also prove that the quotient topology inherited from the underlying p -variation topology is not metrisable.

[Chapter 3](#) considers an important example of a function on unparameterised path space: the terminal time solution of the controlled differential equation

$$dZ_t^\gamma = \sum_{j=1}^d Z_t^\gamma A_j d\gamma_t^j, \quad \text{where } Z_0^\gamma = \text{id},$$

where both Z^γ and each A_j are $\mathbb{C}^{N \times N}$ -valued. When the matrices A_j are constrained to live in a Lie algebra $\mathfrak{g}_N$, the solution Z^γ is called the *development* of γ into the Lie group $G_N = \exp(\mathfrak{g}_N)$. When the matrices A_j are also randomised according to some distribution, we term Z^γ a randomised path development into G_N . A recent result of [\[MLS23\]](#) showed that when the entries of each A_j are i.i.d. $\mathcal{N}(0, \frac{1}{N})$ random variables, the inner product $\frac{1}{N} \langle Z_T^\gamma, Z_T^\sigma \rangle$ converges in L^2 to the inner product of their signatures as $N \rightarrow \infty$. By leveraging techniques from random matrix theory we show that the Gaussian and identical distribution assumptions may be relaxed in favour of mild moment conditions, and that independence between certain entries may be replaced by a zero covariance assumption. Motivated by the use of randomised path developments as a discriminator between probability measures on path space, we prove that the scaling limit of suitably randomised unitary path developments solves a novel quadratic functional equation. The derivation of this equation relies on the Schwinger-Dyson equations arising from mathematical physics and non-commutative probability theory. We also study various extensions concerning sparsity of the driving matrices and randomisation of these matrices by perturbations of the Gaussian potential on the space of Hermitian matrices.

In [Chapter 4](#) we examine another function on unparameterised path space that has recently received extensive study. Letting $\mathcal{S}(\gamma)^n$ denote the n^{th} term of the signature of γ , we study the limit $\lim_{n \rightarrow \infty} \frac{n! \|\mathcal{S}(\gamma)^n\|}{\text{Length}(\gamma)^n}$. The study of this limit sits within a broader literature that attempts to recover properties of the representative of each equivalence class of minimal length – called the tree-reduced representative – from its signature. For example, if the norm $\|\cdot\|$ is in a class of admissible tensor norms, then whenever the tree-reduced representative is C^1 (when parameterised at unit speed) the limit is equal to 1. Whether this holds for all equivalence classes is

an open question, known as the length conjecture. We are interested in a variation where the tensor norms are instead given by the Hilbert-Schmidt norm. In this case if γ is of unit length and C^3 when parameterised at unit speed, the limit was shown in [HL10] to be given by

$$\lim_{n \rightarrow \infty} n! \|\mathcal{S}(\gamma)^n\| = \mathbb{E} \left[\exp \left(\int_0^1 (B_{0,s})^2 \langle \gamma'_s, \gamma_s''' \rangle ds \right) \right]^{1/2},$$

where B is a Brownian bridge. We prove that the C^3 assumption may be relaxed to C^2 with some restrictions by showing that the left-hand-side may be expressed as a sequence of sums of Wasserstein distances between copies of a specific measure related to the path γ . The right-hand side follows by applying results of [DBGU05] that relate the asymptotics of these Wasserstein distances to the fluctuations of Brownian bridges.

The final chapter of this thesis is less directly related to the others and concerns the task of generating faithful samples from the joint distribution of a Brownian increment and its Lévy area. Typically such estimators fall into two camps: truncated Fourier series expansions of Brownian motion, or custom random variables that attempt to match the low order moments of the joint distribution. We introduce a third option in Chapter 5 that we call LévyGAN. We leverage the power of generative modelling to learn the joint distribution without requiring access to any approximate sample from the distribution. To do this, we first prove that the joint distribution of a Brownian increment and its Lévy area is a fixed point of an explicit map on the space of distributions. We then construct a generative adversarial network (GAN) that attempts to minimise the distance between successive applications of this map. The resulting generator helps bridge the gap between the cheap-to-sample moment matching estimators and the expensive-to-sample truncated Fourier series. We finally show that estimators such as LévyGAN have applications to multilevel Monte Carlo (MLMC) simulation of high order SDE solvers that use approximate samples of Lévy area. Indeed, the coupling performed at each level of MLMC with Lévy area requires that the Lévy area samples satisfy the fixed point relation that our generator is trained to minimise.

While most chapters share a common theme, they cover distinct topics that have been separately published or released as preprints. Most chapters are light edits of their published or preprint version, but we have unified the notation across the thesis to allow it to be read in its entirety. At the start of each chapter we include a brief introduction and a comment on how the chapter falls under the umbrella

of *functions on unparameterised path space*. In [Chapter 1](#) we rigorously introduce unparameterised path space, signatures, tensor algebras and rough paths. Other than unparameterised path space, we do not attempt to provide a comprehensive overview of each topic and introduce only the notions that are required for this thesis.

– CONTENTS –

List of Figures	14
List of Tables	15
1 Signatures, rough paths and tensor algebras	16
1.0.1 Tensor algebras and group-like elements	21
1.0.2 Rough paths and signatures	23
1.1 Unparameterised path space	31
1.1.1 Avoiding the quotient	34
1.2 Point set topology	34
2 Topologies on unparameterised path space	36
2.0.1 Candidate topologies	39
2.1 Unparameterised paths of bounded variation	41
2.1.1 Basic topological properties	44
2.1.2 Complete metrisability of candidate topologies?	46
2.2 Uniform approximation and the expected signature model	55
2.2.1 Compactness	56
2.2.2 Revisiting the expected signature model	59
2.3 A comment on open problems	62
2.4 Topologies on unparameterised rough path space	63
2.4.1 Properties of topologies on unparameterised rough path space	70
2.4.2 The totality of unparameterised rough paths	75
3 Randomised path developments	77
3.0.1 Outline	83
3.0.2 The signature kernel	84
3.0.3 Letters, words and graphs	85
3.0.4 Non-commutative laws and Schwinger-Dyson equations	86
3.1 The signature kernel is a universal limit	90

3.2	Universal limit of Wigner unitary developments	98
3.2.1	Non-crossing partitions, Dyck words and trees	100
3.2.2	A limiting functional equation	103
3.3	Numerical Schemes	109
3.3.1	Explicit Scheme	109
3.3.2	Implicit Scheme	113
3.3.3	Comparing Numerical Solvers	115
3.4	Pauli string path developments	117
3.5	Functional equation for interacting matrix models	132
4	Signature asymptotics and Wasserstein distances	137
4.0.1	Contributions	139
4.1	The Hambly-Lyons limit and Wasserstein distances	141
4.1.1	The B-G-U theorem	141
4.1.2	Existence and characterisation of admissible measures . .	144
4.2	Signature projections and Wasserstein distances	148
4.2.1	A probabilistic expression for signature inner products . .	149
4.2.2	Lower bound on signature norms in terms of Wasserstein distances	152
4.2.3	Upper bound on signature norms in terms of Wasserstein distances	153
4.2.4	Generalising the Hambly-Lyons Limit theorem	154
4.2.5	Computing the Hambly-Lyons limit explicitly	157
5	Generative modelling of Lévy area	160
5.0.1	Contributions	163
5.0.2	Outline	164
5.1	The GAN architecture	166
5.2	Generator	167
5.2.1	Bridge-Flipping	168
5.2.2	The pair-net generator	172
5.3	Discriminator	174
5.3.1	Unitary characteristic function	175
5.3.2	Chen training	177
5.4	LévyGAN	183
5.5	Numerical experiments	185
5.5.1	SDE example	187
5.5.2	Numerical results	188
5.6	Conclusion and open directions	192

Bibliography	194
Appendices	208
A Technical results in random matrix theory	208
A.1 Sub-Gaussian tail of the GUE	208
B Appendix to LévyGAN	212
B.1 Some properties of the characteristic function distance	212
B.2 Characteristic function and joint moments of a Brownian increment and Lévy area	214
B.3 Test metrics	217
B.4 Training procedure and hyperparameter tuning	218
B.5 Foster’s moment-matching method	219
B.6 Numerical methods for SDEs	220
B.6.1 Weak vs strong error	220
B.6.2 Discretisation schemes	221
B.6.3 Antithetic MLMC	222
B.6.4 The log-Heston SDE	223

– LIST OF FIGURES –

3.1	Convergence of numerical schemes for the Schwinger-Dyson kernel .	116
3.2	Computational time versus path length and dimension for numerical solvers of the Schwinger-Dyson kernel	117
3.3	Computational time versus hidden dimension for the Monte-Carlo estimator of the Schwinger-Dyson kernel	117
5.1	Lévy area	162
5.2	Schematic of traditional GAN for Lévy area generation	167
5.3	A schematic of the Pair-net architecture when $d = 3$	173
5.4	Schematic of LévyGAN	184
5.5	Multilevel variance and empirical error	190

– LIST OF TABLES –

5.1	Marginal distribution metrics for estimators of Lévy area	186
5.2	Joint moment metrics for estimators of Lévy area	186
5.3	Approximate bias introduced in MLMC by approximate Lévy area .	191
5.4	Average time taken to attain target RMSE in MLMC	192

CHAPTER 1

SIGNATURES, ROUGH PATHS AND TENSOR ALGEBRAS

Throughout this thesis, V and W will denote finite-dimensional vector spaces equipped with inner products. A common task in applications is to model the terminal-time response of a non-linear dynamical system to a given control path. Controlled differential equations (CDEs) of the form

$$dy_t = f(y_t) d\gamma_t, \text{ started at } y_0, \tag{1.1}$$

with a non-linear map $f : W \rightarrow \mathcal{L}(V, W)$ and a control $\gamma : [0, T] \rightarrow V$, provide a natural framework for describing many real-world phenomena while also offering a setting in which these interactions can be analysed theoretically.

If γ is smooth, [Eq. \(1.1\)](#) reduces to an ordinary differential equation (ODE), but many real-world signals — such as stock prices — are far from smooth. In such cases, it is convenient to view [Eq. \(1.1\)](#) as an integral equation:

$$y_t = y_0 + \int_0^t f(y_s) d\gamma_s.$$

To see how iterated integrals naturally arise in this context, we briefly review an example from [\[CS24, Chapter 1\]](#). Assume f is linear, i.e., $f \in \mathcal{L}(W, \mathcal{L}(V, W))$, and that γ is continuously differentiable. Consider the Picard iterates

$$y_t^{(0)} = y_0, \quad y_t^{(n+1)} = y_0 + \int_0^t f(y_s^{(n)}) d\gamma_s.$$

The n -th Picard iterate can be written as

$$y_t^{(n)} = \sum_{k=0}^n \int_{0 \leq t_1 \leq \dots \leq t_k \leq t} f^{\otimes k}(y_0)(\dot{\gamma}_{t_1}, \dots, \dot{\gamma}_{t_k}) dt_1 \dots dt_k,$$

where $f^{\otimes k}(y_0) : V^k \rightarrow W$ is recursively defined by

$$f^{\otimes k}(y_0)(v_1, \dots, v_k) := f\left(f^{\otimes(k-1)}(y_0)(v_1, \dots, v_{k-1})\right)(v_k), \quad f^{\otimes 0}(y_0) = \text{id}_W.$$

By the universal property of the tensor product, there exists a unique linear map $\widetilde{f^{\otimes k}(y_0)} : V^{\otimes k} \rightarrow W$ such that

$$f^{\otimes k}(y_0)(v_1, \dots, v_k) = \widetilde{f^{\otimes k}(y_0)}(v_1 \otimes \dots \otimes v_k).$$

Hence, using the linearity of the integral,

$$y_t^{(n)} = \sum_{k=0}^n \widetilde{f^{\otimes k}(y_0)} \left(\int_{0 \leq t_1 \leq \dots \leq t_k \leq t} \dot{\gamma}_{t_1} \otimes \dots \otimes \dot{\gamma}_{t_k} dt_1 \dots dt_k \right).$$

Define

$$\mathcal{S}(\gamma)^k := \int_{0 \leq t_1 \leq \dots \leq t_k \leq T} \dot{\gamma}_{t_1} \otimes \dots \otimes \dot{\gamma}_{t_k} dt_1 \dots dt_k \in V^{\otimes k},$$

then the *signature* of γ is given by

$$\mathcal{S}(\gamma) := (1, \mathcal{S}(\gamma)^1, \mathcal{S}(\gamma)^2, \dots) \in T((V)) := \prod_{k=0}^{\infty} V^{\otimes k}.$$

Suppose for simplicity we are interested in $\langle u, y_T \rangle$ for some $u \in W$. Setting

$$\ell_k := \langle u, \widetilde{f^{\otimes k}(y_0)}(\cdot) \rangle \in (V^{\otimes k})^*,$$

and

$$\ell^{(n)} := (\ell_0, \dots, \ell_n, 0, \dots) \in T(V^*) \cong \bigoplus_{k=0}^{\infty} (V^{\otimes k})^* \subset T((V))^*,$$

we see that

$$\langle u, y_T \rangle = \lim_{n \rightarrow \infty} \ell^{(n)}(\mathcal{S}(\gamma)). \quad (1.2)$$

That is, the terminal-time solution of a (linear) CDE can be expressed as the pointwise limit of linear functionals on the signature. Under suitable regularity assumptions, the same holds for a broad class of non-linear CDEs. Consequently, studying the CDE in [Eq. \(1.1\)](#) can be reduced to studying linear functionals on the signature. This perspective is further validated by the following result.

Lemma 1.0.1. *Let*

$$\Lambda := \{\Phi_\ell : C^1 \rightarrow \mathbb{R} \mid \Phi_\ell(\gamma) := \ell(\mathcal{S}(\gamma)), \ell \in T(V^*)\}.$$

Then Λ is an algebra.

By an application of the Stone-Weierstrass theorem, one might hope that Λ is dense in $C(K, \mathbb{R})$ in the uniform topology for compact $K \subset C^1$. If true, learning a non-linear dynamical system could be reduced to linear regression on the signature.

However, several gaps remain:

- 1) The assumption $\gamma \in C^1$ is too restrictive. In fact, [Eq. \(1.2\)](#) actually holds for all continuous paths of finite length – and even this assumption may be further relaxed.
- 2) For Λ to be dense in $C(K, \mathbb{R})$, it must first be contained in $C(K, \mathbb{R})$. We will see below that equipping K with the uniform topology is insufficient.
- 3) The signature is not injective; it is invariant under “tree-like equivalence”, which is a generalised notion of reparameterisation. Any function not invariant under reparameterisation cannot then be well approximated by a linear functional on the signature.

We now demonstrate the difficulty of defining a topology on path space which makes $\gamma \mapsto \int d\gamma \otimes d\gamma$ continuous. Let $\gamma, \sigma : [0, 1] \rightarrow \mathbb{R}$ be smooth. Standard Riemann-Stieltjes integration gives

$$\int_0^t \gamma_s d\sigma_s = \lim_{|\mathcal{P}| \rightarrow 0} \sum_{t_i \in \mathcal{P}} \gamma_{t_i} (\sigma_{t_{i+1}} - \sigma_{t_i}).$$

If $\gamma^n \rightarrow \gamma$ and $\sigma^n \rightarrow \sigma$ uniformly, one may hope to be able to define the integral of γ against σ by

$$\int_0^t \gamma_s d\sigma_s := \lim_{n \rightarrow \infty} \int_0^t \gamma_s^n d\sigma_s^n.$$

However, the following example shows that this fails.

Example 1.0.2. *Let $\sigma_t \equiv 0$ and define*

$$\gamma_t^n = \frac{1}{n\sqrt{\pi}} \cos(2\pi n^2 t), \quad \sigma_t^n = \frac{1}{n\sqrt{\pi}} \sin(2\pi n^2 t).$$

Then $\gamma^n, \sigma^n \rightarrow o$ uniformly, but

$$\int_0^t \gamma_s^n d\sigma_s^n \rightarrow t \neq 0.$$

Thus, integration is not continuous in the uniform topology. Any theory requiring continuity of integration must use a stronger topology, and we will use the p -variation topology. Once the path space is appropriately defined, the non-injectivity of the signature must also be addressed.

Since we are interested in the response of dynamical systems to a control path, it is natural to quotient path space by all equivalence relations that leave the system response unchanged. For CDEs the signature fully characterises this response, and so we define the space of *unparameterised paths* as

$$\mathcal{C} := \{\gamma : \gamma \text{ is continuous of finite length}\} / \sim, \quad \gamma \sim \sigma \iff \mathcal{S}(\gamma) = \mathcal{S}(\sigma).$$

Equipping $\mathcal{C}$ with a topology that makes the signature continuous, we obtain the following universal approximation theorem.

Theorem 1.0.3 (Uniform Approximation by the Signature). *Let $(\mathcal{C}, \chi)$ be the space of unparameterised paths equipped with a topology χ . Assume that K is a compact subset of $\mathcal{C}$ and for $\ell \in T(V^*)$ denote by Φ_ℓ the function given by*

$$\begin{aligned} \Phi_\ell &: \mathcal{C} \rightarrow \mathbb{R}; \\ [\gamma] &\mapsto \ell \circ \mathcal{S}([\gamma]). \end{aligned}$$

Let $C(K)$ denote the space of continuous functions on K with the topology of uniform convergence. If the set $\Phi_K := \{\Phi_\ell : \ell \in T(V^*)\}$ is contained in $C(K)$, then it is a dense subset.

The preceding approximation result has appeared in various guises over the years, perhaps most prominently in [LLN16]. This thesis primarily addresses the following questions:

- 1) What topology χ should be placed on the space of unparameterised paths?
- 2) What functions are there on $\mathcal{C}$ that are either of practical relevance, or help us recover geometric information about certain representatives of the equivalence classes $[\gamma]$?

Paths of infinite length

Real-world signals are often highly irregular, and the assumption of finite length breaks down. Returning to [Example 1.0.2](#), we see a simple case of how such paths arise. Consider the two dimensional path $x^n = (\gamma^n, \sigma^n)$, where γ^n and σ^n are as defined in [Example 1.0.2](#). For each n , x^n winds round a circle of radius $\frac{1}{n\sqrt{\pi}}$ a total of n^2 times over the time interval $[0, 1]$. As such, the length of x_n diverges to ∞ while its supremum norm $\|x^n\|_\infty$ converges to 0. Finally, over the course of its winding, x^n always accumulates a constant area of 1. It is for this reason that the integral $\int \gamma^n d\sigma^n$ does not converge to zero; this indicates the importance of keeping track of area terms for highly oscillatory paths. The theory of rough paths [[Lyo98](#)] allows us to deal with the case of a path X_t evolving in V which only has finite p -variation for some $p \geq 2$. For the case $p \in [2, 3)$, one must first postulate the value of the iterated integrals

$$\mathbb{X}_{s,t}^2 := \int_{s \leq t_1 \leq t_2 \leq t} dX_{t_1} \otimes dX_{t_2} \in V^{\otimes 2}, \quad s \leq t \in [0, T]$$

before one is able to give a proper meaning to (rough) differential equations of the form

$$dY_t = f(Y_t) dX_t, \quad \text{started at } Y_0 = y.$$

While there is some flexibility in the choice of $\mathbb{X}$, it must obey certain algebraic and analytic constraints. The algebraic condition is commonly known as Chen's identity. It turns out that the more irregular the underlying path, the more iterated integrals must be postulated. Once sufficiently many iterated integrals have been postulated, one is able to define iterated integrals of any order, i.e. define

$$\mathbb{X}_{s,t}^n := \int_{s \leq t_1 \leq \dots \leq t_n \leq t} dX_{t_1} \otimes \dots \otimes dX_{t_n} \in (\mathbb{R}^d)^{\otimes n}$$

for any n . The sequence of all iterated integrals $\mathbb{X}^{<\infty} := (1, \mathbb{X}^1, \mathbb{X}^2 \dots)$ is then the signature of the rough path $\mathbb{X}$. The framework of unparameterised path space and uniform approximation naturally extends to the case of rough paths. While we will not require the full machinery of rough paths or the general solution theory to rough differential equations (RDEs), we will investigate how one may choose to topologise *unparameterised rough path space*. The next section gives a brief introduction into signatures, rough paths and the required algebraic structures. We will emphasise the definitions and properties required for this thesis and refer

the reader to [Lyo98; FV10; FH14; CS24] for excellent comprehensive introductions to rough path theory and signatures.

1.0.1 Tensor algebras and group-like elements

As previously mentioned, V will denote a real finite dimensional vector space of dimension $d \geq 2$ equipped with an inner product $\langle \cdot, \cdot \rangle$, its associated norm $|\cdot|$ and an orthonormal basis $e_1, \dots, e_d$. Let the tensor powers of V , $(V^{\otimes n}, n \geq 1)$ be equipped with a family of admissible norms $\|\cdot\|_{V^{\otimes n}}$ that satisfies the following.

- 1) For all $m, n \in \mathbb{N}$ and $u \in V^{\otimes m}, v \in V^{\otimes n}$ it holds that

$$\|u \otimes v\|_{V^{\otimes(m+n)}} \leq \|u\|_{V^{\otimes m}} \|v\|_{V^{\otimes n}};$$

- 2) For any permutation $\sigma \in \Sigma_n$ and $u_1, \dots, u_n \in V$

$$\|u_1 \otimes \dots \otimes u_n\|_{V^{\otimes n}} = \|u_{\sigma(1)} \otimes \dots \otimes u_{\sigma(n)}\|_{V^{\otimes n}}.$$

We will write $\langle \cdot, \cdot \rangle := \langle \cdot, \cdot \rangle_{V^{\otimes n}}$ for the Hilbert-Schmidt inner product on $V^{\otimes n}$ derived from the inner product on V .

Definition 1.0.4 (Tensor algebras). *Define the tensor and extended tensor algebra over V respectively as*

$$T(V) := \bigoplus_{n=0}^{\infty} V^{\otimes n};$$

$$T((V)) := \prod_{n=0}^{\infty} V^{\otimes n}.$$

We endow both $T(V)$ and $T((V))$ with a pair of binary operations: addition and multiplication. For $\mathbf{x} = (x_0, x_1, \dots)$ and $\mathbf{y} = (y_0, y_1, \dots)$ we define

$$\mathbf{x} + \mathbf{y} := (x_0 + y_0, x_1 + y_1, x_2 + y_2, \dots)$$

$$\mathbf{x} \otimes \mathbf{y} := (z_0, z_1, z_2, \dots),$$

where

$$z_n := \sum_{k=0}^n x_k \otimes y_{n-k}.$$

Let $\mathbf{1} := (1, 0, 0, \dots)$ be the identity element of both $T(V)$ and $T((V))$ under the tensor-product.

Any element of $T((V))$ can be written in the form

$$\mathbf{x} = \sum_{\mathbf{w} \in \mathcal{W}_d} \mathbf{x}^{\mathbf{w}} e_{\mathbf{w}},$$

where $\mathcal{W}_d$ is the set of all words in d letters and $e_{\mathbf{w}} := e_{w_1} \otimes \cdots \otimes e_{w_n}$ for a word $\mathbf{w} = w_1 \dots w_n$. We will commonly write

$$\langle \mathbf{x}, \mathbf{w} \rangle := \langle \mathbf{x}, e_{\mathbf{w}} \rangle = \mathbf{x}^{\mathbf{w}}.$$

Certain finite-dimensional subspaces of the tensor algebra are important to the study of signatures.

Definition 1.0.5 (Truncated tensor algebra). *Define the truncated tensor algebra $T^{(n)}(V)$ through an associated projection $\pi^{(n)}$ by*

$$\pi^{(n)} : T((V)) \rightarrow T^{(n)}(V) := \prod_{i=0}^n V^{\otimes i}.$$

We also define the projection onto each level of the tensor algebra as $\pi_n : T((V)) \rightarrow V^{\otimes n}$. Let $\iota^{(n)} : T^{(n)}(V) \hookrightarrow T((V))$ denote the inclusion map $T^{(n)}(V) \hookrightarrow T((V))$.

In the following, we require the subsets of $T((V))$ given by

$$T_i((V)) := \{\mathbf{x} \in T((V)) : x_0 = i\} \quad \text{for } i = 0, 1.$$

Definition 1.0.6 (Functions on tensor algebra). *Define the exponential $\exp$ as*

$$\begin{aligned} \exp : T((V)) &\rightarrow T_1((V)) \\ \mathbf{x} &\mapsto \sum_{n=0}^{\infty} \frac{\mathbf{x}^{\otimes n}}{n!}. \end{aligned}$$

The logarithm of an element of $T_1((V))$ is defined as

$$\log(\mathbf{x}) := \log(x_0) \mathbf{1} + \sum_{n=1}^{\infty} \frac{(-1)^n}{n} \left(\mathbf{1} - \frac{\mathbf{x}}{x_0} \right)^{\otimes n}.$$

Finally the inverse of $\mathbf{x} \in T((V))$ is given by

$$\mathbf{x}^{-1} := \frac{1}{x_0} \sum_{n=0}^{\infty} \left(\mathbf{1} - \frac{\mathbf{x}}{x_0} \right)^{\otimes n}. \tag{1.3}$$

Remark 1.0.7. *The above definitions of the exponential, logarithm and inverse are defined purely algebraically, and it can be checked that each series is locally finite.*

We summarise various properties of these operations below; see [Lyo+07, Chapter 2] for details.

Proposition 1.0.8. *The maps $\exp : T_0((V)) \rightarrow T_1((V))$ and $\log : T_1((V)) \rightarrow T_0((V))$ are inverses. Moreover, the map given by Eq. (1.3) is the inverse of the tensor-product on the space $T((V)) \setminus T_0((V))$.*

The subspace of $T^{(n)}(V)$ where truncated signatures will reside is of equal importance.

Definition 1.0.9 (Step- n free Lie algebra and free nilpotent Lie group). *Consider the commutator $[\mathbf{x}, \mathbf{y}] = \mathbf{x} \otimes \mathbf{y} - \mathbf{y} \otimes \mathbf{x}$. For subspaces $X, Y \subseteq T((V))$ we define*

$$[X, Y] := \text{span}\{[\mathbf{x}, \mathbf{y}] : \mathbf{x} \in X, \mathbf{y} \in Y\}.$$

The step- n free Lie algebra $\mathfrak{g}^{(n)}$ is the smallest sub-algebra of $T^{(n)}(V)$ that contains V and is closed under the commutator $[\cdot, \cdot]$. That is

$$\mathfrak{g}^{(n)} := V \oplus [V, V] \oplus \cdots \oplus \underbrace{[V, [\dots, [V, V]]]}_{n-1 \text{ brackets}}.$$

The step- n free nilpotent Lie group is then given by

$$G^{(n)} := \pi^{(n)} \circ \exp \circ \iota^{(n)}(\mathfrak{g}^{(n)}).$$

Note that $\mathfrak{g}^{(1)} \equiv G^{(1)} \equiv V$. The above definitions can be extended to $T((V))$ to define the free Lie algebra $\mathfrak{g}^{(*)}$ and the free Lie group $G^{(*)}$.

Definition 1.0.10 (Free Lie algebra and free Lie group). *Define the free Lie algebra as*

$$\mathfrak{g}^{(*)} := \{\mathbf{x} \in T((V)) : \pi^{(n)}(\mathbf{x}) \in \mathfrak{g}^{(n)} \text{ for all } n \geq 0\},$$

and the free Lie group by

$$G^{(*)} := \exp(\mathfrak{g}^{(*)}).$$

From now on, to ease notation, we will write $\mathbf{xy} := \mathbf{x} \otimes \mathbf{y}$.

1.0.2 Rough paths and signatures

Central to the concept of rough paths is the notion of p -variation.

Definition 1.0.11. Let (E, ρ) be a metric space, then we say that $\gamma : [0, T] \rightarrow E$ has finite p -variation for $p \geq 1$ if

$$\|\gamma\|_{p\text{-var};[0,T]} := \sup_{\mathcal{P}} \left(\sum_{t_i \in \mathcal{P}} \rho(\gamma_{t_i}, \gamma_{t_{i+1}})^p \right)^{\frac{1}{p}} < \infty, \quad (1.4)$$

where $\mathcal{P}$ is a partition of $[0, T]$. When γ possesses finite 1-variation, we say that γ is of finite, or bounded, variation. For paths of finite variation evolving in V , the iterated integrals of γ are well defined as Young-integrals. The collection of all iterated integrals of γ is called its *signature*.

Definition 1.0.12 (Continuous paths of bounded variation). Let $C^{1\text{-var}}([0, T], V)$ be the space of continuous paths $\gamma : [0, T] \rightarrow V$ for which $\|\gamma\|_{1\text{-var};[0,T]} < \infty$ with $\gamma_0 = 0$.

Definition 1.0.13 (Time-Simplex). For any $n \in \mathbb{N}$, define the n -simplex over an interval $[a, b]$ by

$$\Delta_{[a,b]}^n := \{(t_1, \dots, t_n) : a \leq t_1 \leq \dots \leq t_n \leq b\}.$$

Definition 1.0.14 (The signature). Let $\gamma \in C^{1\text{-var}}([0, T], V)$ and define the signature map

$$\mathcal{S}_{\cdot, \cdot}(\gamma) : \Delta_{[0,T]}^2 \rightarrow T((V)),$$

where $\mathcal{S}_{s,s}(\gamma) = (1, 0, \dots)$ for all $s \in [0, T]$ and

$$\mathcal{S}_{s,t}(\gamma)^n := \int_{\Delta_{[s,t]}^n} d\gamma_{t_1} \otimes \dots \otimes d\gamma_{t_n} \in V^{\otimes n},$$

for $n \geq 1$ with $\mathcal{S}_{s,t}(\gamma)^0 := 1 \in \mathbb{R}$. For a word $\mathbf{w} = w_1 \dots w_n \in \mathcal{W}_d$ with $|\mathbf{w}| = n$, we also define the coordinate iterated integral

$$\mathcal{S}_{s,t}^{\mathbf{w}}(\gamma) := \int_{\Delta_{[s,t]}^n} d\gamma_{t_1}^{w_1} \dots d\gamma_{t_n}^{w_n} \in \mathbb{R},$$

where $\mathcal{S}_{s,t}^{\emptyset}(\gamma) := \mathcal{S}_{s,t}(\gamma)^0 = 1$. The truncated signature map is defined by $\mathcal{S}_{s,t}(\gamma)^{(n)} := \pi^{(n)} \mathcal{S}_{s,t}(\gamma)$. Finally, we call $\mathcal{S}(\gamma) := \mathcal{S}_{0,T}(\gamma)$ the signature of γ .

In an applied situation dealing with streamed data, one often has to deal with the (infinite dimensional) group of symmetries known as reparameterisations. For example, the rate at which a letter is drawn will not change whether it is the letter a or the letter b . The signature has the benefit of being invariant under this group of symmetries; see [Yan+16] for the application of reparameterisation invariance to the task of recognising hand-drawn Chinese characters.

Definition 1.0.15. A function $\alpha : [0, T] \rightarrow [0, T]$ is called a reparameterisation if α is a non-decreasing surjection.

Lemma 1.0.16 (Lemma 1.2.1 [CS24]). Let α be a reparameterisation and $\gamma \in C^{1-var}([0, T], V)$, then $\mathcal{S}(\gamma) = \mathcal{S}(\gamma \circ \alpha)$.

Depending on the context, reparameterisation invariance allows the signature map to remove an infinite dimensional group of undesired symmetries. Another key property of the signature is that the product of two linear functionals on the signature is itself a linear functional. This fact is derived from the ordinary integration by parts rule satisfied by continuous paths of bounded variation. This is encoded in their signatures through the *shuffle identity*.

Definition 1.0.17. The shuffle product $\sqcup : T(V) \times T(V) \rightarrow T(V)$ is the commutative operation on tensors defined by $\mathbf{u} \sqcup \emptyset = \emptyset \sqcup \mathbf{u} = \mathbf{u}$ and the recursive definition

$$\mathbf{u}i \sqcup \mathbf{v}j = (\mathbf{u} \sqcup \mathbf{v}j)i + (\mathbf{v}i \sqcup \mathbf{u})j,$$

for all words $\mathbf{u}$ and $\mathbf{v} \in \mathcal{W}_d$ and extended to $T(V)$ by linearity. We say that an element $\mathbf{x} \in T_1((V))$ satisfies the shuffle identity if

$$\langle \ell_1, \mathbf{x} \rangle \langle \ell_2, \mathbf{x} \rangle = \langle \ell_1 \sqcup \ell_2, \mathbf{x} \rangle, \quad (1.5)$$

for every $\ell_1, \ell_2 \in T(V)$.

Proposition 1.0.18. Let $\gamma \in C^{1-var}([0, T], V)$. Then $\mathcal{S}(\gamma)$ satisfies the shuffle identity.

The signatures of continuous finite variation paths and the shuffle identity provide alternative characterisations of the step- n free nilpotent Lie group.

Proposition 1.0.19. The following spaces are identical.

- 1) $G^{(n)}$;
- 2) $\{\mathcal{S}(\gamma)^{(n)} : \gamma \in C^{1-var}([0, T], V)\}$;
- 3) $\{\mathbf{x} \in T^{(n)}(V) : \mathbf{x}^\emptyset = 1 \text{ and } \mathbf{x} \text{ satisfies the shuffle identity}\}$;
- 4) $\pi^{(n)}(G^{(*)})$.

Note that as a consequence, the projection $\pi^{(n)}(\mathbf{x})$ of any $\mathbf{x} \in G^{(*)}$ is the truncated signature of some path $\gamma \in C^{1-var}([0, T], V)$.

Remark 1.0.20. The characterisation of $G^{(n)}$ in terms of the shuffle identity extends to $G^{(*)}$, that is $\mathbf{x} \in T_1((V))$ is group-like if and only if it satisfies the shuffle identity. We refer the interested reader to [Gen] for further characterisations of group-like elements.

In order to define the space of rough paths elegantly, we will require a metric on $G^{(n)}$. An important concept is that of a symmetric, sub-additive homogenous norm.

Definition 1.0.21. A norm $\|\cdot\| : G^{(n)} \rightarrow [0, \infty)$ is symmetric, sub-additive and homogenous if

- 1) $\|\mathbf{g}\| = 0$ if and only if $\mathbf{g} = \mathbf{1} \in G^{(n)}$;
- 2) $\|\mathbf{g}\| = \|\mathbf{g}^{-1}\|$;
- 3) $\|\mathbf{gh}\| \leq \|\mathbf{g}\| + \|\mathbf{h}\|$;
- 4) $\|\delta_\lambda \mathbf{g}\| = |\lambda| \|\mathbf{g}\|$, where δ_λ is the dilation operator defined by

$$\pi_n(\delta_\lambda \mathbf{g}) = \lambda^n \pi_n(\mathbf{g}), \quad \lambda \in \mathbb{R}.$$

To every symmetric sub-additive homogeneous norm we may associate a metric. For more details and examples of homogenous norms see [FV10, Chapter 7]. We record here only the Carnot-Caratheodory norm which we require for our purposes.

Definition 1.0.22 (Carnot-Caratheodory norm). For $\mathbf{g} \in G^{(n)}$ define

$$\|\mathbf{g}\|_{\text{CC}} := \inf \{ \|\gamma\|_{1\text{-var};[0,T]} : \mathbf{g} = \mathcal{S}(\gamma)^{(n)} \}.$$

Proposition 1.0.23 (Proposition 7.40 of [FV10]). The Carnot-Caratheodory norm is symmetric, sub-additive and homogenous.

Definition 1.0.24. Let d_{CC} be the metric on $G^{(n)}$ associated with the Carnot-Caratheodory norm, which is given by

$$d_{\text{CC}}(\mathbf{g}, \mathbf{h}) := \|\mathbf{g}^{-1}\mathbf{h}\|.$$

Remark 1.0.25. Note that on $G^{(1)}$, d_{CC} reduces to the Euclidean metric.

We may now equip $G^{(n)}$ with d_{CC} and define the space of weakly geometric rough paths.

Definition 1.0.26 (Weakly geometric rough paths). We let $C^{p\text{-var}}([0, T], G^{(n)})$ be the space of continuous paths of finite p -variation which start at the unit element of $G^{(n)}$. The space of weakly geometric rough paths $W\Omega_p(V)$ is defined to be $C^{p\text{-var}}([0, T], G^{(\lfloor p \rfloor)})$

Remark 1.0.27. We will typically use the letters γ, σ to denote paths in $C^{1\text{-var}}([0, T], V)$ when we do not need to consider lifts into $C^{p\text{-var}}([0, T], G^{(n)})$. Additionally, we will use

$\mathbb{X}, \mathbb{Y}$ for paths in $C^{p\text{-var}}([0, T], G^{(n)})$ with $p > 1$ or paths in $C^{1\text{-var}}([0, T], V)$ whose lifts we will require.

An important consequence of the Carnot-Caratheodory metric is the following result.

Proposition 1.0.28. *Let $\mathbb{X}^{\leq n} \in C^{p\text{-var}}([0, T], G^{(n)})$, then the p -variation norm of $\mathbb{X}^{\leq n}$ in $C^{p\text{-var}}([0, T], G^{(n)})$ is greater than or equal to the p -variation norm of the projection $\mathbb{X}^{\leq m} \in C^{p\text{-var}}([0, T], G^{(m)})$ for all $m \leq n$.*

We now proceed to topologise the space of weakly geometric rough paths.

Remark 1.0.29. *This is straightforward in the case $n = 1$, since $C^{p\text{-var}}([0, T], V)$ is a linear space, and so we may define*

$$d_{p\text{-var}}^{(1)}(\mathbb{X}^{\leq 1}, \mathbb{Y}^{\leq 1}) := \left\| \mathbb{X}^{\leq 1} - \mathbb{Y}^{\leq 1} \right\|_{p\text{-var}; [0, T]}.$$

Since $G^{(n)}$ is not a linear space, we require a more general notion of the “increments” of a path evolving in $G^{(n)}$.

Definition 1.0.30. *For $\mathbb{X}^{\leq n} \in C^{p\text{-var}}([0, T], G^{(n)})$ we define its increments for $s, t \in \Delta_{[0, T]}^2$ by*

$$\mathbb{X}_{s, t}^{\leq n} := (\mathbb{X}_s^{\leq n})^{-1}(\mathbb{X}_t^{\leq n}).$$

Definition 1.0.31 (Inhomogenous p -variation distance). *For $\mathbb{X}^{\leq n}, \mathbb{Y}^{\leq n} \in C^{p\text{-var}}([0, T], G^{(n)})$ we define the inhomogenous p -variation distance by*

$$d_{p\text{-var}}^{(n)}(\mathbb{X}^{\leq n}, \mathbb{Y}^{\leq n}) = \max_{1 \leq i \leq n} \sup_{\mathcal{P}} \left(\sum_{t_j \in \mathcal{P}} \left\| \pi_i(\mathbb{X}_{t_j, t_{j+1}}^{\leq n} - \mathbb{Y}_{t_j, t_{j+1}}^{\leq n}) \right\|^{\frac{p}{i}} \right)^{\frac{i}{p}}. \quad (1.6)$$

Definition 1.0.32 (Homogenous p -variation distance). *For $\mathbb{X}^{\leq n}, \mathbb{Y}^{\leq n} \in C^{p\text{-var}}([0, T], G^{(n)})$ we define the homogenous p -variation distance by*

$$\tilde{d}_{p\text{-var}}^{(n)}(\mathbb{X}^{\leq n}, \mathbb{Y}^{\leq n}) = \sup_{\mathcal{P}} \left(\sum_{t_j \in \mathcal{P}} d_{\text{CC}}(\mathbb{X}_{t_j, t_{j+1}}^{\leq n}, \mathbb{Y}_{t_j, t_{j+1}}^{\leq n})^p \right)^{\frac{1}{p}}. \quad (1.7)$$

By [FV10, Theorem 8.10], both the inhomogenous and homogenous distances generate the same topology and notions of Cauchy sequences. This allows us to use the two metrics interchangeably in Chapter 2. For two weakly geometric rough paths $\mathbb{X}^{\leq [p]}, \mathbb{Y}^{\leq [p]} \in W\mathcal{G}\Omega_p(V)$, we will omit the dependence on n and simply

write $d_{p\text{-var}}(\mathbb{X}^{\leq [p]}, \mathbb{Y}^{\leq [p]}) := d_{p\text{-var}}^{([p])}(\mathbb{X}^{\leq [p]}, \mathbb{Y}^{\leq [p]})$. Additionally, when the level to which a path is defined is clear from the context, we will write $\mathbb{X} \equiv \mathbb{X}^{\leq [p]}$. The following theorem of Lyons informally states that for a continuous path of finite p -variation, if the first $[p]$ iterated integrals are provided as input (subject to certain algebraic and analytic constraints) then the remaining stack of iterated integrals is uniquely determined. This result may be found in [FV10, Theorem 9.5] for finite-dimensional V and [Cas+16, Corollary 3.9] for infinite dimensional V , see also [Lyo98, Theorem 2.2.1].

Theorem 1.0.33 (Lyons' Extension Theorem). *Let $\mathbb{X}^{\leq [p]} \in WG\Omega_p(V)$, then there exists a unique path $\mathcal{S}_{0,\cdot}(\mathbb{X}^{\leq [p]}) : [0, T] \rightarrow G^{(*)}$ such that*

- 1) $\mathcal{S}_{0,0}(\mathbb{X}^{\leq [p]}) = \mathbf{1}$;
- 2) $\pi^{(n)}(\mathcal{S}_{0,t}(\mathbb{X}^{\leq [p]}))$ defines a continuous map from $WG\Omega_p(V)$ to $C^{p\text{-var}}([0, T], G^{(n)})$;
- 3) $\pi^{([p])}(\mathcal{S}_{0,t}(\mathbb{X}^{\leq [p]})) = \mathbb{X}_t^{\leq [p]}$.

We call $\mathcal{S}(\mathbb{X}^{\leq [p]}) := \mathcal{S}_{0,T}(\mathbb{X}^{\leq [p]})$ the signature of the rough path $\mathbb{X}^{\leq [p]}$.

We note that the Lyons' lift defines a canonical mapping $WG\Omega_p(V) \hookrightarrow C^{p\text{-var}}([0, T], G^{(n)})$ for $n \geq [p]$ given by

$$\mathbb{X}^{\leq [p]} \mapsto \mathbb{X}^{\leq n} := \pi^{(n)}(\mathcal{S}_{0,\cdot}(\mathbb{X}^{\leq [p]})).$$

Additionally, we will write $\mathbb{X}^{<\infty} := \mathcal{S}_{0,T}(\mathbb{X}^{\leq [p]})$. We denote by $\mathcal{S}_p$ the image of $WG\Omega_p(V)$ in $G^{(*)}$ under the map $\mathcal{S}$ and write

$$\mathcal{S} := \bigcup_{p \geq 1} \mathcal{S}_p \subseteq G^{(*)},$$

for the space of all signatures. An important subset of weakly geometric rough paths are the *geometric* paths. This is the subset of paths of $WG\Omega_p(V)$ which may be approximated in $d_{p\text{-var}}$ by paths of finite q -variation for some $q < p$. Formally, we use the following definition.

Definition 1.0.34. *The space of geometric rough paths $G\Omega_p(V) \subset WG\Omega_p(V)$ is the set of continuous paths $\mathbb{X}^{\leq [p]} : [0, T] \rightarrow G^{([p])}$, starting at the unit, for which there exists a sequence of paths $(\mathbb{X}_n^{\leq 1})_{n=1}^\infty \subset WG\Omega_1(V)$ with*

$$d_{p\text{-var}}(\mathbb{X}_n^{\leq [p]}, \mathbb{X}^{\leq [p]}) \rightarrow 0.$$

Remark 1.0.35. Clearly in the case $p = 1$, we have by definition that $WG\Omega_1(V) = G\Omega_1(V)$. However, the equivalent notion of “geometric” rough paths is the space $C^{0,1-var}([0, T], V)$ of absolutely continuous paths, given by the closure of smooth paths in the 1-variation topology.

We now record several standard (but important) facts concerning the spaces $G\Omega_p(V)$, $WG\Omega_p(V)$ and their signatures. First, we note that $G\Omega_p(V)$ is a closed and separable subset of $WG\Omega_p(V)$, see [FV10, Proposition 8.25, Corollary 1.35] for a proof.

Lemma 1.0.36. The spaces $G\Omega_p(V)$ for $p > 1$, and $C^{0,1-var}([0, T], V)$ for $p = 1$, are Polish spaces with the lifts of piecewise linear paths forming a dense subset.

Lemma 1.0.37 (Properties of the p -variation metric). Fix $q \geq p$.

1) Suppose $\mathbb{X}^{\leq n}, \mathbb{Y}^{\leq n} \in C^{p-var}([0, T], G^{(n)})$, then

$$d_{q-var}^{(n)}(\mathbb{X}^{\leq n}, \mathbb{Y}^{\leq n}) \leq d_{p-var}^{(n)}(\mathbb{X}^{\leq n}, \mathbb{Y}^{\leq n});$$

2) If $\mathbb{X}_n^{\leq [p]} \rightarrow \mathbb{X}^{\leq [p]}$ in $WG\Omega_p(V)$, then $\mathbb{X}_n^{\leq [q]} \rightarrow \mathbb{X}^{\leq [q]}$ in $WG\Omega_q(V)$;

3) We have the strict inclusions $G\Omega_p(V) \subset WG\Omega_p(V) \subset G\Omega_q(V)$.

Proof. The first item is a consequence of the definition of Eq. (1.6) and the inequality

$$\left(\sum |a_j|^{\frac{q}{i}} \right)^{\frac{i}{q}} \leq \left(\sum |a_j|^{\frac{p}{i}} \right)^{\frac{i}{p}}.$$

For the second item, Lyons’ lift defines a continuous map from $WG\Omega_p(V) \hookrightarrow C^{p-var}([0, T], G^{(n)})$ for $n \geq [p]$, see [FV10, Corollary 9.11]. The claimed convergence is a consequence of this continuity and the first item. The final item is standard; see [FV10, Corollary 8.24]. $\square$

Definition 1.0.38. For every $\mathbb{X} \in WG\Omega_p(V)$, we define the control $\omega_{\mathbb{X},p}(s, t) := \|\mathbb{X}\|_{p-var;[s,t]}^p$. Let

$$\varphi(t) := \frac{\omega_{\mathbb{X},p}(0, t)}{\omega_{\mathbb{X},p}(0, T)} T,$$

and $\tilde{\mathbb{X}}$ the path implicitly defined by the relation $\mathbb{X} = \tilde{\mathbb{X}} \circ \varphi$. We call $\tilde{\mathbb{X}}$ the “Hölder-control reparameterisation” of $\mathbb{X}$.

Proposition 1.0.39. For any $\mathbb{X} \in WG\Omega_p(V)$, its Hölder-control reparameterisation $\tilde{\mathbb{X}}$ satisfies

$$1) \quad \|\tilde{\mathbb{X}}\|_{p-var;[0,t]}^p = \frac{t}{T} \|\tilde{\mathbb{X}}\|_{p-var;[0,T]}^p.$$

2) $\tilde{\mathbb{X}}$ is $\frac{1}{p}$ -Hölder continuous with $\|\tilde{\mathbb{X}}\|_{\frac{1}{p}\text{-Höl}} \leq \frac{1}{T^{1/p}} \|\tilde{\mathbb{X}}\|_{p\text{-var};[0,T]}$.

Proof. Both points are well-known and can be found in [FV10, Proposition 5.14], [Lyo+07] or [Che16, Lemma 3.2.2]. $\square$

Remark 1.0.40. Note that if $\mathbb{X} \in WG\Omega_p(V)$ satisfies $\|\mathbb{X}\|_{p\text{-var};[0,t]}^p = \frac{t}{T} \|\mathbb{X}\|_{p\text{-var};[0,T]}$, then $\mathbb{X} = \tilde{\mathbb{X}}$ since $\phi(\cdot)$ will be the identity.

Remark 1.0.41. In the case $p = 1$, Hölder-control parameterisation is equivalent to constant-speed parameterisation, i.e. $|\gamma'_t| = \frac{1}{T} \|\gamma\|_{1\text{-var};[0,T]}$ almost-everywhere. Indeed, if γ has Hölder-control parameterisation then its weak derivative $\gamma' \in L^1([0, T], V)$ is uniquely defined and must satisfy

$$\frac{t}{T} \|\gamma\|_{1\text{-var};[0,T]} = \|\gamma\|_{1\text{-var};[0,t]} = \int_0^t |\gamma'_u| du.$$

Hence $|\gamma'_u| = \frac{1}{T} \|\gamma\|_{1\text{-var};[0,T]}$ almost-everywhere by the Lebesgue Differentiation Theorem.

An important concept in the study of signatures is that of path-concatenation and reversal.

Definition 1.0.42. Given $\mathbb{X}, \mathbb{Y} \in WG\Omega_p(V)$, we define their concatenation as

$$(\mathbb{X} \star \mathbb{Y})(t) := \begin{cases} \mathbb{X}(\frac{2t}{T}), & t \in [0, \frac{T}{2}], \\ \mathbb{X}(T) \mathbb{Y}(\frac{2t-T}{2}), & t \in [\frac{T}{2}, T], \end{cases}$$

and the reversal of a path by

$$\overleftarrow{\mathbb{X}}_t := (\mathbb{X}_T)^{-1} \mathbb{X}_{T-t}.$$

The following result is commonly known as Chen's identity [Che54] and states that the set of signatures is closed under multiplication and inverses. We will denote by o the constant path in $WG\Omega_p(V)$ for which $o(t) = \mathbf{1}$ for all $t \in [0, T]$.

Proposition 1.0.43. For $\mathbb{X}, \mathbb{Y} \in WG\Omega_p(V)$, the signature map satisfies

1) *Chen's identity:* $(\mathbb{X} \star \mathbb{Y})^{<\infty} = \mathbb{X}^{<\infty} \mathbb{Y}^{<\infty}$;

2) *Inverses:* $\mathbb{X}^{<\infty} \overleftarrow{\mathbb{X}}^{<\infty} = o^{<\infty} = \mathbf{1}$.

The following is a well-known fact, whose proof follows essentially the same proof of the fact that $C^{p\text{-var}}([0, T], V)$ is a Banach space under the p -variation norm.

Lemma 1.0.44. Fix $p \geq 1$, and let $\mathbb{X}, \mathbb{Y} \in WG\Omega_p(V)$ then

$$\|\mathbb{X} \star \mathbb{Y}\|_{p\text{-var}} \leq \|\mathbb{X}\|_{p\text{-var}} + \|\mathbb{Y}\|_{p\text{-var}}.$$

Proof. Fix $\mathcal{P}$ to be a partition of $[0, T]$ and let $t_j \in \mathcal{P}$ be such that $\frac{1}{2} \in [t_j, t_{j+1}]$. Notice that

$$d((\mathbb{X} \star \mathbb{Y})_{t_j}, (\mathbb{X} \star \mathbb{Y})_{t_{j+1}}) = d(\mathbb{X}_{2t_j}, \mathbb{X}_T \mathbb{Y}_{2t_{j+1}-T}) \leq d(\mathbb{X}_{2t_j}, \mathbb{X}_T) + d(\mathbb{Y}_0, \mathbb{Y}_{2t_{j+1}-T}), \quad (1.8)$$

where the inequality follows from the triangle inequality and the definition of concatenation. Then

$$\begin{aligned} \left(\sum_{t_i \in \mathcal{P}} d((\mathbb{X} \star \mathbb{Y})_{t_i}, (\mathbb{X} \star \mathbb{Y})_{t_{i+1}})^p \right)^{\frac{1}{p}} &\leq \left(\sum_{t_i < t_j} d(\mathbb{X}_{2t_i}, \mathbb{X}_{2t_{i+1}})^p + d(\mathbb{X}_{2t_j}, \mathbb{X}_T)^p \right)^{\frac{1}{p}} \\ &\quad + \left(d(\mathbb{Y}_0, \mathbb{Y}_{2t_{j+1}-T})^p + \sum_{t_i > t_j} d(\mathbb{Y}_{2t_i-T}, \mathbb{Y}_{2t_{i+1}-T})^p \right)^{\frac{1}{p}} \\ &\leq \|\mathbb{X}\|_{\text{p-var}} + \|\mathbb{Y}\|_{\text{p-var}}, \end{aligned}$$

where the first inequality follows from Eq. (1.8) and the sub-additivity of ℓ_p norms. The second inequality follows from taking the supremum over partitions of the right-hand side. We conclude by taking supremums over partitions $\mathcal{P}$ on the left-hand-side. $\square$

1.1 Unparameterised path space

An important challenge in machine learning is the task of learning some continuous function $f : K \rightarrow \mathbb{R}$, where K is the domain in which your data lives. Typically, one considers a parametric family of models $\mathcal{X} = \{f_\theta : \theta \in \Theta\}$ and then one attempts to find

$$\hat{\theta} = \operatorname{argmin}_{\theta \in \Theta} \mathcal{L}(f_\theta, f),$$

where $\mathcal{L}$ is some loss function. While there may be no true minimiser $\hat{\theta}$, one might hope that for each $\varepsilon > 0$ one can find some $\theta_\varepsilon \in \Theta$ for which $\mathcal{L}(f_{\theta_\varepsilon}, f) < \varepsilon$. The challenge then is two-fold: using a suitable family of models $\{f_\theta : \theta \in \Theta\}$ that is capable of approximating f , and then efficiently find a good approximator within the space of models. The archetypal example of this is where $K \subset \mathbb{R}^n$ is a compact set and $\mathcal{L}(f, g) = \sup_{x \in K} |f(x) - g(x)|$. Defining

$$\Phi(x) = (1, x, \frac{x^{\otimes 2}}{2!}, \frac{x^{\otimes 3}}{3!}, \dots) \in T((V)),$$

then the family $\mathcal{X} = \{\langle \Phi(\cdot), \mathbf{x} \rangle : \mathbf{x} \in T(V)\}$ is simply the space of polynomials from $\mathbb{R}^n$ to $\mathbb{R}$ which is known to be dense in $C(K, \mathbb{R})$ by Stone-Weierstrass. An

approximator may be found simply by choosing the maximal degree of the polynomial and then applying linear regression. It is natural to ask whether this approach may be extended to the case where K is a subset of paths.

By the shuffle identity we know that the family of linear functionals on signatures of bounded variation paths ($\{\langle \mathcal{S}(\cdot), \mathbf{x} \rangle : \mathbf{x} \in T(V)\}$) is an algebra that contains the constant functions, by virtue of the fact that $\mathcal{S}(\gamma)^0 = 1$ for any path γ . In order to apply Stone-Weierstrass, one must ask the question: does the signature map separate points? Unfortunately, this is not the case. For example, for any $\mathbb{X} \neq o$ we have that $\mathbb{X} \star \overleftarrow{\mathbb{X}} \neq o$, while from [Proposition 1.0.43](#) we have

$$\mathcal{S}(o) = \mathbf{1} = \mathcal{S}(\mathbb{X} \star \overleftarrow{\mathbb{X}}).$$

Alternatively, as we have already seen, for a reparameterisation $\alpha : [0, T] \mapsto [0, T]$ we know that $\mathcal{S}(\mathbb{X} \circ \alpha) = \mathcal{S}(\mathbb{X})$, namely the signature is invariant under reparameterisation. This means that one cannot apply the Stone-Weierstrass theorem directly to compact subsets of paths, but it is actually often desirable in application to use a model invariant to the effect of time reparameterisation. Indeed, suppose that one is trying to classify a time series of a person drawing a digit as either “3” or “not 3”. Clearly, the speed at which the digit is drawn does not impact whether or not the final digit is a 3, and so any method that removes this symmetry is desirable. This approach has proven successful in a number of settings, e.g. [\[LLN16\]](#), [\[Yan+16\]](#) in which the invariance of the signature under $\sim_\tau$ provides a form of dimensionality reduction. It is then important to understand exactly when two paths have equal signatures.

The seminal work of Hambly & Lyons [\[HL10\]](#) proves that the signature of a path $\gamma \in C^{1\text{-var}}([0, T], V)$ characterises γ up to an equivalence relation on $C^{1\text{-var}}([0, T], V)$ called tree-like equivalence (denoted by $\sim_\tau$). The signature map $\mathcal{S}$ accordingly becomes well defined and injective on the $\sim_\tau$ -equivalence classes: the space of so-called unparameterised paths $\mathcal{C}_1$. Tree-like equivalence may be thought of as a generalised notion of reparameterisation. Indeed, loosely speaking, two paths γ and σ are tree-like equivalent when the time- T solutions to all differential equations of the form

$$dy_t = f(y_t)d\gamma_t, \quad y_0 = x \quad \text{and} \quad dz_t = f(z_t)d\gamma_t, \quad z_0 = x$$

satisfy $y_T = z_T$. In this thesis, we will mainly work with the space $C^{1\text{-var}}([0, T], V)$ of continuous paths of bounded variation. However, in [Section 2.4.1](#), we will study the more general setting of $WG\Omega_p(V)$ for any $p > 1$. As such, we will provide

the notions tree-like equivalence and unparameterised path space in this general setting.

Definition 1.1.1. A path $\mathbb{X} \in W\Omega_p(V)$ is called *tree-like* if there exists a real tree $\mathcal{T}$ such that $\mathbb{X}$ admits a factorisation $\mathbb{X} = \phi \circ \rho$ through a pair of continuous maps $\rho : [0, T] \mapsto \mathcal{T}$ and $\phi : \mathcal{T} \mapsto V$, where $\rho(0) = \rho(T)$. For two paths $\mathbb{X}, \mathbb{Y} \in W\Omega_p(V)$, if $\mathbb{X} \star \overleftarrow{\mathbb{Y}}$ is tree-like, we call $\mathbb{X}$ and $\mathbb{Y}$ *tree-like equivalent* and write $\mathbb{X} \sim_\tau \mathbb{Y}$.

Remark 1.1.2. In fact, in the original paper [HL10] a different definition of tree-like paths was used which is equivalent for $C^{1-var}([0, T], V)$. However, this notion was not general enough to handle the case of weakly geometric p -rough paths later studied in [Boe+16]. We will use the general definition.

Paths devoid of tree-like pieces are called *tree-reduced* and will play a key role in Chapter 2. The formal definition of this property is given in the following.

Definition 1.1.3 ([Boe+16]). A path $\mathbb{X} \in W\Omega_p(V)$ is called *tree-reduced* if the path $t \mapsto \mathcal{S}_{0,t}(\mathbb{X}) \in G^{(*)}$ is injective.

In [HL10] for the case when $p = 1$, and later in [Boe+16] for the case $p > 1$ the following collection of results are proved.

Theorem 1.1.4 ([HL10; Boe+16]). Let $p \geq 1$, then

- 1) The relation $\sim_\tau$ defines an equivalence relation on $W\Omega_p(V)$;
- 2) For all $\mathbb{X}, \mathbb{Y} \in W\Omega_p(V)$ it holds that $\mathbb{X}^{<\infty} = \mathbb{Y}^{<\infty}$ if and only if $\mathbb{X} \sim_\tau \mathbb{Y}$;
- 3) Each $\sim_\tau$ equivalence class contains a unique tree-reduced representative (up to reparameterisation).

We will use $[\mathbb{X}]$ to denote the tree-like equivalence class of a path $\mathbb{X} \in W\Omega_p(V)$. The following result provides an explicit characterisation of the tree-reduced representative of an equivalence class in the case $p = 1$.

Proposition 1.1.5 ([HL10]). When $p = 1$, the tree-reduced representative of an equivalence class $[\gamma]$ is the unique (up to reparameterisation) minimiser of the length functional.

Definition 1.1.6 (Unparameterised Path Space). The space of unparameterised rough paths $\mathcal{C}_p$ is defined as the quotient space $W\Omega_p / \sim_\tau$. The quotient map is denoted by π .

Remark 1.1.7. We can canonically identify $\mathcal{C}_p$ with the space $\mathcal{S}_p$ and will occasionally use them interchangeably where appropriate.

Returning to the question of approximating a continuous function $f : K \rightarrow \mathbb{R}$ with the family $\{\mathcal{S}(\cdot, \mathbf{x}) : \mathbf{x} \in T(V)\}$, it is now clear that K should be a compact subset of unparameterised path space. Chapter 2 deals with the task of defining topologies

on unparameterised path space so that notions of continuity and compactness are well defined. Then in [Chapters 3 and 4](#), we will study two canonical examples of functions on unparameterised path space: the terminal time solution to a CDE and the asymptotics of the signature map.

1.1.1 Avoiding the quotient

A popular approach to avoid the issues created by quotienting by $\sim_\tau$ is to add the time parameterisation as an additional coordinate, i.e. setting $\hat{\gamma}_t = (t, \gamma_t)$, see for example [\[Bay+23; Boe+16; CO22\]](#). The monotone coordinate ensures the signatures of $\hat{\gamma}$ and $\hat{\sigma}$ coincide if and only if $\gamma = \sigma$. This transformation presents two issues. First, the signature will no longer remove the symmetry of time reparameterisation. This is appropriate in some cases such as in financial applications [\[Bay+23\]](#) where the parameterisation is important, but not in others [\[Yan+16\]](#). Secondly, care must be taken in how the space of paths is defined: the concatenation of paths does not commute with $\hat{\cdot}$ when all paths are parameterised on a pre-defined interval, so Chen's identities fail to hold. In this chapter, as is customary in the literature, see for example [\[CF19\]](#), we work with paths over a fixed pre-defined interval (which we take to be $[0, T]$ for convenience). A plausible alternative might be to take account for paths defined on possibly different compact subintervals of $\mathbb{R}$. While this would leave some results of [Chapter 2](#) unchanged, it adds an additional challenge of topologising the underlying path space so that the signature map itself is continuous. We will briefly return to this discussion in [Section 2.1](#).

1.2 Point set topology

The analysis of [Chapter 2](#) will require some familiarity with certain topological definitions and we gather here the various notions that will be used. We write (X, χ) for a generic space X equipped with a topology χ

Definition 1.2.1 (Complete-metrisability and Polishness). *A topological space (X, χ) is completely-metrisable if there exists a metric generating χ which is complete. (X, χ) is called Polish if it is separable and completely-metrisable.*

Remark 1.2.2. *Completeness is a property of a metric and not the topology it generates. There may be many metrics generating the same topology, where some of them are complete and others not.*

Definition 1.2.3. *Let X be a metric space. We say that a subset $A \subseteq X$ is*

- 1) *Nowhere dense if the interior of the closure of A is empty;*

- 2) Meagre in X if it can be written as the countable union of nowhere dense sets;
- 3) G_δ if A is the countable intersection of open sets;
- 4) σ -compact if it is the countable union of compact sets.

Definition 1.2.4. A topological space (X, χ) is

- 1) First countable if for every $x \in X$ there is a sequence of neighbourhoods $\{N_i\}_{i=0}^\infty$ such that for any neighbourhood N of x , $N_i \subseteq N$ for some i ;
- 2) Regular if for any closed $F \subseteq X$ and $x \in X \setminus F$, there exists a neighbourhood U of F and neighbourhood V of x that are disjoint;
- 3) Completely regular if for any closed $F \subseteq X$ and $x \in X \setminus F$, there exists some continuous $f : X \rightarrow \mathbb{R}$ for which $f(x) = 1$ and $f(F) = 0$;
- 4) A Baire space if every meagre subset has empty interior;
- 5) A Lusin space if there exists a topology $\tilde{\chi} \supseteq \chi$ for which $(X, \tilde{\chi})$ is Polish.

Finally, we recall the Baire category theorem. There are many equivalent statements of the following, and the exact statement we provide is often viewed as a consequence of the Baire category theorem. However, it will suffice for the results of this thesis.

Theorem 1.2.5 (Baire category theorem). *The following two statements are true:*

- 1) Every completely metrisable topological space (X, χ) is a Baire space;
- 2) Every locally compact regular topological space (X, χ) is a Baire space.

CHAPTER 2

TOPOLOGIES ON UNPARAMETERISED PATH SPACE

As noted in [HL10], there is no canonical topology with which to equip the space of unparameterised paths: which properties should one privilege? The quotient topology inherited from the underlying path space gives the universal property, while defining convergence of unparameterised paths by the convergence of their signatures aligns with their usage in application. One can alternatively define metrics based on representatives of the equivalence classes. As we will see in this chapter, each candidate has dramatically different topological properties and no candidate is clearly superior to the others. Our aim is not to propose one single topology, but rather to explore the benefits and drawbacks of each so that an informed selection may be made for the application at hand. This chapter is based on the papers [CT24b; CT24c].

As discussed in [Section 1.1](#), the algebra of functions $\{\mathcal{S}(\cdot, \mathbf{x}) : \mathbf{x} \in T(V)\}$ allows for an approximation theory to be developed on unparameterised path space. The most basic form of such a theory provides that any continuous function f on a (locally) compact subset K of $\mathcal{C}_1$ can be uniformly approximated by (the restriction of) a linear functional on the tensor algebra. The choice of topology on the space of unparameterised paths which is necessary to a complete understanding of this method, is often left unspecified in applications or is otherwise chosen to suit the application at hand.

The purpose of this chapter is to broaden this discussion by evaluating and comparing the properties of three topologies which are arrived at by leveraging different properties of $C^{1\text{-var}}([0, T], V)$. The qualities we emphasise are:

- 1) The injectivity of the signature map from $C^{1\text{-var}}([0, T], V)$ onto a subset of $T((V))$. This allows one to equip the range of the signature map with a (subspace) topology, and hence to transfer this topology onto $\mathcal{C}_1$ by requiring the signature map to be an embedding. When $T((V))$ is equipped with the product topology, we refer to the resulting topology on $\mathcal{C}_1$ as the product topology.
- 2) The fact that tree-like equivalence on the space of continuous bounded variation paths is an equivalence relation means that $\mathcal{C}_1$ can be endowed with the quotient topology derived from the 1-variation topology.
- 3) The existence for each equivalence class $[\gamma]$ in $\mathcal{C}_1$ of a so-called tree-reduced representative [HL10], which is characterised by having minimal length. By considering the constant speed parameterised version of this representative γ^* , we can define the metric topology on $\mathcal{C}_1$ induced by

$$d([\gamma], [\sigma]) := \|\gamma^* - \sigma^*\|_{1\text{-var};[0,T]}.$$

The product and quotient topologies can be seen as two alternatives which can be determined without extra structure. The former only utilises a topology on the range of the signature map, while the latter only relies on the given topology on the path space domain. By contrast, the metric topology relies on additional properties of the space of unparameterised paths. Variants of (1) may also be of interest, e.g. where a topology on $\mathcal{C}_1$ is induced from a different subspace topology on the extended tensor algebra, but the principle in these cases remains the same [Che16; CL16]. In fact, if $\mathcal{C}_1$ is endowed with any metrisable topology for which the signature map is (appropriately) continuous, the conclusions for the product topology may be carried over to the resulting induced topology on $\mathcal{C}_1$. We draw special attention to the product topology as it is the weakest one for which linear functionals on the tensor algebra are continuous.

Given the central role played by uniform approximation to a now-growing body of applications, it seems important to establish the basic topological features of these spaces. The main conclusions of Section 2.1 are captured in the following list:

- i) All of the three topologies listed above are separable and Hausdorff.
- ii) The collections of open sets are strictly ordered by inclusion, where the product topology is the weakest and the metric topology is the strongest.
- iii) Open sets in the product and quotient topologies are unbounded in the metric d .

- iv) The product topology is metrisable and σ -compact, but not a Baire space, and so neither Polish nor locally compact.
- v) A completion of the product topology is given by the subspace $G^{(*)} \subset T((V))$ of group-like elements.
- vi) The quotient topology is not metrisable.
- vii) The metric d is not complete, but it generates a completely metrisable topology.

The uniform approximation theory referred to above invites the study of two points. The first is the availability of compact subsets of $C^{1\text{-var}}([0, T], V)$ – which should ideally be explicitly describable – and the second is to understand the classes of continuous functions on these subsets. With respect to both points, the underlying topology has an unignorable bearing. We investigate this to find that:

- i) The subset $B(r)$ of unparameterised paths with tree reduced length bounded by $r < \infty$ is a compact subset of $C^{1\text{-var}}([0, T], V)$ in the product topology.
- ii) The fixed-time solution of a differential equation $dy_t = f(y_t)d\gamma_t$, under suitable regularity and growth conditions on f , induces a well-defined function on $\mathcal{C}_1$. This function is continuous on $B(r)$.

The relevance of these results is accounted for by the uses of the signature: physical limits on the ability to record and store data impose a bound on the (tree-reduced) lengths of the paths that can practically be considered. From this point of view, restricting attention to functions on $B(r)$ is a natural step. The second item ensures that a rich class of input-output response pairs is admissible as the true underlying causal dynamics, e.g. in regression analysis.

The properties of the topology on unparameterised path space are also relevant for certain open problems. In 2009, Levin and Lyons [LL09] conjectured the existence of a signed measure on unparameterised path space associated to higher order differential operators (in the same way that the Wiener measure corresponds to the half-Laplacian). They provided partial evidence for this conjecture, but the full picture remains unresolved. Among the missing ingredients is a more advanced understanding of the available measure-theoretic tools on unparameterised path space. Of particular importance is the ability to characterise weak limits and suitable criteria to ensure (pre)compactness for families of measures. In [Section 2.3](#) we briefly introduce this problem and discuss connections with the results of this chapter.

From the perspectives of the above applications and from the point of view of studying probability on $\mathcal{C}_1$ more generally, it can be desirable to work with measures defined on the Borel σ -algebra of a Polish topology, see e.g. the monograph [Gar18] for a detailed overview. The results above exclude having this structure for both the product and the quotient topologies, but allow for the use of the metric topology. Moreover, the σ -compactness of the product topology still offers a way to obtain some of the benefits of Polishness, notably versions of Ulam's and Lusin's Theorems still hold. In Section 2.2 we illustrate how this approach can be used to validate and extend the framework of the expected signature model proposed in [LLN16].

2.0.1 Candidate topologies

Our aim is now to explore the basic properties of three topologies on the spaces $\mathcal{C}_p$ of unparameterised paths. In the context of topologies on $\mathcal{C}_1$, this will allow us to put forward a framework within which one can understand uniform approximation results for continuous functions defined on compact subsets of $\mathcal{C}_1$, as well as the expected signature model of [LLN16]. As noted in [HL10], there is no canonical topology, but there are at least three principled approaches to constructing one:

- 1) To use the injectivity of the signature map $\mathcal{S} : \mathcal{C}_p \rightarrow \mathcal{S}_p \subset T((V))$.
- 2) To use the quotient topology on $\mathcal{C}_1$ inherited from the p -variation norm topology on $WG\Omega_p(V)$.
- 3) To define a metric using the tree-reduced representatives parameterised at constant speed.

To this end, we now introduce our candidate topologies.

Definition 2.0.1. Fix $p \geq 1$. We consider the following topologies on unparameterised path space.

- 1) Let m be any metric on $\mathcal{C}_p$ for which $\mathcal{S} : WG\Omega_p \rightarrow (\mathcal{C}_p, m)$ is continuous. Let χ_m denote the topology on $\mathcal{C}_p$ induced by m .
- 2) The quotient topology χ_τ on $\mathcal{C}_p$, inherited from the underlying topology on $WG\Omega_p(V)$.
- 3) The metric topology χ_d defined by the metric

$$d([\mathbb{X}], [\mathbb{Y}]) := d_{p\text{-var}}(\mathbb{X}^*, \mathbb{Y}^*), \quad (2.1)$$

where $\mathbb{X}^*$ is the Hölder-control parameterisation of the unique tree-reduced representative of $[\mathbb{X}]$.

Remark 2.0.2. Recall that product topology on a countable product of metric spaces is metrisable. Equipping $T((V))$ with the product topology we can then impose the subspace topology on $\mathcal{S}_p$ and pull this back to induce a metrisable topology on $\mathcal{C}_p$ of the type 1). We draw specific attention to this topology and as such denote it by χ_{pr} .

Remark 2.0.3. A possible alternative metric (suggested by Dr. Xi Geng) between equivalence classes is defined by

$$d_\star([\mathbb{X}], [\mathbb{Y}]) := \|(\mathbb{X} \star \overleftarrow{\mathbb{Y}})^\star\|_{p\text{-var};[0,T]}. \quad (2.2)$$

This metric is clearly symmetric and satisfies the triangle inequality since

$$\begin{aligned} \|(\mathbb{X} \star \overleftarrow{\mathbb{Y}})^\star\|_{p\text{-var}} &= \left\| \left((\mathbb{X} \star \overleftarrow{\mathbb{Z}})^\star \star (\mathbb{Z} \star \overleftarrow{\mathbb{Y}})^\star \right)^\star \right\|_{p\text{-var};[0,T]} \\ &\leq \|(\mathbb{X} \star \overleftarrow{\mathbb{Z}})^\star \star (\mathbb{Z} \star \overleftarrow{\mathbb{Y}})^\star\|_{p\text{-var};[0,T]} \\ &\leq \|(\mathbb{X} \star \overleftarrow{\mathbb{Z}})^\star\|_{p\text{-var};[0,T]} + \|(\mathbb{Z} \star \overleftarrow{\mathbb{Y}})^\star\|_{p\text{-var};[0,T]} \\ &= d_\star([\mathbb{X}], [\mathbb{Z}]) + d_\star([\mathbb{Z}], [\mathbb{Y}]), \end{aligned}$$

where the second inequality follows from the sub-additivity of p -variation. The resulting topology is limited in practical use however, since it fails even to be separable. Indeed, consider the set of canonical lifts of the V -valued paths $K = \{\frac{t}{T}v : \|v\| = 1\}$ into $G\Omega_p(V)$. For $\mathbb{X}^{\leq [p]} \in K$, the definition of the p -variation norm and [Proposition 1.0.28](#) implies that

$$\|\mathbb{X}^{\leq [p]}\|_{p\text{-var}}^p \geq \|\pi_1(\mathbb{X}_1 - \mathbb{X}_0)\|^p = \|v\|^p = 1.$$

Given two distinct paths $\mathbb{X}, \mathbb{Y} \in K$ the path $(\mathbb{X} \star \overleftarrow{\mathbb{Y}})^{\leq 1}$ is a piecewise linear irreducible path and hence tree-reduced, so that the path $(\mathbb{X} \star \overleftarrow{\mathbb{Y}})^{\leq [p]}$ is also tree-reduced. The distance between these paths is then lower bounded by

$$\begin{aligned} d_\star([\mathbb{X}^{\leq [p]}], [\mathbb{Y}^{\leq [p]}])^p &= \|(\mathbb{X} \star \overleftarrow{\mathbb{Y}})^{\leq p}\|_{p\text{-var}}^p \\ &\geq \|\mathbb{X}^{\leq [p]}\|_{p\text{-var}}^p + \|\mathbb{Y}^{\leq [p]}\|_{p\text{-var}}^p \\ &\geq 2, \end{aligned}$$

by the super-additivity of controls. Since K is an uncountable set, the topology induced by this metric cannot be separable.

We now turn our attention to the specific case of $\mathcal{C}_1$. The case of unparameterised rough path space is more subtle.

2.1 Unparameterised paths of bounded variation

Tree-reduced paths will be central to our analysis. Indeed when $p = 1$, the characterisation of the tree-reduced representative minimising length will be crucial to certain results.

Remark 2.1.1. Recall [Proposition 1.0.39](#) that if $\gamma \in C^{1-var}([0, T], V)$ is parametrised at constant speed then γ is Lipschitz continuous with Lipschitz constant $\frac{1}{T} \|\gamma\|_{1-var;[0,T]}$.

We can manufacture simple examples of paths that are already tree-reduced by considering axis paths defined with respect to an orthonormal basis. Given v in V we let γ_v denote the linear path in $C^{1-var}([0, T], V)$ whose derivative $\gamma'_v \equiv \frac{1}{T}v$ on $[0, T]$.

Example 2.1.2. For a collection of $v_1, v_2, \dots, v_m$ in V the piecewise linear path $\gamma : [0, T] \mapsto V$ in $C^{1-var}([0, T], V)$ defined by $\gamma = \gamma_{v_1} \star \gamma_{v_2} \star \dots \star \gamma_{v_m}$ has length $L = \|\gamma\|_{1-var;[0,T]} = |v_1| + |v_2| + \dots + |v_m|$. When parameterised at constant speed it is defined by

$$\gamma'_t = \frac{L}{T} \hat{v}_i := \frac{L}{T} \cdot \frac{v_i}{|v_i|} \text{ for } t \text{ in } [t_{i-1}, t_i], \quad (2.3)$$

where $t_0 = 0$ and

$$t_i = \frac{T}{L} \sum_{j=1}^i |v_j| \text{ for } i = 1, \dots, m.$$

From now on, unless stated otherwise, we assume that all piecewise linear paths as in [Example 2.1.2](#) are parameterised at constant speed. The following result is known more generally for C^2 paths [\[HL10\]](#) and irreducible piecewise linear paths [\[LX17\]](#), but the arguments rely delicately on estimates of the hyperbolic development of the path. In the case of axis paths it is possible to present a simplified and direct proof using elementary tools. The use of axis paths will be fundamental to our subsequent discussion, and so we include a proof for completeness.

Lemma 2.1.3. Let $\gamma : [0, T] \rightarrow V$ in $C^{1-var}([0, T], V)$ be of the form $\gamma = \gamma_{v_1} \star \gamma_{v_2} \star \dots \star \gamma_{v_m}$ as in [Example 2.1.2](#). Assume further that for every consecutive pair of vectors $\langle v_i, v_{i+1} \rangle = 0$ for $i = 1, \dots, m - 1$. Then γ is tree reduced.

Proof. We prove the result indirectly not through the definition of tree-like equivalence, but by using asymptotic properties of the signature. We recall that the projective tensor norm is the largest cross norm on $V^{\otimes n}$. For a tensor A in the

algebraic n -fold tensor product space it is defined by

$$\|A\|_{V,n} = \inf \left\{ \sum_{i=1}^r |u_1^i| \dots |u_n^i| : A = \sum_{i=1}^r u_1^i u_2^i \dots u_n^i \right\}.$$

It can be deduced that if $A = \sum_i u_1^i u_2^i \dots u_{2n}^i$ then

$$\|A\|_{V,2n} \geq \varphi(A) := \sum_i \langle u_1^i, u_2^i \rangle \dots \langle u_{2n-1}^i, u_{2n}^i \rangle.$$

Note here that φ is a bounded linear map on $V^{\otimes n}$. Consider γ with the constant-speed parameterisation described in [Example 2.1.2](#). Then, using the previous inequality, we obtain the lower bound

$$\begin{aligned} (2n)! \|S(\gamma)^n\|_{V,2n} &\geq (2n)! \varphi \left(\int_{\Delta_{[0,T]}^{2n}} \gamma'_{t_1} \otimes \gamma'_{t_2} \otimes \dots \otimes \gamma'_{t_{2n}} dt_1 dt_2 \dots dt_{2n} \right) \\ &= (2n)! \int_{\Delta_{[0,T]}^{2n}} \varphi(\gamma'_{t_1} \otimes \gamma'_{t_2} \otimes \dots \otimes \gamma'_{t_{2n}}) dt_1 dt_2 \dots dt_{2n} \\ &\geq (2n)! \int_{\Delta_{[0,T]}^{2n}} \prod_{j=1}^n \langle \gamma'_{t_{2j-1}}, \gamma'_{t_{2j}} \rangle dt_1 dt_2 \dots dt_{2n} \\ &= \mathbb{E} \left[T^{2n} \prod_{j=1}^n \langle \gamma'_{U_{(2j-1)}}, \gamma'_{U_{(2j)}} \rangle \right] =: \mathbb{E}[X_n], \end{aligned}$$

where $0 \leq U_{(1)} \leq U_{(2)} \leq \dots \leq U_{(2n)} \leq 1$ are the order statistics of an i.i.d. sample of $2n$ uniform $[0, T]$ random variables defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let N_i , for $i = 1, \dots, m$, denote the number of realisations in this sample which are contained in the interval $[t_{i-1}, t_i]$. We define three events by

$$A = \bigcap_{i=1}^m \{N_i \text{ is even}\}, \quad B = \bigcup_{i=1}^m \{N_i = 0\}, \quad C = B^c \cap \bigcup_{i=1}^m \{N_i \text{ is odd}\}.$$

The event B is required, since it is possible for X_n to be non-zero even if one of the N_i is odd: the order statistics may “skip” a segment. Their union is Ω , and it can easily be checked that

$$X_n = \begin{cases} \frac{L^{2n}}{T^{2n}}, & \text{on } A \\ 0, & \text{on } C, \end{cases}$$

and therefore that

$$(2n)! \|S(\gamma)^{2n}\|_{V,2n} \geq L^{2n} \mathbb{P}(A) - L^{2n} \mathbb{P}(B). \quad (2.4)$$

Letting $\{E_i : i = 1, 2, \dots, m\}$ be a collection of i.i.d Rademacher random variables and $r = \min_{i=1, \dots, m} |t_i - t_{i-1}| \in (0, T)$, simple arguments can be employed to show

$$\mathbb{P}(A) = \mathbb{E} \left[\left(\sum_{i=1}^m E_i \frac{|v_i|}{L} \right)^{2n} \right] \geq \frac{2}{2^m}, \text{ and } \mathbb{P}(B) \leq m(1-r)^{2n}.$$

Using these inequalities together with [Eq. \(2.4\)](#) we obtain easily that

$$\limsup_{n \rightarrow \infty} \left(n! \left\| S(\gamma)^{2n} \right\|_{V,n} \right)^{\frac{1}{n}} \geq L.$$

For any path $\tilde{\gamma}$ in $C^{1\text{-var}}([0, T], V)$ with length $\tilde{L}$ such that $S(\gamma) = S(\tilde{\gamma})$ we deduce

$$L \leq \limsup_{n \rightarrow \infty} \left(n! \left\| S(\gamma)^n \right\|_{V,n} \right)^{\frac{1}{n}} \leq \tilde{L},$$

where the second inequality follows from [\[Lyo+07, Proposition 2.2\]](#), and hence γ is tree-reduced. $\square$

The following construction can be found in [\[LX17\]](#). It illustrates how the class of paths considered in the previous lemma can be used to gain insight into properties of the signature. It will be an integral tool in the next section.

Example 2.1.4 ([\[LX17\]](#)). *Let W be a two-dimensional vector subspace of V which is identified with $\mathbb{R}^2$ through an orthonormal basis $\{v_i : i = 1, 2\}$. We can define two sequences $\{\rho_n : n = 1, 2, \dots\}$ and $\{\sigma_n : n = 1, 2, \dots\}$ of so-called axis paths by $\rho_1 := \gamma_{v_1} \star \gamma_{v_2}$, $\sigma_1 := \gamma_{v_2} \star \gamma_{v_1}$ and then for $n = 2, 3, \dots$ by*

$$\rho_n = \rho_{n-1} * \sigma_{n-1} \text{ and } \sigma_n = \sigma_{n-1} * \rho_{n-1}.$$

If two consecutive line segments v_i and v_{i+1} are positively collinear, then by replacing $\gamma_{v_i} * \gamma_{v_{i+1}}$ with $\gamma_{v_i+v_{i+1}}$, we may write each path in the form required for [Lemma 2.1.3](#). By virtue of this lemma, each of these paths is tree-reduced. For every n , the paths ρ_n and σ_n have length 2^n and it was further shown in [\[LX17\]](#) that the terms in their signatures coincide up to degree n , i.e.

$$S(\rho_n)^k = S(\sigma_n)^k \text{ for } k = 1, 2, \dots, n.$$

Moreover, it can be shown that $S(\rho_n)^{n+1}$ and $S(\sigma_n)^{n+1}$ are not equal for any n . Consequently, each tree-reduced path $\Gamma_k := \sigma_k \star \overleftarrow{\rho}_k$ has length 2^{k+1} and satisfies

$$S(\Gamma_k)^m = S(\sigma_k * \overleftarrow{\sigma}_k)^m = 0 \text{ for } m = 1, 2, \dots, k, \text{ while } S(\Gamma_k) \neq \mathbf{1},$$

since the projection at level k of the algebraic inversion and multiplication depends only on the level k projections of the inputs.

2.1.1 Basic topological properties

Let us now show some simple properties of these topological spaces. The following proposition shows that the three topologies we have proposed are distinct. In fact, the collections of open sets are strictly ordered with χ_m defining the weakest and χ_d the strongest of the topologies.

Proposition 2.1.5. *We have the strict inclusions $\chi_m \subset \chi_\tau \subset \chi_d$.*

Proof. The first (non-strict) inclusion follows immediately from the universal property of the quotient topology. We leave the strictness for general m until [Theorem 2.1.10](#), but we provide a direct proof for the special case χ_{pr} . To see this strict inclusion we consider a two-dimensional subspace W of V spanned by two orthonormal vectors $\{v_1, v_2\}$. In this basis, we consider the sequence of axis paths $(\Gamma_k)_{k=1}^\infty$ constructed in [Example 2.1.4](#). Define the set $A := \{[\Gamma_k] : k = 1, 2, \dots\} \subset \mathcal{C}_1$. Then A is not closed in the product topology, since for every m we have

$$\mathcal{S}(\Gamma_k)^m \rightarrow 0 \text{ as } k \rightarrow \infty$$

while A does not contain the equivalence class of tree-like paths $[o]$. On the other hand, A is closed in the quotient topology because any sequence $(\gamma_n)_{n=1}^\infty$ in $\pi^{-1}(A)$ which converges to γ in $C^{1-var}([0, T], V)$ must be such that $\sup_n \|\gamma_n\|_{1-var; [0, T]} < \infty$. Using the fact that each Γ_k is tree-reduced and $\|\Gamma_k\|_1 = 2^{k+1}$ then shows that there exists $N \in \mathbb{N}$ such that $\{\gamma_n : n \in \mathbb{N}\}$ is a subset of $\cup_{k=1}^N [\Gamma_k] = \cup_{k=1}^N \pi^{-1}(\{[\Gamma_k]\})$. This latter set is closed in $C^{1-var}([0, T], V)$ since singletons are closed in any Hausdorff topology which χ_τ must be since $\chi_{pr} \subseteq \chi_\tau$. The limit γ is then also in this set which is a subset of $\pi^{-1}(A)$.

To prove that $\chi_\tau \subseteq \chi_d$, we let $[\gamma]$ be in $\mathcal{C}_1$ and prove that every open neighbourhood N of $[\gamma]$ in χ_τ is also a neighbourhood of $[\gamma]$ in χ_d . By definition, the pre-image $\pi^{-1}(N) \supset [\gamma]$ is an open subset of $C^{1-var}([0, T], V)$. The tree-reduced representative γ^* belongs to $[\gamma]$ and consequently there exists $\delta > 0$ such that $B_{\|\cdot\|_1}(\gamma^*, \delta) = \{\sigma \in C^{1-var}([0, T], V) : \|\gamma^* - \sigma\|_1 < \delta\} \subset \pi^{-1}(N)$. The result then follows because

$$N \supset \{[\sigma] : \|\gamma^* - \sigma\|_1 < \delta\} \supseteq \{[\sigma] : \|\gamma^* - \sigma^*\|_1 < \delta\} = B_d([\gamma], \delta).$$

To prove the strict inclusion we find a neighbourhood of the constant path $[o]$ of the form $B_d([o], \delta)$ such that every χ_τ -neighbourhood of $[o]$ has non-empty intersection with the complement $B_d([o], \delta)^c := \mathcal{C}_1 \setminus B_d([o], \delta)$. To do so, we consider as above a two-dimensional subspace of V spanned by an orthonormal set $\{v_1, v_2\}$. Using this basis we define a family of axis paths given by:

$$\gamma_\varepsilon = \gamma_{\varepsilon v_2} * \gamma_{v_1} * \gamma_{-\varepsilon v_2} * \gamma_{-v_1} \text{ for } \varepsilon \geq 0.$$

For every $\varepsilon > 0$ the curve γ_ε is tree-reduced by [Lemma 2.1.3](#). On the other hand, γ_0 is tree-like equivalent to the constant path o . It follows that

$$d([\gamma_0], [\gamma_\varepsilon]) = \| \gamma_\varepsilon \|_1 \geq 2, \quad (2.5)$$

while an easy calculation using [Eq. \(2.3\)](#) shows that $\| \gamma_\varepsilon - \gamma_0 \|_1 \leq 6\varepsilon$. Hence any χ_τ -neighbourhood of $[o]$ must contain $[\gamma_\varepsilon]$ for some $\varepsilon > 0$. In view of [Eq. \(2.5\)](#), the set $B_d([o], \delta)$ can never contain such a neighbourhood whenever $\delta < 2$. $\square$

The inclusions $\chi_m \subset \chi_\tau \subset \chi_d$ and the fact that χ_m is a Hausdorff space immediately yields the following corollary.

Corollary 2.1.6. *All three candidate topologies are Hausdorff.*

For the specific case of χ_{pr} we may obtain the following result.

Proposition 2.1.7. *$(\mathcal{C}_1, \chi_{pr})$ with the operations $[\gamma] \cdot [\sigma] := [\gamma * \sigma]$ and $[\gamma]^{-1} := [\overleftarrow{\gamma}]$ forms a topological group. This is not the case for the metric topology χ_d .*

We provide below a proof from first principles, however the continuity may also be seen from existing results in the literature: for example by using that $G^{(n)}$ is a closed Lie subgroup of the closed (with respect to the product topology) set $\pi^{(n)}T_1((V))$. The manifold topology of the Lie group is the same as the subspace topology of the product topology; for further details we refer the reader to [\[FV10, Chapter 7\]](#).

Proof. We first show continuity of the group operations on $(\mathcal{C}_1, \chi_{pr})$. For $[\cdot]^{-1}$, we recall the algebraic inversion $\mathbf{x}^{-1}$ defined for $\mathbf{x} \in T_1((V))$. We may define the map

$$\begin{aligned} \psi : T((V)) &\rightarrow T((V)), \\ \mathbf{x} &\mapsto \sum_{n=1}^{\infty} (\mathbf{1} - \mathbf{x})^{\otimes n}, \end{aligned}$$

which is well-defined for all $\mathbf{x} \in T((V))$ and restricts to the inversion operation on $T_1((V))$. By definition of the subspace topology and the fact that $\mathcal{S}_1 = \psi^{-1}(\mathcal{S}_1)$, it suffices to show continuity of ψ on $T((V))$. Moreover, by definition of the product topology, it is enough to show continuity in each factor. Since the projection onto $V^{\otimes n}$ of $\psi(\mathbf{a})$ depends only on $\pi^{(n)}\mathbf{a}$, we may consider the factorisation

$$\begin{array}{ccc} T((V)) & & \\ \pi^{(n)} \downarrow & \searrow \psi(\cdot)_n & \\ T^{(n)}(V) & \xrightarrow{\psi_n} & V^{\otimes n} \end{array}$$

where $\psi_n(\mathbf{b}) := \pi_n \circ \psi \circ \iota^{(n)}(\mathbf{b})$. By continuity of $\pi^{(n)}$ we only need to show continuity of ψ_n . This follows from the continuity of the tensor product, the canonical isomorphisms $V^{\otimes k} \otimes V^{\otimes l} \cong V^{\otimes k+l}$, and the continuity of addition. The continuity of group multiplication follows similarly.

We now prove the discontinuity of group multiplication with respect to the metric topology. Let $\{v_1, v_2\}$ be orthonormal vectors and define the two sequences of axis paths

$$\rho_n = \gamma_{\frac{v_2}{n}} \star \gamma_{v_1}, \quad \sigma_n = \gamma_{-\frac{v_2}{n}} \star \gamma_{-v_1} \text{ for } n \in \mathbb{N}.$$

For each n , the path ρ_n (*resp.* σ_n) is tree reduced so that $\rho_n^* = \rho_n$ (*resp.* $\sigma_n^* = \sigma_n$). Moreover the concatenation $\rho_n \star \sigma_n$ is tree reduced and parameterised at constant speed. It follows that

$$\mathcal{d}([\rho_n] \cdot [\sigma_n], [o]) = \|\rho_n \star \sigma_n - o\|_{1\text{-var};[0,T]} \geq 2.$$

On the other hand, for every n

$$\mathcal{d}([\rho_n], [\gamma_{v_1}]) = \mathcal{d}([\sigma_n], [\gamma_{-v_1}]) \leq \frac{3}{n},$$

and so ρ_n (*resp.* σ_n) converges to $[\gamma_{v_1}]$ (*resp.* $[\gamma_{-v_1}]$) in χ_d . But since $\gamma_{v_1} \star \gamma_{-v_1}$ is tree-like, we have

$$\mathcal{d}([\gamma_{v_1}] \cdot [\gamma_{-v_1}], [o]) = \mathcal{d}([o], [o]) = 0.$$

Hence multiplication is not continuous with respect to χ_d . □

2.1.2 Complete metrisability of candidate topologies?

We start by answering this question in the negative for the product topology $(\mathcal{C}_1, \chi_{\text{pr}})$, where again a direct proof is possible.

Proposition 2.1.8. *The product topology $(\mathcal{C}_1, \chi_{pr})$ is not a Baire space.*

Proof. The proof has two parts: first we show that every non-empty open set in χ_{pr} is unbounded with respect to the metric d , and then that $\mathcal{C}_1$ may be written as the countable union of closed (with respect to χ_{pr}) sets that are bounded (with respect to d).

Consider the two dimensional subspace of V spanned by two orthonormal vectors $\{v_1, v_2\}$, and let $(\Gamma_k)_{k=1}^\infty$ be the sequence of axis paths defined in [Example 2.1.4](#). Each Γ_k is tree reduced by [Lemma 2.1.3](#), thus

$$d([o], [\Gamma_k]) = 2^{k+1}.$$

As shown in the proof of [Proposition 2.1.5](#), $[\Gamma_k]$ converges to $[o]$ in χ_{pr} , and so every open neighbourhood of $[o]$ contains infinitely many terms of the sequence $(\Gamma_k)_{k=1}^\infty$. Hence, every open neighbourhood of $[o]$ is unbounded in d . Now let $[\gamma] \in \mathcal{C}_1$ and U be any open neighbourhood of γ . Define the map

$$\begin{aligned} \lambda : U &\longrightarrow \mathcal{C}_1 \\ [\sigma] &\longmapsto [\gamma]^{-1} \cdot [\sigma], \end{aligned}$$

which is continuous by [Proposition 2.1.7](#) and bijective with its image $V := \lambda(U)$. The inverse map

$$\begin{aligned} \lambda^{-1} : V &\longrightarrow U \\ [\sigma] &\longmapsto [\gamma] \cdot [\sigma], \end{aligned}$$

is also continuous by [Proposition 2.1.7](#). Therefore $\lambda(U)$ is an open neighbourhood of $[o]$. If U were bounded in d from $[o]$ by some constant $K > 0$, then by construction of λ , the length of the tree-reduced representative of every $[\sigma]$ in V is at most $2K$. This would imply that V were bounded in d , a contradiction.

For any $r > 0$, define the set

$$B(r) := \{[\gamma] \in \mathcal{C}_1 : \|\gamma^*\|_1 \leq r\}.$$

Later in [Proposition 2.2.2](#), that $B(r)$ is compact in χ_{pr} , and so closed by the Hausdorff property. Since each $B(r)$ is bounded in d , it must have empty interior by the preceding. Writing

$$\mathcal{C}_1 = \bigcup_{r=1}^{\infty} B(r),$$

we see that $\mathcal{C}_1$ is the countable union of nowhere dense sets and so is not a Baire space. $\square$

As a consequence of the above and item (2) of the Baire category theorem ([Theorem 1.2.5](#)) we obtain the following corollary.

Corollary 2.1.9. *The product topology $(\mathcal{C}_1, \chi_{pr})$ is not locally compact.*

The following is one of the main results of this chapter.

Theorem 2.1.10. *The three topologies on the space $\mathcal{C}_1$ of unparameterised paths are separable and have the following properties.*

- 1) *The product topology $(\mathcal{C}_1, \chi_{pr})$ is metrisable, but not completely metrisable.*
- 2) *The quotient topology $(\mathcal{C}_1, \chi_\tau)$ is not metrisable.*
- 3) *The metric d is not complete, but is completely metrisable.*

Proof. We will show separability of χ_d and use the inclusions $\chi_{pr} \subset \chi_\tau \subset \chi_d$ to conclude separability of χ_{pr} and χ_τ . By [Lemma 1.0.36](#), the space of absolutely continuous paths $C^{0,1-var}([0, T], V)$ is separable. Consider the subspace A of $C^{1-var}([0, T], V)$ of tree-reduced paths parameterised at constant speed. That is, all representatives of equivalence classes seen by the metric d . Since every path parameterised at constant speed is Lipschitz, we may consider A as a subspace of absolutely continuous paths. Since any subspace of a separable metric space is separable, A is separable with respect to $\|\cdot\|_{1-var;[0,T]}$. This is equivalent to the separability of χ_d .

Item (1) is a consequence of the definition of χ_{pr} , [Proposition 2.1.8](#) and [Theorem 1.2.5](#).

We will prove item (2) using the fact that any metrisable space must be both first countable and regular [[Nag50](#)]. By assuming first-countability we show regularity cannot hold. Thus we assume that there exists $\{U_i\}_{i=1}^\infty$, a countable neighbourhood basis of $[o]$ which without loss of generality is assumed to consist of open sets. As before, we consider sequences of axis paths in a two-dimensional subspace of V defined relative to two orthonormal vectors $\{v_1, v_2\}$. The first sequence will consist of tree-like paths; for every $n \in \mathbb{N}$ we take

$$\gamma_n := \gamma_{nv_1} \star \gamma_{-nv_1} = \gamma_{nv_1} \star \overleftarrow{\gamma_{nv_1}}.$$

Every γ_n belongs to the equivalence class $[o]$ and hence

$$r_n = \sup \{r > 0 : B(\gamma_n, r) \subset \pi^{-1}(U_n)\} > 0.$$

We introduce a strictly decreasing sequence of positive real numbers $\{a_n\}_{n=1}^\infty$ by

$$a_1 := \frac{1}{2} \min\{r_1, 1\}, \quad a_{n+1} = \frac{1}{2} \min\{r_{n+1}, a_n\} \quad \text{for } n = 2, 3, \dots$$

so that $\pi(B(\gamma_n, a_n)) \subset U_n$ for every n . We let Γ_n be the path

$$\Gamma_n = \gamma_{\varepsilon_n v_2} \star \gamma_{nv_1} \star \gamma_{-\varepsilon_n v_2} \star \gamma_{-nv_1}, \quad \text{where } \varepsilon_n = \frac{a_n}{6}.$$

For every n , the path Γ_n is tree-reduced by [Lemma 2.1.3](#) and has length $\|\Gamma_n\|_{1\text{-var};[0,T]} = 2n + 2\varepsilon_n$. By the same argument in the proof of [Proposition 2.1.5](#) these properties yield that $F = \cup_{n=1}^\infty \{\Gamma_n\}$ is closed in $(\mathcal{C}_1, \chi_\tau)$. It is also clear that $[o] \notin F$. On the other hand, we can obtain the estimate

$$\|\Gamma_n - \gamma_n\|_{1\text{-var};[0,T]} < 6\varepsilon_n = a_n. \quad (2.6)$$

The collection $\{U_n\}_{n=1}^\infty$ is a neighbourhood basis at $[o]$, and therefore any set U in χ_τ which contains $[o]$ must contain U_k as a subset for some k . Using the estimate [Eq. \(2.6\)](#) above and the definition of the sequence $\{a_n\}_{n=1}^\infty$, we have that $\Gamma_k \in B(\gamma_k, a_k)$ and therefore $[\Gamma_k] \in U_k$. In other words, we have shown that any neighbourhood of $[o]$ must have non-empty intersection with F and therefore $(\mathcal{C}_1, \chi_\tau)$ cannot be regular.

To prove that $\mathcal{d}$ is not complete, we define another sequence of paths

$$\rho_n = \gamma_{\frac{v_2}{n}} \star \gamma_{v_1} \star \gamma_{-\frac{v_2}{n}} \star \gamma_{-v_1} \quad \text{for } n \in \mathbb{N}.$$

Again ρ_n is tree-reduced, i.e. $\rho_n^* = \rho_n$, and there exists a constant $c > 0$ such that for any $1 \leq n < m$ we have

$$\mathcal{d}([\rho_n], [\rho_m]) = \|\rho_n - \rho_m\|_{1\text{-var};[0,T]} \leq \frac{c}{n} \quad (2.7)$$

so that $([\rho_n])_{n=1}^\infty$ is a $\mathcal{d}$ -Cauchy sequence. However, because $\mathcal{S}(\rho_n)^k \rightarrow 0$ as $n \rightarrow \infty$ for every $k \in \mathbb{N}$ and since $\chi_{\text{pr}} \subset \chi_{\mathcal{d}}$, the only possible limit point of the sequence is $[o]$. This cannot happen owing to

$$\mathcal{d}([o], [\rho_n]) \geq 2,$$

which holds for all n . The proof that the topology χ_d is completely metrisable is more involved. Recall the following topological facts:

- 1) A subspace of a completely metrisable space is completely metrisable if and only if it is G_δ [Kec95, Theorem 3.11];
- 2) For a topological space X , the set of continuity points of a function $f : X \rightarrow \mathbb{R}$ is G_δ [Kec95, Proposition 3.6];
- 3) A closed subset of a metric space is G_δ [Kec95, Proposition 3.7];
- 4) The intersection of two G_δ sets is G_δ .

By Lemma 1.0.36, $C^{0,1\text{-var}}([0, T], V)$ is completely metrisable with the set of piecewise linear paths forming a dense subset. We split the proof into two components:

- 1) The set of continuity points of the map $\gamma \mapsto \|\gamma^*\|_{1\text{-var};[0,T]}$ is the set of tree-reduced paths;
- 2) The set of paths parameterised at constant speed is closed in $C^{0,1\text{-var}}([0, T], V)$.

For the first item, suppose that γ is tree-reduced with $\gamma_n \rightarrow \gamma$ in 1-variation, implying $\|\gamma_n\|_{1\text{-var};[0,T]} \rightarrow \|\gamma\|_{1\text{-var};[0,T]}$. Suppose for a contradiction that $\|\gamma_n^*\|_{1\text{-var};[0,T]}$ does not converge to $\|\gamma\|_{1\text{-var};[0,T]}$; then there exists a subsequence (which we also denote by γ_n^*) for which $\|\gamma_n^*\|_{1\text{-var};[0,T]} \rightarrow \kappa \geq 0$ where $\kappa < \|\gamma^*\|_{1\text{-var};[0,T]}$. Since the sequence $(\gamma_n^*)_{n=1}^\infty$ is equicontinuous and bounded, it holds by Arzelà-Ascoli that there exists a uniformly convergent subsequence converging to a path $\tilde{\gamma}$ with finite 1-variation, for which $\gamma_n^* \rightarrow \tilde{\gamma}$ in $WG\Omega_p(V)$ for some $p \in (1, 2)$. From [FV10, Lemma 1.15], we have the bound

$$\|\tilde{\gamma}\|_{1\text{-var};[0,T]} \leq \liminf_{n \rightarrow \infty} \|\gamma_n^*\|_{1\text{-var};[0,T]} = \kappa.$$

By continuity of the signature map $\mathcal{S} : WG\Omega_p(V) \rightarrow T((V))$ (with $T((V))$ equipped with the product topology), we have the following

$$\mathcal{S}(\gamma) = \lim_{n \rightarrow \infty} \mathcal{S}(\gamma_n) = \lim_{n \rightarrow \infty} \mathcal{S}(\gamma_n^*) = \mathcal{S}(\tilde{\gamma}).$$

And so $\tilde{\gamma} \in [\gamma]$, implying that γ does not have shortest length in its equivalence class: a contradiction. For the reverse, suppose that $\gamma \in C^{0,1\text{-var}}([0, T], G^{(1)})$ is not tree-reduced. By density, we can approximate γ in 1-variation by a sequence $(\gamma_n)_{n=1}^\infty$ of piecewise linear paths. Replacing any re-traced pieces with suitable approximations via axis paths, this sequence can be assumed to be tree-reduced.

It follows that

$$\lim_{n \rightarrow \infty} \|\gamma_n^*\|_{1\text{-var};[0,T]} = \lim_{n \rightarrow \infty} \|\gamma_n\|_{1\text{-var};[0,T]} = \|\gamma\|_{1\text{-var};[0,T]} > \|\gamma^*\|_{1\text{-var};[0,T]},$$

so that the map is discontinuous at any path which is not tree-reduced.

For the second part, let $(\gamma_n)_{n=1}^\infty$ be a collection of paths parameterised at constant speed, which we recall is defined by $\|\gamma\|_{1\text{-var};[0,t]} = \frac{t}{T} \|\gamma\|_{1\text{-var};[0,T]}$. Suppose that $\gamma_n \rightarrow \gamma$ in $C^{0,1\text{-var}}([0, T], V)$, so that $\|\gamma_n\|_{1\text{-var}} \rightarrow \|\gamma\|_{1\text{-var}}$. It must also hold that $\gamma_n|_{[0,t]}$ converges to $\gamma|_{[0,t]}$ in $C^{0,1\text{-var}}([0, t], V)$. We conclude by taking the limits

$$\|\gamma\|_{1\text{-var};[0,t]} = \lim_{n \rightarrow \infty} \|\gamma_n\|_{1\text{-var};[0,t]} = \lim_{n \rightarrow \infty} \frac{t}{T} \|\gamma_n\|_{1\text{-var};[0,T]} = \frac{t}{T} \|\gamma\|_{1\text{-var};[0,T]}.$$

□

Since $\mathcal{C}_1$ with χ_d is separable and completely metrisable we have the following.

Corollary 2.1.11. *The space $(\mathcal{C}_1, \chi_d)$ is Polish.*

Remark 2.1.12. *The strict inclusions demonstrated in [Proposition 2.1.5](#) can now be concluded from the lack of metrisability of χ_τ , the metrisability of χ_m , and the inclusions $\chi_m \subseteq \chi_\tau$.*

Remark 2.1.13. *As noted in the introduction, there are other definitions of unparameterised path space. For example, [\[CO22\]](#) considers the set of paths $C^{1\text{-var}} := \bigcup_{T>0} C^{1\text{-var}}([0, T], V)$. For any γ in $C^{1\text{-var}}$, denote by $[0, T_\gamma]$ the interval on which γ is parameterised. In this case, $C^{1\text{-var}}$ may be equipped with the pseudo-metric*

$$\tilde{d}(\gamma, \rho) := \inf_{\tau: [0, T_\gamma] \rightarrow [0, T_\rho]} \|\gamma - \rho \circ \tau\|_{1\text{-var};[0, T_\gamma]},$$

where the reparameterisations τ are required to be strictly increasing surjections. The resulting space of unparameterised paths is the quotient space $\mathcal{C}(1) := C^{1\text{-var}} / \sim_\tau$, endowed with the quotient topology $\chi_{\tilde{d}}$, derived from the topology on $\mathcal{C}(1)$ generated by $\tilde{d}$. The canonical bijection from $(\mathcal{C}_1, \chi_\tau)$ to $(\mathcal{C}(1), \chi_{\tilde{d}})$ is continuous, implying that $\chi_{\tilde{d}}$ is no stronger than χ_τ . Moreover, the techniques from the proof of [Theorem 2.1.10](#) can be employed to conclude that $\chi_{\tilde{d}}$ is also not metrisable.

We now generalise many of the results for χ_{pr} to a broader class of topologies which are of the form χ_m with an additional continuity criterion. We first require the following lemma.

Lemma 2.1.14. *Every open set in χ_τ is unbounded in d .*

Proof. Let $U \in \chi_\tau$ be any open neighbourhood of some point $[\gamma]$ in $\mathcal{C}_1$. The tree-reduced representative γ^* is in $\pi^{-1}(U)$, and so [FV10, Corollary 1.34] and Chen's identities imply that $\pi^{-1}(U)$ contains an irreducible (and hence tree-reduced [LX17]) piecewise linear path parameterised at constant speed, which we again denote by γ^* for convenience. Let $\{v_1, v_2\}$ be a pair of orthonormal vectors not collinear to the last linear piece of γ^* . For each $\varepsilon > 0$, define the sequences of paths $(\sigma_{n,\varepsilon})_{n=1}^\infty$ and $(\rho_{n,\varepsilon})_{n=1}^\infty$ by

$$\begin{aligned}\sigma_{n,\varepsilon} &= \gamma^* \star \gamma_{\varepsilon v_2} \star \gamma_{nv_1} \star \gamma_{-\varepsilon v_2} \star \gamma_{-nv_1}, \\ \rho_{n,\varepsilon} &= \gamma^* \star \gamma_{(n+\varepsilon)v_1} \star \gamma_{-(n+\varepsilon)v_1}.\end{aligned}$$

For every n and ε , $\rho_{n,\varepsilon}$ is in the equivalence class $[\gamma^*]$, which itself is a subset of $\pi^{-1}(U)$. By construction

$$\begin{aligned}\|\sigma_{n,\varepsilon} - \rho_{n,\varepsilon}\|_{1\text{-var};[0,T]} &= \|\gamma^* - \gamma^*\|_{1\text{-var};[0,T]} \\ &\quad + \|\gamma_{\varepsilon v_2} \star \gamma_{nv_1} \star \gamma_{-\varepsilon v_2} \star \gamma_{-nv_1} - \gamma_{(n+\varepsilon)v_1} \star \gamma_{-(n+\varepsilon)v_1}\|_{1\text{-var};[0,T]} \\ &\leq 0 + 4\varepsilon.\end{aligned}$$

Hence for every n , $\sigma_{n,\varepsilon}$ lies in $\pi^{-1}(U)$ for some $\varepsilon > 0$. To conclude, we note that $\sigma_{n,\varepsilon}$ is tree-reduced by [LX17], with $d([\sigma_{n,\varepsilon}], [o]) \geq 2n$. $\square$

Corollary 2.1.15. *Assume that χ_m satisfies the following condition: there exists some $p \in (1, 2)$ for which*

$$\mathcal{S} : (C^{1\text{-var}}([0, T], V), d_{p\text{-var}}^{(1)}) \rightarrow (\mathcal{C}_1, \chi_m) \quad (2.8)$$

is continuous. Then the topological space $(\mathcal{C}_1, \chi_m)$ is

- 1) *Separable and metrisable;*
- 2) *Not a Baire space, so neither Polish nor locally compact;*
- 3) *σ -compact.*

Remark 2.1.16. *The condition in Eq. (2.8) translates to the following: if $(\gamma_n)_{n=1}^\infty$ is a sequence of continuous paths of bounded variation which converges to a path $\gamma \in C^{1\text{-var}}([0, T], V)$ with respect to the topology generated by the p -variation metric for some $p \in (1, 2)$ then the sequence $([\gamma_n])_{n=0}^\infty$ should converge to $[\gamma]$ in $(\mathcal{C}_1, \chi_m)$. It is a consequence of various fundamental results in the theory of rough paths that χ_{pr} satisfies this condition.*

Proof. The separability of χ_m follows from the inclusion $\chi_m \subset \chi_\tau$ and separability of the question topology. The same arguments found in Proposition 2.2.2 show

the sets $B(r)$ are compact in χ_m , and since every open set in χ_τ is unbounded in $\mathcal{d}$, they have empty interior in χ . An application of the Baire category theorem proves item (2). Item (3) is immediate from the compactness of $B(r)$. $\square$

Example 2.1.17. In the recent paper [CLX24], the authors introduce the general concept of weighted signature kernels, defined through weighted inner products on $T((V))$. This leads to a class of topologies on (a subspace of) $T((V))$ that satisfy the conditions of Corollary 2.1.15. Let $\langle \cdot, \cdot \rangle_n$ be the canonical Hilbert-Schmidt inner product on $V^{\otimes n}$ derived from the inner product $\langle \cdot, \cdot \rangle$ on V . For $\mathbf{x}, \mathbf{y} \in T((V))$, define

$$\langle \mathbf{x}, \mathbf{y} \rangle_\phi := \sum_{n=0}^{\infty} \phi(n) \langle x_n, y_n \rangle_n,$$

with $\phi : \mathbb{N} \cup \{0\} \rightarrow \mathbb{R}_+$ some weight function. Set $T_\phi(V) := \{\mathbf{x} \in T((V)) : \langle \mathbf{x}, \mathbf{x} \rangle_\phi < \infty\}$, a Hilbert space equipped with the norm topology derived from the inner product $\langle \cdot, \cdot \rangle_\phi$. Suppose there exists some $\delta > 0$ for which

$$\sum_{n=0}^{\infty} \phi(n) \frac{C^n}{\Gamma\left(\frac{n}{1+\delta} + 1\right)^2} < \infty, \quad (2.9)$$

for every $C > 0$, where $\Gamma(\cdot)$ denotes the Gamma function. Then [Lyo+07, Theorems 3.7 and 3.10] ensure that $\mathcal{S}_p \subseteq T_\phi(V)$ and that $\mathcal{S} : C^{p\text{-var}}([0, T], V) \rightarrow T_\phi(V)$ is continuous for each $p \in [1, 1 + \delta)$. We note here the slight difference between (2.9) and the summability condition presented in [CLX24, Lemma 2.3]. This is to guarantee suitable continuity of the signature map as required for Corollary 2.1.15. An important example of weight function is the case $\phi(i) = \Gamma\left(\frac{i}{2} + 1\right)$, for which [CLX24] establishes a connection between the expected signature of Stratonovich Brownian motion and the hyperbolic development of a path.

A recent area of study [Boe+21; CL16; CO22; Faw03] is the moment problem for path valued data: to what extent the expected signature of a stochastic process characterises its law. Due to the non-uniqueness of the signature, it is standard to consider the stochastic process as being $\mathcal{C}_1$ valued. The choice of topology on this space is then tied to the class of Borel probability measures. The authors of [CO22] impose a topology of the form found in Example 2.1.17, with $\phi \equiv 1$. However, as illustrated in the aforementioned example, there exists a broad family of induced topologies one may wish to consider, all of which lie inside χ_τ . As such, it is of interest to know whether the results of [CO22] may be extended to this setting. Indeed, the underlying reference necessary to their result [Gil71, Theorem 4.6] requires only complete regularity (or, equivalently, uniformisability)

and not metrisability. To the best of our knowledge, it is not known whether the quotient topology is completely regular. With respect to this question, our proof of non-metrisability does not yield much insight, except that first countability would then imply a lack of (complete) regularity. Alternatively, since χ_τ is Hausdorff, local compactness would imply complete regularity. However, under the assumption of the length conjecture, it is not locally compact. The length conjecture hypothesises that

$$\lim_{n \rightarrow \infty} \|n! \mathcal{S}(\gamma)^n\|^{1/n} = \|\gamma^*\|_{1\text{-var};[0,T]}.$$

The existence of the limit has been shown [CLN18b; BG19b], and the correct limiting value is known in special cases, e.g. when γ^* is C^1 [HL10], or for certain planar paths [BG23]. This hypothesis yields an upper bound on the length of γ^* in terms of asymptotic behaviour of $\mathcal{S}(\gamma^*)$, and so it may be shown that $\pi^{-1}(B(r))$ is closed in $\mathcal{C}_1$. As such the $B(r)$ are closed in χ_τ , and have empty interiors by Lemma 2.1.14. Since χ_τ is Hausdorff, the Baire category theorem implies that the quotient topology would not be locally compact. Finally, continuity of the group operations on $(\mathcal{C}_1, \chi_\tau)$ would show complete regularity, as every Hausdorff topological group is completely regular.

The non-metrisability of the quotient topology may present challenges when used for advanced applications in probability and stochastic analysis. As pointed out in the monograph [Gar18], many probability measures in practice are Borel measures on Polish spaces. This assumption greatly simplifies the analysis in many applications by circumventing the complexities of working with general topological spaces. The development of theory reflects this. Useful tools are available in this setting including the facts that finite Borel measures on Polish spaces are tight (Ulam's theorem), any Borel measurable function on such a space is continuous on a large compact set (Lusin's theorem), and a family of Borel measures is tight if and only if it is relatively compact in the topology of weak convergence (Prokhorov's theorem). In the case of the product topology χ_{pr} , Ulam's theorem, Lusin's theorem, and the sufficiency of tightness for relative compactness in Prokhorov's theorem still hold, whereas all theorems hold for χ_d . In the next section we illustrate how these methods can be used to support an analysis of the expected signature model of [LLN16].

It may still be desirable to identify a Polish topology on $\mathcal{C}_1$ for which $\mathcal{S} : C^{1\text{-var}}([0, T], V) \rightarrow \mathcal{C}_1$ is continuous. Theorem 2.1.10 excludes χ_τ as a possibility, while the signature map $\mathcal{S} : C^{1\text{-var}}([0, T], V) \rightarrow (\mathcal{C}_1, \chi_d)$ is not continuous. However, χ_{pr} can be metrised by using any metric which induces the product

topology on $T((V))$. Taking the completion of $\mathcal{C}_1$ with respect to any such metric is a direct way to generate a Polish space, where the completion is easily identifiable.

Proposition 2.1.18. *Let ρ be any metric on that induces the product topology on $T((V))$, and let $\rho|_{\mathcal{S}_1}$ denote its restriction to $\mathcal{S}_1$. Then, the set of group-like elements $G^{(*)}$ equipped with the metric $\rho|_{G^{(*)}}$ is a valid $\rho|_{\mathcal{S}_1}$ -completion of $(\mathcal{C}_1, \chi_{pr})$.*

Proof. It is enough to prove the following: the set $G^{(*)}$ is closed in $T((V))$ with respect to this topology and $\mathcal{S}_1 \subseteq G^{(*)} = \bar{\mathcal{S}}_1$. To see that $G^{(*)}$ is closed we show that the limit $\mathbf{x}$ of any convergent sequence $(\mathbf{x}_m)_{m=1}^\infty$ in $G^{(*)}$ is again in $G^{(*)}$. Let $\mathbf{x}_m = (x_{m,0}, x_{m,1}, \dots, x_{m,j}, \dots)$, the convergence $\lim_{m \rightarrow \infty} \mathbf{x}_m = \mathbf{x}$ in the product topology holds if and only if $x_{m,j} \rightarrow x_j$ in $V^{\otimes j}$ as $m \rightarrow \infty$ for every $j = 0, 1, 2, \dots$. For any n it follows that $\pi^{(n)} \mathbf{x}_m \rightarrow \pi^{(n)} \mathbf{x}$ as $m \rightarrow \infty$, and since it is known that $G^{(n)} = \pi^{(n)}(G^{(*)})$ is closed in the truncated tensor algebra, see for example [FV10, Theorem 7.30] or [Lyo+07, Proposition 2.25], it follows that $\pi^{(n)} \mathbf{x}$ is also in $G^{(n)}$. By definition this implies $\mathbf{x}$ is then in $G^{(*)}$. To see the final statement, it is clear by Proposition 1.0.19 that $\mathcal{S}_1 \subseteq G^{(*)}$. Then if $\mathbf{x}$ is in $G^{(*)}$, for every n there exists by definition γ_n such that $\mathcal{S}^{(n)}(\gamma_n) = \pi_n \mathbf{x}$. The sequence determined by $\mathbf{x}_n = \mathcal{S}(\gamma_n)$ in $\mathcal{S}$ then converges to $\mathbf{x}$ in $T((V))$. It follows that $G^{(*)} = \bar{\mathcal{S}}_1$. $\square$

Remark 2.1.19. *It is well known that the inclusion $\mathcal{S}_1 \subset G^{(*)}$ is in fact strict, and the results of Section 2.1 provide an alternative (non-constructive) proof. Indeed, $G^{(*)}$ is closed in $T((V))$ whereas $\mathcal{S}_1$ cannot be since it is not completely metrisable.*

2.2 Uniform approximation and the expected signature model

The contemporary use of the signature in applications to regression analysis, as well as its wider use in statistical and machine learning applications, is in large part underpinned by the following fundamental result. The result has appeared in various guises in the literature; perhaps most notably in [LLN16], but we will use the statement found in [CS24].

Theorem 2.2.1 (The Fundamental Theorem of Uniform Approximation by the Signature). *Let $(\mathcal{C}_1, \chi)$ be the space of unparameterised paths equipped with a topology χ . Assume that the K is a compact subset of $\mathcal{C}_1$ and for $\ell \in T((V))^*$ denote by Φ_ℓ the function given by*

$$\begin{aligned} \Phi_\ell : \mathcal{C}_1 &\rightarrow \mathbb{R}; \\ [\gamma] &\mapsto \ell \circ \mathcal{S}([\gamma]). \end{aligned}$$

Let $C(K)$ denote the space of continuous functions on K with the topology of uniform convergence. If the set $\Phi_K := \{\Phi_\ell|_K : \ell \in T((V))^*\}$ is contained in $C(K)$, then it is a dense subset.

While this result appears in numerous works, the choice of topology is rarely addressed explicitly; an exception as noted earlier is the paper [CO22]. The matter of which space to work with in the context of Theorem 2.2.1 merits some consideration. Our aim is not to advocate for any particular choice, but to audit our selected choice of candidates. We will prioritise three questions:

- 1) How readily can compact subspaces of the chosen topology be found?
- 2) Given a compact subspace, how easy is it to exhibit continuous functions on these subspaces?
- 3) To what extent do these pairs of sets (compact subspaces and their set of continuous functions) relate to the practical use of Theorem 2.2.1?

We address these points below. Our focus will be mainly on χ_m and specifically the product topology, where it is simplest to provide positive answers to these questions.

2.2.1 Compactness

The inclusion maps $\iota_1 : (\mathcal{C}_1, \chi_\tau) \rightarrow (\mathcal{C}_1, \chi_m)$ and $\iota_2 : (\mathcal{C}_1, \chi_d) \rightarrow (\mathcal{C}_1, \chi_\tau)$ are continuous by Proposition 2.1.5; compact subsets of $(\mathcal{C}_1, \chi_m)$ will therefore be at least as plentiful as for the alternatives. The next result describes some explicit examples of such sets. In the context of geometric rough paths, the following two results may also be found in [CL16, Proposition 4.7]. We translate these proofs into the context and language of topologies on unparameterised path space.

Proposition 2.2.2. *For any $r > 0$, the set*

$$B(r) = \{[\gamma] \in \mathcal{C}_1 : \|\gamma^*\|_1 \leq r\}$$

is a compact subspace of $(\mathcal{C}_1, \chi_m)$, provided χ_m satisfies the conditions of Corollary 2.1.15. Under these conditions, $(\mathcal{C}_1, \chi_m)$ is σ -compact.

Proof. Since χ_m is metrisable, it suffices to prove sequential compactness. Let $([\gamma_n])_{n=1}^\infty$ be a sequence in $B(r)$. Then, by definition of the path γ^* and the set $B(r)$, the sequence $(\gamma_n^*)_{n=1}^\infty$ satisfies

$$|\gamma_n^*(t) - \gamma_n^*(s)| \leq \frac{r}{T} (t - s) \text{ for all } s \leq t \text{ in } [0, T] \text{ and } n \in \mathbb{N}; \quad (2.10)$$

that is, the paths are equicontinuous. The Arzelà-Ascoli theorem gives a uniformly convergent subsequence which we again call $(\gamma_n^*)_{n=1}^\infty$ for convenience. We write γ for the limit and note from Eq. (2.10) that $\|\gamma\|_{1\text{-var};[0,T]} \leq r$ and therefore $\|\gamma^*\|_{1\text{-var};[0,T]} \leq r$ so that $[\gamma]$ is in $B(r)$. Let $1 < p < 2$, then a standard inequality gives

$$\|\gamma_n^* - \gamma_m^*\|_{p\text{-var};[0,T]} \leq 2r^{1/p} \|\gamma_n^* - \gamma_m^*\|_\infty^{1-1/p}$$

allowing us to conclude that $(\gamma_n^*)_{n=1}^\infty$ is a Cauchy sequence and thus convergent to γ in p -variation. The continuity conditions of Corollary 2.1.15 then ensures that $[\gamma_n] \rightarrow \gamma$ in χ_m . $\square$

The following corollaries are direct consequences of σ -compactness.

Corollary 2.2.3. *The range of the signature $\mathcal{S}_1$ is measurable with respect to the Borel sigma-algebra on $G^{(*)}$, if the latter is equipped with the subspace topology of the product topology on $T((V))$.*

Proof. Since $G^{(*)}$ is metrisable and each $B(r)$ is closed, each $B(r)$ will be metrisable. Thus $\mathcal{S}_1$ is a countable union of measurable sets, hence it too is measurable. $\square$

Corollary 2.2.4. *Let $\mathcal{C}_1$ be equipped with a topology satisfying the conditions of Corollary 2.1.15. Then $\mathcal{C}_1$ is a Lusin space.*

Proof. Every compact metric space is Lusin, and a countable union of Lusin subspaces of a metric space is a Lusin space. $\square$

Remark 2.2.5. *When $\mathcal{C}_1$ is equipped with χ_m , it is not difficult to exhibit compact subsets $K \subset \mathcal{C}_1$ which are not contained in any $B(r)$, for $r > 0$. For instance, by using the example of Proposition 2.1.5, the set $K = \cup_{k=1}^\infty [\Gamma_k] \cup [o]$ is compact since any (non-trivial) sequence has a subsequence converging to $[o]$ in the product topology. On the other hand, each of the tree-reduced paths Γ_k has length 2^{k+1} .*

An important example of a function on $\mathcal{C}_1$ is the solution of a controlled differential equation. Suppose that W is a finite-dimensional vector space and let $\mathcal{T}(W)$ denote the space of smooth vector fields on W . After fixing a bases for V and W , for each $i = 1, \dots, \dim(V)$ let $f_i \in \mathcal{T}(W)$ and write $f : V \rightarrow \mathcal{T}(W)$ for the linear map defined by $f_v = \sum_{i=1}^{\dim(V)} v^i f_i$. We can then uniquely solve

$$dy_t = f_{d\gamma_t}(y_t), \text{ started at } y_0. \quad (2.11)$$

We briefly review how to extend $f : V \rightarrow \mathcal{T}(W)$ to a linear map $f : T(V) \rightarrow \mathcal{T}(W)$. We associate with each word $w \in \mathcal{W}_{\dim(V)}$ a vector field f_w defined recursively by

$f_\emptyset = \text{id}$, and for $\mathbf{w} = w_1 \dots w_n$ by the recursion

$$f_{\mathbf{w}} = f_{w_1} \triangleright f_{w_2 \dots w_n},$$

where for smooth $\varphi : \mathbb{R}^e \rightarrow \mathbb{R}$

$$(f \triangleright g)\varphi(y) := \sum_{i,j=1}^{\dim(W)} f^i(y) (\partial_i g^j(y)) \partial_j \varphi(y).$$

We can then view $(f \triangleright g)$ acting on smooth $\varphi : W \rightarrow \mathbb{R}$ by considering its action coordinate-wise.

Example 2.2.6. Let $V = \mathbb{R}^d$ and $W = \mathbb{R}^e$ and suppose that $f(y) = Ay$ and $g(y) = By$ for matrices A and B and $\varphi = \text{id}$. Consider the application on the k^{th} coordinate, that is

$$\begin{aligned} (f \triangleright g)\varphi^k(y) &= \sum_{i,j=1}^e \left(\sum_{p=1}^e A_{ip} y^p \right) \left(\partial_i \sum_{q=1}^e B_{jq} y^q \right) \partial_j y^k \\ &= \sum_{i=1}^e \sum_{p=1}^e A_{ip} y^p B_{ki} \\ &= (BAy)^k, \end{aligned}$$

So that $(f \triangleright g)\varphi(y) = BAy$.

The resulting family of maps $\{\mathbf{w} \mapsto f_{\mathbf{w}}(\cdot) \in \mathcal{T}(W) : \mathbf{w} \in \mathcal{W}_{\dim(V)}\}$ may be represented as a single linear map $f : T(V) \rightarrow \mathcal{T}(W)$ defined by

$$\mathbf{x} \mapsto \langle f(\cdot), \mathbf{x} \rangle := f_{\mathbf{x}}(\cdot),$$

where the pairing amounts to viewing f as a formal sum $\sum_{\mathbf{w} \in \mathcal{W}_{\dim(V)}} f_{\mathbf{w}} e_{\mathbf{w}}$. In the following, $f_{\mathbf{w}}(y)$ will be taken to mean $f_{\mathbf{w}\text{id}}(y)$. We are now able to prove the following.

Proposition 2.2.7. Suppose that W is a finite-dimensional vector space. Let $f : V \rightarrow \mathcal{T}(W)$ be linear, and assume that there exists $C < \infty$ such that for every $k \in \mathbb{N}$

$$\sup_{\substack{\mathbf{x} \in V^{\otimes k} \\ \|\mathbf{x}\|=1}} \sup_{y \in W} \|f_{\mathbf{x}}(y)\| \leq C^k. \quad (2.12)$$

Define a function (the Itô-map) by

$$\Psi_{y_0, f} : C^{1\text{-var}}([0, T], V) \rightarrow W, \quad \gamma \mapsto y_T$$

where y is the unique solution over $[0, T]$ of the differential equation [Eq. \(2.11\)](#). Then

- 1) The function $\Psi_{y_0, f}$ is constant on every equivalence class of $\sim_\tau$ and for every γ in $C^{1-var}([0, T], V)$, $\Psi_{y_0, f}(\gamma)$ is given by the convergent series

$$\Psi_{y_0, f}(\gamma) = y_0 + \sum_{k=1}^{\infty} f_{\mathcal{S}(\gamma)^k}(y_0). \quad (2.13)$$

- 2) The Itô-map is a well-defined function from $\mathcal{C}_1$ into W . For every $r > 0$, the restriction of this function to $B(r)$ is continuous with respect to the topology χ_{pr} on $\mathcal{C}_1$.

Proof. The signature is invariant on the tree-like equivalence classes and hence so will be the function $\Psi_{y_0, f}$ once we prove [Eq. \(2.13\)](#). To see this, we observe first by iterated use of the change-of-variable formula that for any $N \geq 1$ we have

$$\Psi_{y_0, f}(\gamma) = \Psi_{y_0, f}^N(\gamma) + \int_{0 < t_1 < \dots < t_{N+1} < 1} f_{d\gamma_{t_1} d\gamma_{t_2} \dots d\gamma_{t_N} d\gamma_{t_{N+1}}}(y_{t_1}),$$

where

$$\Psi_{y_0, f}^N(\gamma) := y_0 + \sum_{k=1}^N f_{\mathcal{S}(\gamma)^k}(y_0).$$

Condition [Eq. \(2.12\)](#) ensures

$$\left| \Psi_{y_0, f}(\gamma) - \Psi_{y_0, f}^N(\gamma) \right| \leq \frac{C^{N+1} \|\gamma\|_{1-var; [0, T]}^{N+1}}{(N+1)!},$$

so that indeed [Eq. \(2.13\)](#) holds.

Abusing notation we still use $\Psi_{y_0, f}$ to denote the well-defined function on $\mathcal{C}_1$ defined by $\Psi_{y_0, f} : [\gamma] \mapsto \Psi_{y_0, f}(\gamma)$. To prove continuity of $\Psi_{y_0, f}$ on $B(r)$ we take a sequence $[\gamma_n] \rightarrow [\gamma]$ in $B(r)$ in the product topology, and then note that for any $N \geq 1$

$$\left| \Psi_{y_0, f}([\gamma]) - \Psi_{y_0, f}([\gamma_n]) \right| \leq \sum_{k=1}^N \left| f_{\mathcal{S}(\gamma^*)^k}^{(k)}(y_0) - f_{\mathcal{S}(\gamma_n^*)^k}^{(k)}(y_0) \right| + \frac{2C^{N+1}r^{N+1}}{(N+1)!}.$$

Letting $n \rightarrow \infty$ and then $N \rightarrow \infty$ we learn that $\Psi_{y_0, f}([\gamma_n]) \rightarrow \Psi_{y_0, f}([\gamma])$ as $n \rightarrow \infty$. $\square$

2.2.2 Revisiting the expected signature model

The chief motivation of [Theorem 2.2.1](#) is to provide a theoretical justification for the so-called expected signature model introduced by Levin, Lyons and Ni in [\[LLN16\]](#).

To explore the relation between this model and the foundations developed above, suppose that Γ is a random variable taking values in the space $(\mathcal{S}_1, \chi_{\text{pr}})$. Let Y be another random variable, defined on the same probability space as Γ , which takes values in a finite dimensional vector space W . The goal of regression analysis here is to learn the conditional expectation $\mathbb{E}[Y|\Gamma]$, where Y is interpreted as the response of some system to the input Γ . In other words, we want to approximate the Borel-measurable function $f : \mathcal{S}_1 \rightarrow W$ defined by

$$f(g) = \mathbb{E}[Y|\Gamma = g]. \quad (2.14)$$

Suppose the law of Γ , a Borel probability measure on $\mathcal{S}_1$, is denoted by μ . Then as we will see in [Corollary 2.2.9](#), by a version of Lusin's Theorem [[Bog07](#)], the function f in [Eq. \(2.14\)](#) is almost continuous in the sense that for any $\delta > 0$ there exists a compact set $K = K_\delta$ such that $\mu(\mathcal{S}_1 \setminus K) < \delta$ and such that f is continuous on K . By [Theorem 2.2.1](#), it is reasonable to adopt the model

$$Y = L(\Gamma) + \varepsilon,$$

where L is the restriction to $\mathcal{S}$ of a linear function from $T((V))$ to W and ε is a W -valued random variable satisfying $\mathbb{E}[\varepsilon|\Gamma] = 0$. This obviates the need to find an explicit compact set and continuous function relating the independent and dependent variables.

The typical case is where $\Gamma = \mathcal{S}(\gamma)$ for some stochastic process γ in $C^{1\text{-var}}([0, T], V)$, so that the probability measure $\mu = \mathcal{S}_* \mu_\gamma$ is the push-forward of the law of γ under the signature map and where Y is the solution to an ordinary differential equation driven by γ . The following result describes conditions for the well-posedness and measurability of the functions on $\mathcal{C}_1 \rightarrow W$ which result from this construction.

Proposition 2.2.8. *Let W be a finite-dimensional vector space. Suppose that $f : V \rightarrow \mathcal{T}(W)$ is a linear function which satisfies the following conditions:*

- 1) *For every v in V , $f_v : W \rightarrow W$ is continuous with at most linear growth.*
- 2) *For every $R > 0$ there exists a finite positive $C = C(R)$ such that for every k in $\mathbb{N}$ the following bound holds*

$$\sup_{\substack{\mathbf{x} \in V^{\otimes k} \\ \|\mathbf{x}\|=1}} \sup_{|y| \leq R} \|f_{\mathbf{x}}(y)\| \leq C(R)^k.$$

Then the Itô map $\gamma \mapsto y_T$ derived from the ordinary differential equation $dy_t = f_{d\gamma_t}(y_t)$ starting at y_0 is invariant with respect to the tree-like equivalence relation

on $C^{1-var}([0, T], V)$ and induces a well-defined Borel measurable function $\Psi_{y_0, f} : \mathcal{C}_1 \rightarrow W$, where $\mathcal{C}_1$ is equipped with any of the candidate topologies. Additionally for χ_{pr} and χ_d , there exists a Borel measurable function $\bar{\Psi}_{y_0, f} : \bar{\mathcal{C}}_1 \rightarrow W$ which agrees with $\Psi_{y_0, f}$ on $\mathcal{C}_1$.

Proof. The ordinary differential equation has a unique solution for every γ in $C^{1-var}([0, T], V)$. By Theorem 3.7 of [FV10] and the linear growth condition, there is a finite positive constant $K = K(|y_0|, \|\gamma\|_{1-var;[0, T]})$ such that

$$\|y\|_\infty \leq K(|y_0|, \|\gamma\|_{1-var;[0, T]}).$$

A repetition of the same argument in the proof of Proposition 2.2.7 leads to the estimate

$$y_T = y_0 + \sum_{k=1}^N f_{S(\gamma)^k}(y_0) + E_{N+1},$$

$$\text{where } |E_N| \leq \frac{C \left(K(|y_0|, \|\gamma\|_{1-var;[0, T]}) \right)^N \|\gamma\|_{1-var;[0, T]}^N}{N!},$$

which allows one to deduce invariance on the equivalence classes by taking the limit $N \rightarrow \infty$. To see that the resulting function $\Psi_{y_0, f} : \mathcal{C}_1 \rightarrow W$ is measurable with respect to $\mathcal{B}(\mathcal{C}_1)$ (the Borel sigma algebra of $\mathcal{C}_1$) and $\mathcal{B}(W)$, we notice that it is the pointwise limit of the functions $L_N \circ \mathcal{S}$ where L_N denotes the (restriction of the) linear function

$$L_N : (a_0, a_1, \dots) \mapsto y_0 + \sum_{k=1}^N f_{a_k}(y_0) \in W$$

to $\mathcal{S}$. The existence of a measurable extension follows from the fact that $(W, \mathcal{B}(W))$ is a Polish space; see [Sho83, Theorem 1] noting that $\mathcal{B}(\mathcal{C}_1) = \{A \cap \mathcal{C}_1 : A \in \mathcal{B}(\bar{\mathcal{C}}_1)\}$. $\square$

A general version of Lusin's theorem yields the following corollary for the product topology χ_{pr} .

Corollary 2.2.9. *Consider the setup as in Proposition 2.2.8 and let μ be a Borel probability measure on $(\mathcal{C}_1, \chi_{pr})$. Then for every $\delta > 0$, there is a compact set $K \subset \mathcal{C}_1$ such that $\mu(\mathcal{C}_1 \setminus K) < \delta$ and $\Psi_{y_0, f}|_K$ is continuous.*

Proof. Since $\mathcal{C}_1$ is σ -compact with respect to χ_{pr} , μ is tight. Separability of $(\mathcal{C}_1, \chi_{pr})$ and [Bog07, Proposition 7.2.2] imply that μ is a Radon measure. Measurability of $\Psi_{y_0, f}$ and Lusin's Theorem for finite Radon measures [Bog07, Theorem 7.1.13] show the existence of compact sets with the desired property. $\square$

2.3 A comment on open problems

Relationships exist between the results of this chapter and some conjectures in the field of rough analysis. For example (as was mentioned in [Section 2.1.2](#)) a positive answer to the length conjecture would imply the quotient topology is not locally compact as the balls $B(r)$ would be closed.

A deeper example where a careful understanding of the topology is relevant arises in the conjecture of Levin and Lyons [[LL09](#)] on the construction of a signed measure on unparameterised path space associated to higher order differential operators. This conjecture proposes to extend the correspondence between second-order differential operators and (the laws of) Markov diffusions to higher order operators. The classical approach works by associating to a second order differential operator L , a Markov semigroup $(P_t)_{t \geq 0}$ with $P_t := e^{tL}$ and a Markov process $(X_t^x)_{t \geq 0}$ satisfying the property

$$P_t f(x) = \mathbb{E}[f(X_t^x)].$$

The function $u_t := P_t f(x)$ then solves $\partial_t u = Lu$ with $u|_{t=0} = f$. It is well known that the Markov process corresponding $L = \frac{1}{2}\Delta$ is Brownian motion (or Wiener measure). In the case of higher order operators (even if they are constant coefficient and translation invariant) one encounters several problems, most notably that the resulting semigroup $(P_t)_{t \geq 0}$ is no longer positivity-preserving. As such, one cannot hope to associate a probability measure to such a differential operator. Nevertheless, this can be achieved in some cases if one is prepared to work with signed-measures, as noticed in an old work of Krylov [[Kry60](#)] and later by Hochberg [[Hoc78](#)]. Even in the absence of positivity for P_t , [[LL09](#)] shows how the semigroup property can be used to define a consistent family of signed measures $\{\mu_D\}$ on path space indexed by a partition D of $[0, T]$. Even though each resulting signed measure μ_D has unit mass, the total variation $\|\mu_D\|_{TV}$ is unbounded as the mesh size goes to zero. This observation would seem to preclude the existence of a well-defined measure on classical path space $\mu = \lim_{\text{mesh}(D) \rightarrow 0} \mu_D$. The work of Levin and Lyons however, shows the existence of

$$\lim_{\text{mesh}(D) \rightarrow 0} \int \mathcal{S}(\gamma^D)^n d\mu_D(\gamma)$$

for all terms in the signature, where γ^D is the piecewise linear interpolation of γ on the partition D . Moreover, these limits may be computed explicitly in terms of the differential operator. In the special case of the half Laplacian, the corresponding limit is exactly the expected signature of Brownian motion.

By recalling that the coordinate iterated integrals are the monomials on *unparameterised* path space, Levin and Lyons conjecture that there exists a signed measure μ on unparameterised path space such that μ , in an appropriate sense, is the weak limit of the μ_D . Conjectured in 2009, little progress has been made save for the work of Nejad [Nej18]. Part of the difficulty being the absence of results in the literature describing compactness criteria for a family of measures on unparameterised path space. We hope that the results of this chapter will reinvigorate activity in this direction.

2.4 Topologies on unparameterised rough path space

It is natural to ask whether the results of Section 2.1 may be extended to topologies on unparameterised weakly-geometric rough path space; the most general setting under which the uniqueness of the signature is known. While the majority of the results may be extended without too much difficulty, certain subtleties arise when dealing with rough paths as opposed to bounded variation paths.

Indeed, the p -variation norm of the concatenation of two paths is only sub-additive (see Lemma 1.0.44) whereas the 1-variation norm is additive. Moreover, the function $\omega_{\mathbb{X},p}(s, t)$ is only super-additive as opposed to additive whenever $p > 1$. The lack of additivity of these functions, combined with the fact that $WG\Omega_p(V)$ is not a linear space for $p \geq 2$, means that deriving inequalities of the type Eq. (2.7) may become substantially more challenging.

While there exists a tree-reduced representative of each equivalence class whose parameterisation may be fixed as the constant Hölder-control parameterisation, there is no characterisation of the representative in terms of minimal p -variation. Indeed, consider the following example as found in [Dri13].

Example 2.4.1. Consider $V \equiv \mathbb{R}^2$, $p \in (1, 2)$ and $\varepsilon > 0$ suitably small. Let $(e_i)_{i=1}^2$ be the Cartesian basis of V and define

$$\gamma = \gamma_{(1+\varepsilon)e_1} \star \gamma_{-2\varepsilon e_1} \star \gamma_{(1+\varepsilon)e_1}.$$

Then clearly γ is not tree-reduced in the sense of Definition 1.1.1 and in the equivalence class $[\gamma_{2e_1}]$, but one may compute $\|\gamma\|_{p\text{-var};[0,T]} = \|\gamma_{2e_1}\|_{p\text{-var};[0,T]} = 2$.

Recall that the characterisation of the tree-reduced representative in terms of minimal length was critical to the proof of the complete-metrisability of χ_d in Theorem 2.1.10. It is true, however, that the tree-reduced representative achieves the minimal p -variation in its equivalence class while not necessarily being the

unique minimiser. That is, if $\mathbb{X}$ is the tree-reduced representative of an equivalence class, then it satisfies [Boe+16]

$$\|\mathbb{X}\|_{p\text{-var};[0,T]} = \min_{\mathbb{Y} \in [\mathbb{X}]} \|\mathbb{Y}\|_{p\text{-var};[0,T]}.$$

However, perhaps the most challenging obstacle is the absence of a countable subset that is clearly dense with respect to the χ_d , which was necessary in proving separability of various topologies. Recall that whenever $\gamma \in C^{1\text{-var}}([0, T], V)$, we had that $\gamma^* \in C^{0,1\text{-var}}([0, T], V)$ which itself is a separable subspace of $C^{1\text{-var}}([0, T], V)$. The key was that paths which are parameterised with 1-Hölder control (i.e, Lipschitz at constant speed) are absolutely continuous. The equivalent statement is not true for $p > 1$, that is the p -Hölder control parameterisation of a path $\mathbb{X} \in W\Omega_p(V)$ is not necessarily in $G\Omega_p(V)$. However, one might still hope that in every $\sim_\tau$ equivalence class there is some representative in $G\Omega_p(V)$; perhaps under some complex reparameterisation or removal of tree-like pieces any weakly geometric rough path is tree-like equivalent to a geometric one? Equivalently, we would hope that $\mathcal{S}(G\Omega_p(V)) = \mathcal{S}(W\Omega_p(V)) = \mathcal{S}_p$. Unfortunately this is not the case, and it seems to be an accepted fact that the inclusion $\mathcal{S}(G\Omega_p(V)) \subset \mathcal{S}(W\Omega_p(V))$ is strict. However, to the best of our knowledge there is not a well-known proof of this result. As such, we provide a proof from the definition of tree-like equivalence and Wiener's characterisation, see e.g. [FV10, Theorem 8.22].

Theorem 2.4.2 (Wiener's Characterisation). *Let $\mathbb{X} \in W\Omega_p(V)$ for $p > 1$. The following statements are equivalent.*

- 1) $\mathbb{X} \in G\Omega_p(V)$;
- 2) $\lim_{\delta \rightarrow 0} \sup_{|\mathcal{P}| < \delta} \sum_{t_i \in \mathcal{P}} \|\mathbb{X}\|_{p\text{-var};[t_{i-1}, t_i]}^p = 0$;
- 3) $\lim_{\delta \rightarrow 0} \sup_{|\mathcal{P}| < \delta} \sum_{t_i \in \mathcal{P}} d_{CC}(\mathbb{X}_{t_{i-1}}, \mathbb{X}_{t_i})^p = 0$.

Using Wiener's characterisation, we are able to prove the following.

Theorem 2.4.3. *For $p > 1$, The space $\mathcal{S}(G\Omega_p(V))$ is a strict subset of $\mathcal{S}_p$.*

We break the proof into various smaller results, some of which are well known, but we include them for completeness.

Proposition 2.4.4. *Let α be a reparameterisation of $[0, T]$ and $\mathbb{X} \in W\Omega_p(V)$. If $\mathbb{X} \circ \alpha \in W\Omega_p(V) \setminus G\Omega_p(V)$, then $\mathbb{X} \in W\Omega_p(V) \setminus G\Omega_p(V)$.*

Proof. Note that α is uniformly continuous on $[0, T]$ since it is a non-decreasing surjection on a compact interval. By Wiener's characterisation

$$\lim_{\delta \rightarrow 0} \sup_{|\mathcal{P}| < \delta} \sum_{t_i \in \mathcal{P}} d_{\text{CC}}(\mathbb{X} \circ \alpha(t_{i-1}), \mathbb{X} \circ \alpha(t_i))^p = \kappa > 0.$$

Fix $\varepsilon > 0$. Since α is uniformly continuous, there is some $\delta > 0$ for which $|s - t| < \delta$ implies $|\alpha(s) - \alpha(t)| < \varepsilon$. Let $\mathcal{P}$ be a partition on $[0, T]$ with $|\mathcal{P}| < \delta$ for which

$$\sum_{t_i \in \mathcal{P}} d_{\text{CC}}(\mathbb{X} \circ \alpha(t_{i-1}), \mathbb{X} \circ \alpha(t_i))^p > \kappa.$$

The partition $\mathcal{Q} := \{\alpha(t_i) : t_i \in \mathcal{P}\}$ will then satisfy $|\mathcal{Q}| < \varepsilon$. Hence,

$$\sum_{s_i \in \mathcal{Q}} d_{\text{CC}}(\mathbb{X}_{t_{i-1}}, \mathbb{X}_{t_i})^p = \sum_{t_i \in \mathcal{P}} d_{\text{CC}}(\mathbb{X} \circ \alpha(t_{i-1}), \mathbb{X} \circ \alpha(t_i))^p > \kappa.$$

Taking limits as $\varepsilon \rightarrow 0$ and applying Wiener's characterisation gives that $\mathbb{X} \in WG\Omega_p(V) \setminus G\Omega_p(V)$. $\square$

The reverse result is not as straightforward. Indeed, a reparameterisation of $[0, T]$ may be constant on countably many subintervals so that a genuine surjective inverse α^{-1} does not exist. As such, we must first prove the existence of an inverse in a suitable sense.

Definition 2.4.5 (Pseudo-inverse). *Let $A \subseteq [0, T]$ be closed and $\alpha : A \rightarrow [0, T]$ be a continuous non-decreasing surjection, then we define α^{-1} by*

$$\alpha^{-1}(t) = \inf\{u \in A : \alpha(u) = t\}.$$

The following basic results are straightforward.

Lemma 2.4.6. *Let $A \subseteq [0, T]$ be closed and $\alpha : A \rightarrow [0, T]$ be a continuous non-decreasing surjection. Then*

- 1) $\alpha \circ \alpha^{-1}(t) = t$ for all $t \in [0, T]$;
- 2) Let $t_n \uparrow t$, then $\alpha^{-1}(t_n) \uparrow \alpha^{-1}(t)$.

Suppose further that

$$\alpha(s) = \alpha(t) \implies A \cap (s, t) = \emptyset,$$

then for all $\varepsilon > 0$ there exists a $\delta > 0$ for which $|s - t| < \delta$ implies

$$\lambda(A \cap [\alpha^{-1}(s), \alpha^{-1}(t)]) < \varepsilon,$$

where λ is the Lebesgue measure on $[0, T]$.

Proof. For the first item, let $u_n \downarrow \alpha^{-1}(t)$ with $\alpha(u_n) = t$. By continuity of α

$$\alpha \circ \alpha^{-1}(t) = \lim_{n \rightarrow \infty} \alpha(u_n) = t.$$

The second item is similar. For the third item, we proceed by contradiction. Suppose there is some $\varepsilon > 0$ such that for every $\delta_n = \frac{1}{n}$ there are $|s_n - t_n| < \delta_n$ for which

$$\lambda(A \cap [\alpha^{-1}(s_n), \alpha^{-1}(t_n)]) \geq \varepsilon.$$

Without loss of generality, by passing to subsequences if necessary, we may assume that $(s_n)_{n=0}^\infty$ and $(t_n)_{n=0}^\infty$ are monotone sequences with $s_n < t_n$ for all n with $s_n, t_n \rightarrow u \in [0, T]$. By definition of α^{-1} , $(\alpha^{-1}(s_n))_{n=0}^\infty$ and $(\alpha^{-1}(t_n))_{n=0}^\infty$ are monotone bounded sequences in a closed set A so that

$$\lim_{n \rightarrow \infty} \alpha^{-1}(s_n) = s \in A \quad \text{and} \quad \lim_{n \rightarrow \infty} \alpha^{-1}(t_n) = t \in A.$$

Note that $\alpha(s) = \alpha(t) = u$ by the first item. If $s = t$, then

$$\varepsilon \leq \lambda(A \cap [\alpha^{-1}(s_n), \alpha^{-1}(t_n)]) < |\alpha^{-1}(s_n) - \alpha^{-1}(t_n)| \rightarrow 0.$$

We now consider the following cases.

- (1): $s_n \uparrow u$ and $t_n \uparrow u$. Then $s = t$ by the second item of this lemma;
- (2): $s_n \downarrow u$ and $t_n \downarrow u$. This implies that for every n there is some k for which $t_k \leq s_n$ so that $\alpha^{-1}(t_k) \leq \alpha^{-1}(s_n)$, implying that $t \leq s$. Hence $s = t$;
- (3): $s_n \uparrow u$ and $t_n \downarrow t$. Hence $\alpha(s) = \alpha(t)$. Then $[\alpha^{-1}(s_n), \alpha^{-1}(t_n)]$ is a shrinking sequence of intervals for which

$$\varepsilon \leq \lim_{n \rightarrow \infty} \lambda(A \cap [\alpha^{-1}(s_n), \alpha^{-1}(t_n)]) = \lambda(A \cap (s, t)) = 0.$$

□

We are now ready to prove the following.

Proposition 2.4.7. *Suppose that α is a reparameterisation of $[0, T]$ and $\mathbb{X} \in W\Omega_p(V) \setminus G\Omega_p(V)$, then $\mathbb{X} \circ \alpha \in W\Omega_p(V) \setminus G\Omega_p(V)$.*

The idea is to use Wiener's characterisation for $\mathbb{X}$ and look at the inverse partition $\alpha^{-1}(\mathcal{P})$. By carefully designing $\mathcal{P}$ and inserting exactly the right additional points into $\alpha^{-1}(\mathcal{P})$, we can use Wiener's characterisation on $\mathbb{X} \circ \alpha$.

Proof. By [Gen15, Proposition 3.4.1] and the monotonicity of α , there exist countably many disjoint open intervals $I_i = (r_i^1, r_i^2) \subseteq [0, T]$ for which $\alpha_t = \alpha(\inf_{I_i})$ if $t \in I_i$ and

$$\alpha_s = \alpha_t \implies s, t \in \bar{I}_i \text{ for some } i. \quad (2.15)$$

Let $U = \bigcup_{i=1}^{\infty} I_i$ and $A = [0, T] \setminus U$. Note that the restriction $\alpha|_A$ is a continuous non-decreasing surjection with pseudo-inverse α^{-1} satisfying the conditions of Lemma 2.4.6. Now let $\delta > 0$ and $N \in \mathbb{N}$ for which

$$\sum_{i=N+1}^{\infty} \lambda(I_i) < \frac{\delta}{2}.$$

By Lemma 2.4.6 there is a $\varepsilon_1 > 0$ for which

$$\lambda(A \cap [\alpha^{-1}(s), \alpha^{-1}(t)]) < \frac{\delta}{2},$$

whenever $|s - t| < \varepsilon_1$. Order the finitely many intervals $I_1, \dots, I_N$ so that $r_i^2 < r_{i+1}^1$. Let

$$\varepsilon_2 = \min_{1 \leq i \leq N-1} |r_{i+1}^1 - r_i^2|.$$

By Eq. (2.15), it holds that $\varepsilon_2 > 0$. By Wiener's characterisation, we must have

$$\lim_{\delta \rightarrow 0} \sup_{|\mathcal{P}| < \delta} \sum_{t_i \in \mathcal{P}} d_{\text{CC}}(\mathbb{X}_{t_{i-1}}, \mathbb{X}_{t_i})^p = \kappa > 0.$$

Let $\mathcal{P}$ be a partition with $|\mathcal{P}| < \min\{\varepsilon_1, \varepsilon_2\}$ for which

$$\sum_{t_i \in \mathcal{P}} d_{\text{CC}}(\mathbb{X}_{t_{i-1}}, \mathbb{X}_{t_i})^p \geq \kappa.$$

For each (r_i^1, r_i^2) let $\mathcal{Q}_i$ be a partition on $[r_i^1, r_i^2]$ with $|\mathcal{Q}_i| < \delta$. Define the partition

$$\mathcal{Q} = \alpha^{-1}(\mathcal{P}) \cup \bigcup_{i=1}^N \mathcal{Q}_i.$$

Now for $t_j, t_{j+1} \in \mathcal{Q}$, if $t_j, t_{j+1} \in \mathcal{Q}_i$ then $|t_{j+1} - t_j| < \delta$. If $t_j \in [r_i^2, r_{i+1}^1]$ then $t_{j+1} \leq r_{i+1}^1$, implying that

$$\begin{aligned} |t_{j+1} - t_j| &= \lambda([t_j, t_{j+1}] \cap U) + \lambda([t_j, t_{j+1}] \cap A) \\ &< \lambda\left([t_j, t_{j+1}] \cap \bigcup_{i=N+1}^{\infty}\right) + \frac{\delta}{2} \\ &< \frac{\delta}{2} + \frac{\delta}{2}. \end{aligned}$$

Hence $|\mathcal{Q}| < \delta$. Now for $s_j \in \mathcal{P}$, there can be at most one interval I_i with $i \leq N$, contained in (s_{j-1}, s_j) since $|\mathcal{P}| < \varepsilon_2$. Letting $t_j := \alpha^{-1}(s_j) \in \alpha^{-1}(\mathcal{P})$, and assuming there exists such an interval, then

$$\begin{aligned} d_{\text{CC}}(\mathbb{X}_{s_{j-1}}, \mathbb{X}_{s_j})^p &= d_{\text{CC}}(\mathbb{X} \circ \alpha(t_{j-1}), \mathbb{X} \circ \alpha(t_j))^p \\ &\leq \left(d_{\text{CC}}(\mathbb{X} \circ \alpha(s_{j-1}), \mathbb{X} \circ \alpha(r_i^1))^p + d_{\text{CC}}(\mathbb{X} \circ \alpha(r_i^1), \mathbb{X} \circ \alpha(s_j))^p\right) \\ &= \left(d_{\text{CC}}(\mathbb{X} \circ \alpha(s_{j-1}), \mathbb{X} \circ \alpha(r_i^1))^p + d_{\text{CC}}(\mathbb{X} \circ \alpha(r_i^2), \mathbb{X} \circ \alpha(s_j))^p\right) \\ &\leq 2^{p-1} \left(d_{\text{CC}}(\mathbb{X} \circ \alpha(s_{j-1}), \mathbb{X} \circ \alpha(r_i^1))^p + d_{\text{CC}}(\mathbb{X} \circ \alpha(r_i^2), \mathbb{X} \circ \alpha(s_j))^p\right) \\ &= 2^{p-1} \left(d_{\text{CC}}(\mathbb{X} \circ \alpha(s_{j-1}), \mathbb{X} \circ \alpha(r_i^1))^p + d_{\text{CC}}(\mathbb{X} \circ \alpha(r_i^2), \mathbb{X} \circ \alpha(s_j))^p\right) \\ &\quad + \sum_{[u,v] \in \mathcal{Q}_i} d_{\text{CC}}(\mathbb{X} \circ \alpha(u), \mathbb{X} \circ \alpha(v))^p, \end{aligned}$$

where the final equality follows since all terms in the sum are identically zero. It follows that $\mathcal{Q}$ is a partition of $[0, T]$ with mesh-size less than δ for which

$$\begin{aligned} \kappa &\leq \sum_{s_j \in \mathcal{P}} d_{\text{CC}}(\mathbb{X}_{s_{j-1}}, \mathbb{X}_{s_j})^p = \sum_{t_j \in \alpha^{-1}(\mathcal{P})} d_{\text{CC}}(\mathbb{X} \circ \alpha(t_{j-1}), \mathbb{X} \circ \alpha(t_j))^p \\ &\leq 2^{p-1} \sum_{t_j \in \mathcal{Q}} d_{\text{CC}}(\mathbb{X} \circ \alpha(t_{j-1}), \mathbb{X} \circ \alpha(t_j))^p. \end{aligned}$$

And so, by Wiener's characterisation, we must have $\mathbb{X} \circ \alpha \in WG\Omega_p(V) \setminus G\Omega_p(V)$. $\square$

We can now conclude that being strictly weakly geometric is invariant under reparameterisation. Clearly if $\mathbb{X}$ is weakly geometric, it could be tree-like equivalent to a geometric path if the “weakly geometric” parts of the path are tree-like. We now need to show the reverse: one cannot add tree-like pieces to turn a strictly weakly geometric rough path into a geometric one. First recall the following construction of [Gen15; Boe+16]. For $\mathbb{X} \in WG\Omega_p(V)$, the corresponding tree-reduced path is a weakly geometric rough path $\tilde{\mathbb{X}}$ for which the signature path $S_{0,\cdot}(\tilde{\mathbb{X}})$ has loops erased in the following sense: there exists a countable collection of open intervals

$I_i \subset [0, T]$ for which

$$\mathcal{S}_{0,t}(\tilde{\mathbb{X}}) = \begin{cases} \mathcal{S}_{0,t}(\mathbb{X}), & t \notin \bigcup_{i=1}^{\infty} I_i \\ \mathcal{S}_{0,\inf I_i}(\mathbb{X}), & t \in I_i, \end{cases}$$

and satisfies the property that $\mathcal{S}_{0,s}(\mathbb{X}) = \mathcal{S}_{0,t}(\mathbb{X})$ if and only if $s, t \in \bar{I}_i$.

Proposition 2.4.8. *Suppose that $\mathbb{X} \in WG\Omega_p(V)$ and $\tilde{\mathbb{X}} \in WG\Omega_p(V) \setminus G\Omega_p(V)$, then $\mathbb{X} \notin G\Omega_p(V)$.*

Proof. Once again, we will use Wiener's characterisation. We know that

$$\lim_{\delta \rightarrow 0} \sup_{|\mathcal{P}| < \delta} \sum_{t_i \in \mathcal{P}} d_{CC}(\tilde{\mathbb{X}}_{t_{i-1}}, \tilde{\mathbb{X}}_{t_i})^p = \kappa > 0.$$

Fix $\delta > 0$ and let $\mathcal{P}$ be some partition with mesh-size less than δ for which the above sum is greater than or equal to κ . Consider now

$$\sum_{t_j \in \mathcal{P}} d_{CC}(\mathbb{X}_{t_{j-1}}, \mathbb{X}_{t_j})^p.$$

For each (t_i, t_{i+1}) we will add at most two points to the partition $\mathcal{P}$. Let $I_i, 1 \leq i < \infty$ be the intervals used to construct $\tilde{X}$ and let $U = \bigcup_{i=1}^{\infty} I_i$. Construct the partition $\mathcal{Q}$ by adding points to $\mathcal{P}$ according to the following.

- (1): If $t_{j-1}, t_j \notin U$, then $\mathbb{X}_{t_{j-1}} = \tilde{X}_{t_j}$ and $\mathbb{X}_{t_j} = \tilde{X}_{t_j}$, so there is no need to add points;
- (2): If $t_{j-1} \notin U$ and $t_j \in I_i$, then add the point $\inf I_i$. In this case

$$\begin{aligned} d_{CC}(\tilde{X}_{t_{j-1}}, \tilde{X}_{t_j})^p &\leq \left(d_{CC}(\mathbb{X}_{t_{j-1}}, \mathbb{X}_{\inf I_i}) + d_{CC}(\mathbb{X}_{\inf I_i}, \mathbb{X}_{t_j}) \right) \\ &\leq 2^{p-1} \left(d_{CC}(\mathbb{X}_{t_{j-1}}, \mathbb{X}_{\inf I_i})^p + d_{CC}(\mathbb{X}_{\inf I_i}, \mathbb{X}_{t_j})^p \right). \end{aligned}$$

- (3): If $t_{j-1} \in U$ and $t_j \notin U$, then add $\sup I_i$.
- (4): If $t_{j-1}, t_j \in I_i$, there is no need to add points since

$$d_{CC}(\tilde{X}_{t_{j-1}}, \tilde{X}_{t_j})^p \leq d_{CC}(\mathbb{X}_{t_{j-1}}, \mathbb{X}_{t_j})^p.$$

(5): If $t_{j-1} \in I_i$ and $t_j \in I_k$, then add $\sup I_i$ and $\inf I_k$. By Jensen's inequality,

$$d_{\text{CC}}(\tilde{X}_{t_{j-1}}, \tilde{X}_{t_j})^p \leq 3^{p-1} \left(d_{\text{CC}}(\mathbb{X}_{t_{j-1}}, \mathbb{X}_{\sup I_i})^p + d_{\text{CC}}(\mathbb{X}_{\inf I_i}, \mathbb{X}_{\sup I_k})^p + d_{\text{CC}}(\mathbb{X}_{\sup I_k}, \mathbb{X}_{t_j})^p \right).$$

Then $\mathcal{Q}$ is a partition of $[0, T]$ with mesh-size less than δ for which

$$3^{p-1} \sum_{t_j \in \mathcal{Q}} d_{\text{CC}}(\mathbb{X}_{t_{j-1}}, \mathbb{X}_{t_j})^p \geq \sum_{t_j \in \mathcal{P}} d_{\text{CC}}(\tilde{X}_{t_{j-1}}, \tilde{X}_{t_j})^p \geq \kappa.$$

Again, we conclude by applying Wiener's characterisation. $\square$

Proof of Theorem 2.4.3. Fix $\mathbb{X} \in WG\Omega_p(V) \setminus G\Omega_p(V)$ to be tree-reduced and let $\mathbb{Y} \in [\mathbb{X}]$. We first note that $\mathbb{X} = \mathbb{X}^* \circ \alpha$ for some reparameterisation α , so that $\mathbb{X}^*$ is strictly weakly geometric by Proposition 2.4.7. Since the tree-reduced representative is unique up to reparameterisation, it must be that $\tilde{\mathbb{Y}} = \mathbb{X}^* \circ \tau$ for some reparameterisation τ . As such, $\tilde{\mathbb{Y}}$ is strictly weakly geometric by Proposition 2.4.4. Finally, $\mathbb{Y}$ itself must be strictly weakly geometric by Proposition 2.4.8, from which we conclude that $[\mathbb{X}] \subset WG\Omega_p(V) \setminus G\Omega_p(V)$. Hence, $\mathcal{S}(G\Omega_p(V)) \subset \mathcal{S}_p$. $\square$

Remark 2.4.9. It is clear that there is always some $\mathbb{X} \in WG\Omega_p(V) \setminus G\Omega_p(V)$ that is tree-reduced. Indeed, let $\mathbb{Y}$ be a strictly weakly geometric rough path evolving in $\mathbb{R}^{d-1}$, then define $\mathbb{X}_t := (t, \mathbb{Y}_t)$. By Wiener's characterisation, $\mathbb{X}$ will be strictly weakly geometric and tree-reduced, since its signature path will be injective by construction.

The above result says that there is no obvious countably dense subset of unparameterised rough paths whenever $p > 1$. With these obstacles in mind, we have not been able to extend all the results of Section 2.1 to general $\mathcal{C}_p$ for $p > 1$. We do however present the results we have been able to extend, while drawing attention to the open questions.

2.4.1 Properties of topologies on unparameterised rough path space

Since the details of the proofs of many of the results are natural extensions of those found in Section 2.1, we will be relatively terse with our arguments. The topologies we defined in the preceding are ordered as follows.

Proposition 2.4.10. *We have the inclusions*

$$\chi_m \subseteq \chi_\tau \subseteq \chi_d. \quad (2.16)$$

Proof. Both inclusions follow analogously to [Proposition 2.1.5](#). $\square$

The above inclusions are in fact strict and may be shown directly as [Theorem 2.1.10](#) in the case $p = 1$. For brevity we will leave the strictness as a corollary of [Theorem 2.4.16](#).

Corollary 2.4.11. *All three topologies are Hausdorff.*

Proof. Since χ_m is a Hausdorff topology by construction, we immediately obtain that all three are Hausdorff. $\square$

From this point forwards we will assume that χ_m satisfies the following assumption.

Assumption 2.4.12. *There exists some $q \in (p, \lfloor p \rfloor + 1)$ for which*

$$\mathcal{S} : \left(WG\Omega_p(V), d_{q-var}^{(\lfloor p \rfloor)} \right) \rightarrow (\mathcal{C}_p, \chi_m)$$

is continuous.

Our study of the finer topological properties of χ_m relies heavily on the fact that under [Assumption 2.4.12](#), open sets must contain tree-reduced paths of arbitrarily large p -variation.

Lemma 2.4.13. *Suppose χ_m satisfies [Assumption 2.4.12](#). Then non-trivial open sets are unbounded in p -variation in the sense that*

$$\sup_{[\mathbb{X}] \in U} \|\mathbb{X}^*\|_{p-var} = \infty, \quad (2.17)$$

for any non-empty open U .

Proof. Let $U \in \chi_m$ be an open neighbourhood of some point $[\mathbb{X}] \in \mathcal{C}_p$. By [Assumption 2.4.12](#) and [Lemmas 1.0.36](#) and [1.0.37](#), $\pi^{-1}(U)$ contains the lift of some piecewise linear path parameterised at constant speed. With abuse of notation, we also call this path $\mathbb{X}^{\leq 1}$. We now follow the argument of [Lemma 2.1.14](#). Let $\{v_1, v_2\}$ be two orthonormal vectors not collinear to the last linear piece of $\mathbb{X}^{\leq 1}$. For every $\varepsilon > 0$, define the sequences of paths $(\mathbb{Y}_{n,\varepsilon}^{\leq 1})_{n=1}^\infty$ and $(\mathbb{Z}_{n,\varepsilon}^{\leq 1})_{n=1}^\infty$ by

$$\mathbb{Y}_{n,\varepsilon}^{\leq 1} = \mathbb{X}^{\leq 1} \star \gamma_{(n+\varepsilon)v_1} \star \gamma_{-(n+\varepsilon)v_1}, \quad (2.18)$$

$$\mathbb{Z}_{n,\varepsilon}^{\leq 1} = \mathbb{X}^{\leq 1} \star \gamma_{\varepsilon v_2} \star \gamma_{nv_1} \star \gamma_{-\varepsilon v_2} \star \gamma_{-nv_1}, \quad (2.19)$$

both parameterised at constant 1-variation. For every n, ε , it holds that $\mathbb{Y}_{n,\varepsilon}^{\leq \lfloor p \rfloor} \in [\mathbb{X}] \subseteq \pi^{-1}(U)$, whereas $\mathbb{Z}_{n,\varepsilon}^{\leq 1}$ is tree-reduced by construction. Additionally,

$$d_{1\text{-var}}(\mathbb{X}_{n,\varepsilon}^{\leq 1}, \mathbb{Z}_{n,\varepsilon}^{\leq 1}) \leq 4\varepsilon,$$

so that $\mathbb{Z}_{n,\varepsilon}^{\leq \lfloor p \rfloor} \in \pi^{-1}(U)$ for every n and some ε by [Lemma 1.0.37](#) together with continuity of the signature map $\mathcal{S} : WG\Omega_p(V) \rightarrow (\mathcal{C}_p, \chi_m)$. Finally, notice that

$$\begin{aligned} \|\mathbb{Z}_{n,\varepsilon}^{\leq \lfloor p \rfloor}\|_{p\text{-var}}^p &\geq \|\mathbb{Z}_{n,\varepsilon}^{\leq 1}\|_{p\text{-var}}^p \\ &\geq \|\mathbb{X}^{\leq 1}\|_{p\text{-var}}^p + 2\varepsilon^p + 2n^p. \end{aligned}$$

□

Proposition 2.4.14. *Suppose that χ_m satisfies [Assumption 2.4.12](#), then the balls $B_p(r) := \{[\gamma] : \|\gamma^*\|_{p\text{-var}} \leq r\}$ are compact in χ_m .*

The proof follows the method of [Proposition 2.2.2](#) or [[CL16](#), Proposition 5.7].

Proof. Since χ_m is a metrisable topology, it is enough to show that $B_p(r)$ is sequentially compact. Let $([\mathbb{X}_n])_{n=1}^\infty$ be a sequence in $B_p(r)$, then each tree-reduced representative $\mathbb{X}_n^*$ is $\frac{1}{p}$ -Hölder continuous with $\frac{1}{p}$ -Hölder norm bounded by r . Consequently, the sequence $(\mathbb{X}_n^*)_{n=1}^\infty \subset WG\Omega_p(V)$ is equicontinuous and bounded, with uniformly bounded p -variation. And so for any $p < q < \lfloor p \rfloor + 1$, a subsequence converges in $WG\Omega_q(V)$ to some path $\mathbb{X} \in WG\Omega_p(V)$, see [[FV10](#), Proposition 8.17]. By [[FV10](#), Lemma 5.12], $\|\mathbb{X}\|_{p\text{-var}} \leq r$ so that $[\mathbb{X}] \in B_p(r)$. Finally, by [Assumption 2.4.12](#), we have the convergence $[\mathbb{X}_n] \rightarrow [\mathbb{X}]$ in χ_m . □

The preceding results show that we can construct large sets that are nowhere dense in the topology χ_m . Taking a countable union of these closed sets allows us to conclude several topological properties.

Corollary 2.4.15. *If χ_m satisfies [Assumption 2.4.12](#), then the space $(\mathcal{C}_p, \chi_m)$ is*

- 1) *Separable;*
- 2) *σ -compact;*
- 3) *A Lusin space;*
- 4) *Not a Baire space;*
- 5) *Not locally compact.*

Proof. The first item follows from [Assumption 2.4.12](#) and [Lemma 1.0.37](#) and the separability result [Lemma 1.0.36](#). We may write

$$\mathcal{C}_p = \bigcup_{r=1}^{\infty} B_p(r). \quad (2.20)$$

Thus the Hausdorff space $(\mathcal{C}_p, \chi_m)$ is a countable union of compact sets [cf. [Proposition 2.4.14](#)] with empty interior [cf. [Lemma 2.4.13](#)] and so both σ -compact and a Lusin space, but not a Baire space. The final item follows from the Baire category theorem. $\square$

We now state one of our main results.

Theorem 2.4.16 (Complete Metrisability). *Let χ_m satisfy [Assumption 2.4.12](#), then*

- 1) χ_m is not completely-metrisable;
- 2) χ_τ is not metrisable;
- 3) The metric generating χ_d is not complete.

We record here that unlike in the case $p = 1$, the complete metrisability of χ_d for $p > 1$ remains unresolved.

Proof. The first item is an immediate consequence of [Corollary 2.4.15](#) and the Baire category theorem. For the second item, we follow the strategy of [Theorem 2.1.10](#) and show that χ_τ cannot both be first countable and regular, and so not metrisable. Assume that χ_τ has a countable neighbourhood basis $\{U_n\}_{n=1}^{\infty}$ of $[o] \in \mathcal{C}_p$ consisting of (with loss of generality) open sets. Fix two orthonormal vectors $\{v_1, v_2\}$ and consider the sequence of tree-like paths in $WG\Omega_1(V)$ given by

$$\mathbb{Y}_n^{\leq 1} := \gamma_{nv_1} \star \gamma_{-nv_1} = \gamma_{nv_1} \star \overleftarrow{\gamma_{nv_1}}.$$

Then each $\mathbb{Y}_n^{\leq [p]} \in [o] \in \mathcal{C}_p$, since the lift of a tree-like path is tree-like. As such,

$$r_n := \sup \{r > 0 : B_{d_{p\text{-var}}}(\mathbb{Y}_n^{\leq [p]}; r) \subseteq U_n\} > 0.$$

Define the strictly decreasing sequence of positive real numbers $\{\delta_n\}_{n=1}^{\infty}$ by

$$\delta_1 = \min\{r_1, 1\} \quad \text{and} \quad \delta_{n+1} = \frac{1}{2} \min\{\delta_n, r_{n+1}\},$$

for $n \geq 1$ so that $\pi\left(B_{d_{\text{p-var}}}\left(\mathbb{Y}_n^{\leq [p]}; \delta_n\right)\right) \subseteq U_n$ for each n . For $\varepsilon > 0$ let $\mathbb{Z}_{n,\varepsilon}^{\leq 1}$ be the path

$$\mathbb{Z}_{n,\varepsilon}^{\leq 1} = \gamma_{\varepsilon v_2} \star \gamma_{(n-\varepsilon)v_1} \star \gamma_{-\varepsilon v_2} \star \gamma_{-(n-\varepsilon)v_1},$$

so that $d_{1\text{-var}}(\mathbb{Z}_{n,\varepsilon}^{\leq 1}, \mathbb{Y}_n^{\leq 1}) \leq 4\varepsilon$. And so by [Lemma 1.0.37](#)

$$\lim_{\varepsilon \rightarrow 0} d_{\text{p-var}}(\mathbb{Z}_{n,\varepsilon}^{\leq [p]}, \mathbb{Y}_n^{\leq [p]}) = 0.$$

It then follows that $\mathbb{Z}_{n,\varepsilon}^{\leq [p]} \in B_{d_{\text{p-var}}}(\mathbb{Y}_n^{\leq [p]}; \delta_n)$ for suitably small ε_n , resulting in $[\mathbb{Z}_{n,\varepsilon_n}^{\leq [p]}] \in U_n$. Clearly

$$\left\| \mathbb{Z}_{n,\varepsilon_n}^{\leq [p]} \right\|_{\text{p-var}} \uparrow \infty,$$

meaning that

$$A := \bigcup_{n=1}^{\infty} A_n, \text{ with } A_n := \bigcup_{j=1}^n \left\{ [\mathbb{Z}_{j,\varepsilon_j}^{\leq [p]}] \right\},$$

is closed in χ_τ . To see this, note that if $(\mathbb{W}_n)_{n=1}^{\infty} \subset \pi^{-1}(A)$ converges to some $\mathbb{W} \in W\Omega_p(V)$, then $\sup_n \|\mathbb{W}_n\|_{\text{p-var}} < \infty$. Using the fact that each $\mathbb{Z}_{n,\varepsilon_n}^{\leq [p]}$ is tree-reduced with exploding p -variation, there exists some N for which $(\mathbb{W}_n)_{n=1}^{\infty}$ is contained in A_N . A_N is a finite union of equivalence classes each of which is closed by [Corollary 2.4.11](#) and hence $\mathbb{W} \in \pi^{-1}(A)$. We have thus constructed a closed set A , not containing $[o]$, which has non-empty intersection with every neighbourhood of $[o]$; χ_τ cannot then be both first countable and regular, and so is not metrisable.

For the final item let $\{v_1, v_2\}$ be a pair of unit vectors in V which are not collinear. For each $\frac{1}{2} > \varepsilon > 0$ define the following tree-reduced path

$$\mathbb{X}_\varepsilon^{\leq 1} = \gamma_{\varepsilon v_2} \star \gamma_{(1/2-\varepsilon)v_1} \star \gamma_{-\varepsilon v_1} \star \gamma_{-(1/2-\varepsilon)v_1}. \quad (2.21)$$

Let $\mathbb{X}_\varepsilon^{\leq [p]}$ be the canonical lift into $C^{1\text{-var}}([0, T], G^{([p])}) \subset W\Omega_p(V)$ which is then parameterised at $\frac{1}{p}$ -Hölder control. Let $(\varepsilon_n)_{n=1}^{\infty}$ be a sequence satisfying $0 < \varepsilon_n < \frac{1}{2}$ and converges to 0. The sequence $(\mathbb{X}_{\varepsilon_n}^{\leq 1})_{n=1}^{\infty}$ is clearly bounded in 1-variation and equicontinuous by the Hölder control reparameterisation. Hence the sequence $(\mathbb{X}_{\varepsilon_n}^{\leq 1})_{n=1}^{\infty}$ converges uniformly (along a subsequence) to some path $\mathbb{X}^{\leq 1}$ of finite 1-variation, and so converges in q -variation for $q \in (1, 2)$ by a standard interpolation argument. By the form of [Eq. \(2.21\)](#) it is clear that $\mathbb{X}^{\leq 1}$ must be a reparameterisation of the path $\gamma_{v_1} \star \gamma_{-v_1}$. By [Lemma 1.0.37](#) we conclude that $\mathbb{X}_{\varepsilon_n}^{\leq [p]} \rightarrow \mathbb{X}^{\leq [p]}$ in $W\Omega_p(V)$, and so forms a Cauchy sequence with respect to the metric d . By continuity of the quotient map, we have the convergence of equivalence classes $[\mathbb{X}_{\varepsilon_n}^{\leq [p]}] \rightarrow [\mathbb{X}^{\leq [p]}] = [o]$ in χ_τ . By the uniqueness of limits in Hausdorff spaces, the

only possible limit of $[\mathbb{X}_{\varepsilon_n}^{\leq [p]}]$ in χ_d is then $[o]$. However it is clear that

$$\lim_{n \rightarrow \infty} d([\mathbb{X}_{\varepsilon_n}^{\leq [p]}], [o]) \geq 2^{1/p-1},$$

so that $[\mathbb{X}_{\varepsilon_n}^{\leq [p]}]$ does not converge to $[o]$ with respect to the d .

□

As a result of the non-metrisability of the quotient topology, the inclusions from [Proposition 2.4.10](#) are strict.

Corollary 2.4.17. *The inclusions in [Eq. \(2.16\)](#) are strict. That is*

$$\chi_m \subset \chi_\tau \subset \chi_d.$$

The non-completeness of the metric topology has the following consequence.

Corollary 2.4.18. *The space $(\mathcal{C}_p, \chi_d)$ is not locally compact.*

Proof. Using the same techniques as in [Theorem 2.4.16](#), we can show that $B_p(r)$ is not complete for any r . As a result $[o]$ cannot have a compact neighbourhood in this topology, as this would imply $B_p(r)$ is compact for some suitably small r . □

2.4.2 The totality of unparameterised rough paths

We conclude this chapter with a discussion on the totality of all unparameterised rough paths $\mathcal{C}$, that is the quotient of the space

$$WG\Omega(V) := \bigcup_{p \geq 1} WG\Omega_p(V)$$

by tree-like equivalence. Here for $q > p$, two paths $\mathbb{X}^{\leq [p]} \in WG\Omega_p(V)$ and $\mathbb{Y}^{\leq [q]} \in WG\Omega_q(V)$ are called tree-like equivalent if $\mathbb{X}^{\leq [q]}$ and $\mathbb{Y}^{\leq [q]}$ are tree-like equivalent. It is unclear how to topologise $WG\Omega(V)$ in a consistent fashion, and so defining a suitable quotient topology on $\mathcal{C}$ is not straightforward. However, it is still possible to define topologies on $\mathcal{C}$ induced by metrics on subspaces of $G^{(\star)}$.

Remark 2.4.19. *We can canonically identify $\mathcal{C}$ with the space $\mathcal{S}$ (just as we have often identified $\mathcal{C}_p$ with $\mathcal{S}_p$), so that we can view $\mathcal{C}_p$ as a subset of $\mathcal{C}$ through the inclusion $\mathcal{S}_p \subset \mathcal{S}$.*

In a similar spirit to [Section 2.4.1](#) we have the following theorem.

Theorem 2.4.20. *Let χ be any metrisable topology on $\mathcal{C}$ for which $\mathcal{S} : WG\Omega_p(V) \rightarrow (\mathcal{C}, \chi)$ is continuous for all p , then the space $(\mathcal{C}, \chi)$ is*

- 1) *Separable*;
- 2) *σ -compact*;
- 3) *Not locally compact*;
- 4) *Not polish*.

Proof. Recall that the (lift of the) same countable subset of piecewise linear paths is dense in each $G\Omega_p(V)$ for every $p \geq 1$, and that $WG\Omega_p(V) \subset G\Omega_q(V)$ for all $q > p$. By continuity for each $p \geq 1$, the image of these paths under $\mathcal{S}$ will then be dense in $\mathcal{C}$.

For the remaining points, we first claim that every open set in $\mathcal{C}$ is unbounded in p -variation for some $p \geq 1$. First note that for every $p \geq 1$, the subspace topology on $\mathcal{C}_p$ induced by the topology χ on $\mathcal{S}$ results in a topology on $\mathcal{C}_p$ that satisfies [Assumption 2.4.12](#). Label this topology as χ_m^p . Let U be a non-trivial open set in $\mathcal{C}$ containing the signature of a weakly-geometric rough path $\mathbb{X} \in \mathcal{C}_p$. Then $U \cap \mathcal{C}_p \subseteq \mathcal{C}_p$ will be a non-trivial open set in $(\mathcal{C}_p, \chi_m^p)$, and so unbounded in p -variation in the sense of [Eq. \(2.17\)](#) by [Lemma 2.4.13](#). Let $B_p(r)$ be the set of equivalence classes whose tree-reduced representative may be viewed as an element of $WG\Omega_p(V)$ with p -variation bounded by r . By the same argument as in [Proposition 2.4.14](#), each $B_p(r)$ will be compact in $(\mathcal{C}, \chi)$. We may then write

$$\mathcal{C} = \bigcup_{p=1}^{\infty} \bigcup_{r=1}^{\infty} B_p(r).$$

This is now a countable union of compact subsets of a Hausdorff space with empty interior, from which we deduce the remaining points by the Baire category theorem. $\square$

The set of group-like elements $G^{(*)}$ is a closed and completely metrisable subset of $T((V))$ equipped with the product topology (see [Proposition 2.1.18](#)) which itself is completely metrisable. Since $\mathcal{S}$ is contained in $G^{(*)}$ and is itself not completely metrisable, we obtain the following well known result.

Corollary 2.4.21. *The set of all signatures $\mathcal{S}$ is a strict subset of $G^{(*)}$. If $G^{(*)}$ is equipped with the product topology inherited from $T((V))$, then $\mathcal{S}$ is a meagre subset of $G^{(*)}$.*

CHAPTER 3

RANDOMISED PATH DEVELOPMENTS

Our first function on unparameterised path space is the path development. Path developments have long formed a vital tool in understanding the theoretical properties of the signature, itself an example of a path development. In particular, the hyperbolic development was heavily utilised in the proof of uniqueness of the signature in [HL10], the study of signature asymptotics [BG19a] and signature inversion [LX17]. Meanwhile, the unitary development allows for a theory of Fourier transforms of measures on unparameterised path space [CL16] as well as constructive determination of the signature [Li+25]. Recently, there has been a reversal of this trend where path developments have been used as an alternative to the signature transform in certain applications. For example, they have been applied as a layer in an LSTM [LLN24], a discriminator between probability measures on path space [LLN23; TNL24] and as an approximation to the signature kernel [MLS23]. A feature of the latter two applications is that the matrices driving the path development are randomised with the dimension of the target Lie group ideally taken as high as possible. This chapter studies the scaling limits of such randomised path developments when the target Lie group is either the general linear group or the unitary group. This chapter is based on the pair of works [CT24a; Cre+25].

The signature, which we have seen to be an injective map from the space of unparameterised paths into $T((V))$, is an example of a path development. Recall that the signature of $\gamma \in C^{1\text{-var}}([0, T], V)$ is given by $\mathcal{S}(\gamma) = Z_T$, where Z solves the CDE

$$dZ_t = Z_t \otimes d\gamma_t \quad \text{on } [0, T], \quad (3.1)$$

in the space $T((V))$ with initial condition $Z_0 = 1$, the identity in $T((V))$. The structure of Eq. (3.1) resembles the classical operation of exponentiation, and the

signature transform can be loosely viewed as its analogue with respect to the non-commutative product $\otimes$. As we have already seen, the terms in the signature provide a natural way to express functional relationships on path space; a canonical example (see [Proposition 2.2.7](#)) is the map $(y_0, \gamma) \mapsto y$, where y solves the CDE

$$dy_t = f(y_t) d\gamma_t, \quad y(0) = y_0.$$

The solution to [Eq. \(3.1\)](#) can therefore encode a wide class of such maps. In particular, the signature reduces the analysis of paths $\gamma : [0, T] \rightarrow V$ to the study of a sequence of tensors in $\prod_{m=0}^{\infty} V^{\otimes m}$ associated with γ . Since $\dim V^{\otimes m}$ grows exponentially in m , dimensionality considerations remain important. A common approach is to truncate the signature to order L , working instead in the finite-dimensional space $T^{(L)} = \prod_{m=0}^L V^{\otimes m}$. In other contexts, signature kernel methods have become popular [[KO19](#); [Sal+21](#); [Iss+23](#); [CLX24](#)], enabling efficient computation of inner products

$$K(\gamma, \sigma) := \langle \mathcal{S}(\gamma), \mathcal{S}(\sigma) \rangle,$$

for various choices of inner product on $T((V))$ without explicitly computing $\mathcal{S}(\gamma)$.

One promising recent alternative is to use the development of γ into a matrix Lie group [[LLN24](#); [LLN23](#)]. In this formulation the central object is not the signature itself but related representations obtained from the evolution

$$dZ_t = M(d\gamma_t)(Z_t), \quad Z_0 = I_N, \tag{3.2}$$

where Z takes values in a matrix Lie group $G_N \subset \mathcal{M}_N(F)$ and $M \in \text{Hom}(V, \text{Lie}(G_N))$. By identifying the Lie algebra $\mathfrak{g}_N$ with $\text{Lie}(G_N)$ we may equivalently regard $M \in \text{Hom}(V, \mathfrak{g}_N)$ and write

$$dZ_t = Z_t \cdot M(d\gamma_t), \quad Z_0 = I_N, \tag{3.3}$$

where $\cdot$ denotes matrix multiplication. Important examples include:

- 1) $F = \mathbb{R}$ or $\mathbb{C}$, with $G_N = \text{GL}(N; F)$ and $\mathfrak{g}_N = \mathcal{M}_N(F)$; the resulting representations are denoted $\mathcal{G}(\gamma; M) := Z_T$;
- 2) $F = \mathbb{C}$, with $G_N = U(N; \mathbb{C})$ and Lie algebra the anti-Hermitian matrices

$$\mathfrak{u}_N := \{M \in \mathcal{M}_N(\mathbb{C}) : M^* = -M\},$$

yielding unitary features $\mathcal{U}(\gamma; M) := Z_T$.

The $N = 1$ case recovers the characteristic-function: if $M = i \cdot f$ with $f \in V^*$ then

$$Z_T = \exp \left(i f(\gamma_{0,T}) \right), \quad \gamma_{0,T} := \gamma_T - \gamma_0,$$

and for $\gamma_{0,T}$ distributed according to μ we have $\mathbb{E}_\mu[Z_T] = \Phi_\mu(f)$, the classical characteristic function at f .

Lou, Li and Ni [LLN23] leverage these developments for generative modelling by defining

$$\Phi_\mu(M) := \mathbb{E}_{\gamma \sim \mu} [\mathcal{U}(\gamma; M)], \quad d_M(\mu, \nu) := \|\Phi_\mu(M) - \Phi_\nu(M)\|_{\text{HS}},$$

and then randomising the choice of development via a probability measure ξ on the space of maps M to obtain

$$d_\xi(\mu, \nu) := \mathbb{E}_{M \sim \xi} [d_M(\mu, \nu)^2]^{1/2}. \quad (3.4)$$

When $N = 1$, this reduces to the characteristic-function distance of [FSS20; Li+20a]. In the experiments of [LLN23] the measure ξ is taken to be discrete, i.e. $\xi = \frac{1}{k} \sum_{i=1}^k \delta_{M_i}$, and the finite collection $\{M_i\}_{i=1}^k$ is treated as trainable parameters of the discriminator which are learnt adversarially. The authors report state-of-the-art performance across various time-series generation tasks.

Despite these promising results, questions remain which affect the wider use of this approach. The critical issue of dimensionality outlined above is transferred – rather than resolved – from the selection of a truncation parameter in the signature, to the choice of a Lie group of sufficiently high dimension for the method to be effective. The overhead associated with computing d_ξ , as well as the difficulty of optimising the large parameter set forming the discriminator remains and may be prohibitive in some cases. It is desirable therefore to have some direct means of computing Eq. (3.4), especially in the case where N may be large.

One way to tackle this problem is to observe [LLN23, Section B.3] that d_ξ is a maximum mean discrepancy distance (MMD). Indeed

$$\begin{aligned} d_\xi^2(\mu, \nu) &= \mathbb{E}_{M \sim \xi} [\langle \Phi_\mu(M), \Phi_\mu(M) \rangle] + \mathbb{E}_{M \sim \xi} [\langle \Phi_\nu(M), \Phi_\nu(M) \rangle] \\ &\quad - 2 \mathbb{E}_{M \sim \xi} [\langle \Phi_\mu(M), \Phi_\nu(M) \rangle], \end{aligned}$$

so that d_ξ is an MMD with associated kernel

$$\kappa(\gamma, \sigma) := \mathbb{E}_{M \sim \xi} [\langle \mathcal{U}(\gamma; M), \mathcal{U}(\sigma; M) \rangle_{\text{HS}}].$$

As our interest is primarily in the asymptotic regime $N \rightarrow \infty$, this leads us to ask whether we can characterise the large N behaviour of the law of the random variables

$$\langle \mathcal{U}(\gamma; \cdot), \mathcal{U}(\sigma; \cdot) \rangle_{\text{HS}} \text{ under } \xi_N \quad (3.5)$$

for a suitable class of (Borel) probability measures ξ_N on $\mathfrak{u}_N$. While this may appear a challenging technical problem, the theory of non-commutative probability [Voi85; Voi87; Voi86] provides both a general framework for the study of these limits, as well as explicit ways to characterise and calculate them. We recall that a non-commutative C^* -probability space is a pair $(\mathcal{A}, \phi)$ consisting of a C^* -algebra $\mathcal{A}$ having a unit e , and a state (in the sense of [AGZ10, Definition 5.2.0]) $\phi : \mathcal{A} \rightarrow \mathbb{C}$ with $\phi(e) = 1$. Two well-known examples are

- 1) $\mathcal{A}_1 = \mathcal{M}_N(\mathbb{C})$, and $\phi(M) = \frac{1}{N} \text{Tr}(M)$;
- 2) $\mathcal{A}_2 = L^\infty(\Omega; \mathcal{M}_N(\mathbb{C}))$, the space of bounded $\mathcal{M}_N(\mathbb{C})$ -valued random variables, with the linear-functional $\phi(M) = \frac{1}{N} \mathbb{E}[\text{Tr}(M)]$.

From this point forward, we will write Tr for the usual trace of a matrix and $\text{tr} := \frac{1}{N} \text{Tr}$ for the normalised trace. One attractive feature of this theory is the way in which it can be used to identify limits, both almost surely and in expectation, of sequences such as Eq. (3.5). These limits are typically determined by the dependence structure between the matrix entries but are otherwise universal. They do not require ξ_N to belong to a simple or explicit class of distributions – in particular they need not be Gaussian – and instead require only mild moment assumptions. The archetype of this result shows that sequences (of collections) of random Wigner matrices $\mathcal{L}_N := \left\{ \frac{1}{\sqrt{N}} M_i^N \right\}_{i=1}^d \in \mathcal{M}_N(\mathbb{C})$ under very general assumptions (on the distribution of the entries) are asymptotically free and converge in law (in the sense of non-commutative probability) to the semicircle law. These statements hold both in the almost sure sense, viewing $\mathcal{L}_N$ as a random set in $(\mathcal{M}_N(\mathbb{C}), \text{tr}(\cdot))$, and in expectation, by working instead with $\phi(\cdot) = \mathbb{E}[\text{tr}(\cdot)]$.

By using techniques from this theory we are able to define the limit

$$K_{\text{SD}}(\gamma, \sigma) := \lim_{N \rightarrow \infty} \frac{1}{N} \mathbb{E}_{M \sim \xi_N} [\langle \mathcal{U}(\gamma; M), \mathcal{U}(\sigma; M) \rangle_{\text{HS}}],$$

and moreover understand $K_{\text{SD}}(\gamma, \sigma)$ in terms of the signature of the path $y = \gamma * \overleftarrow{\sigma}$

$$K_{\text{SD}}(\gamma, \sigma) = \sum_{k=0}^{\infty} i^k \sum_{|u|=k} \tau(X_u) \mathcal{S}_{0,T}^u(y), \quad (3.6)$$

where $\{X_1, \dots, X_d\}$ are free semicircular random variables with $\tau(X_u)$ the non-commutative moment of the monomial $X_{u_1} \dots X_{u_k}$. To make practical use of this result, we need some means of computing Eq. (3.6) that avoids the direct computation of signatures, by analogy to the signature kernel PDE [Sal+21]. The structure of free semicircular variables offers a way to do this. One way to characterise their properties is through the *Schwinger-Dyson equations* (hence K_{SD}). These equations, also known as master equations or loop equations, originated from quantum field theory in the 1940s and form an infinite hierarchy of equations governing the correlation functions of quantum fields. They also have an important realisation in the context of non-commutative probability. In particular, they arise in the study of the asymptotic behaviour of high-dimensional Wigner matrices. Let $A_1^N, \dots, A_d^N$ be a collection of $N \times N$ Wigner matrices with independent, centred entries with variance $\frac{1}{N}$. Then Voiculescu [Voi91] proved that

$$\lim_{N \rightarrow \infty} \mathbb{E}[\text{tr}(A_{i_1}^N \dots A_{i_n}^N)] = \tau(X_{i_1} \dots X_{i_n}),$$

where $\tau(\cdot)$ is the non-commutative law of d -free semicircular random variables $X_1, \dots, X_d$. That is, a collection of non-commutative random variables each of whose marginal law has density

$$\sigma(x) = \frac{1}{2\pi} \sqrt{4 - x^2} \mathbb{1}_{|x| \leq 2},$$

and whose mixed moments are determined combinatorially by

$$\tau(X_w) = \sum_{\pi \in \text{NC}_2(n)} \prod_{\{i,j\} \in \pi} \mathbb{1}_{w_i = w_j}, \quad (3.7)$$

for any word $w = w_1 \dots w_n$ and where $\text{NC}_2(n)$ denotes the set of non-crossing pair partitions of the set $\{1, \dots, n\}$. One may think of Eq. (3.7) as a non-commutative analogue of Wick's formula for Gaussian random variables. Just as the Gaussian distribution is characterized by the integration-by-parts (or Stein's) identity, free semicircular variables satisfy an analogous identity known as the Schwinger–Dyson equation. In the setting of free semicircular random variables, the equations take the form

$$\tau(X_{wj}) = \sum_{w=ujv} \tau(X_u) \tau(X_v) \quad \text{and} \quad \tau(X_{uv}) = \tau(X_{vu}). \quad (3.8)$$

Since the seminal work of Voiculescu these equations have been studied extensively in various contexts within non-commutative probability, such as free diffusions [GS09], free transport [GS14b], free energy and matrix models [GM07; GM06] as well as in connections to the enumeration of planar maps, see [Gui19] for an

overview and references. We use this structure to prove that $K_{\text{SD}}(\gamma, \sigma) = K_{\text{SD}}^y(0, T)$, where $y = \gamma * \overleftarrow{\sigma}$ and $K_{\text{SD}}^y : \{0 \leq s \leq t \leq T\} \rightarrow \mathbb{R}$ solves the quadratic functional equation

$$\begin{aligned} K_{\text{SD}}^y(s, t) &= 1 - \int_s^t \int_s^r K_{\text{SD}}^y(s, u) K_{\text{SD}}^y(u, r) \langle dy_u, dy_r \rangle, \\ K_{\text{SD}}^y(s, s) &= 1 \text{ for all } s \in [0, T]. \end{aligned} \quad (3.9)$$

We show that this equation (which appears not to have been previously studied) admits a unique solution. We then propose and implement a numerical scheme to solve Eq. (3.9) and establish convergence of the scheme to the true solution together with explicit convergence rates. An alternative numerical approach is to simulate multiple trajectories of the path-development in high dimensions and average the results. In the case where γ is piecewise linear, the solution admits an explicit representation via matrix exponentials and products. Both operations have computational cost approximately $\mathcal{O}(N^3)$ in the matrix dimension and thus become rapidly infeasible for large N . Since all matrices involved are anti-Hermitian or unitary, one can in principle leverage quantum algorithms to simulate these trajectories more efficiently. A caveat is that the driving matrices must be sufficiently sparse and possess “efficient” representations; this is the subject of the recent preprint [Cre+25]. Quantum computing lies beyond the scope of this thesis, but we present the novel random matrix theory involved in Section 3.4.

In a different direction, we also derive the limiting equation when the driving matrices are no longer Wigner matrices but are distributed according to a perturbation of the Gaussian Unitary Ensemble on Hermitian matrices. Such matrix models which arise naturally in the context of zero-dimensional quantum field theory [EKR15], allow for interactions both between entries and between matrices. Under suitable conditions these perturbations also asymptotically satisfy Schwinger–Dyson equations [GM07], and the corresponding functional equation for the path-development is then a perturbation of Eq. (3.9) involving path-dependent derivatives. A brief introduction to matrix models and Schwinger–Dyson equations is given in Section 3.0.4.

As a further application of the use of these methods we consider the evolution Eq. (3.3) in $G_N = \text{GL}(N; \mathbb{C})$ the general linear group. The corresponding limit under these assumptions is the classical signature kernel:

$$K_{\text{sig}}(\gamma, \sigma) = \lim_{N \rightarrow \infty} \frac{1}{N} \mathbb{E}_{M \sim \xi_N} [\langle \mathcal{G}(\gamma; M), \mathcal{G}(\sigma; M) \rangle_{\text{HS}}].$$

For the specific case where ξ_N is the Gaussian measure on $\text{Hom}(V, \mathcal{M}_N(\mathbb{R}))$ with variance $\frac{1}{N}$, this result was obtained in the recent paper of Mua Cirone et al. [MLS23, Theorem 4.4]. By adopting the techniques outlined above we are able to give a concise proof which also covers the non-Gaussian case, allows for complex valued matrices and dispenses with some unnecessary independence conditions. We note that [MLS23] also derives limits for non-linear randomised CDEs, though the proof techniques differ substantially from the linear case. The original motivation in [MLS23] is to analyse the scaling limits of controlled ResNets arising as the Euler discretisation of CDEs; in particular the authors focus on the kernel function of the limiting random processes. We hope our results help broaden this emerging connection between families of kernels on unparameterised path space and scaling limits of controlled ResNets, where infinite width and depth limits are known to commute in certain cases [MLS23].

This chapter also sits within a broader body of literature surrounding the scaling limits of random neural networks and the application of random matrix theory within machine learning. Random matrix theory and free probability theory are swiftly arising as popular tools to analyse the properties of relevant matrices. For example in [PSG17] where the authors employ techniques from free probability theory to derive the singular value distribution of the input-output Jacobian of a network. Furthermore, the eigenvalue distribution of certain non-linear random matrix models was studied in [BP21] and the limiting spectrum of the neural tangent kernel (NTK) has also received much attention [AP20; WZ23]. We refer the reader to [Ge+21] for a survey and references on recent results in this area.

3.0.1 Outline

In Section 3.1 we leverage techniques from non-commutative probability theory to provide a universal version of [MLS23, Theorem 4.4]. We adapt the methods from the proof of [AGZ10, Theorem 5.4.2], which deals with Wigner random matrices, to the setting of $\mathcal{M}_N(\mathbb{C})$ -valued random matrices. The core of the approach in [AGZ10] is the association of undirected graphs with words from an alphabet of N letters. These words then correspond to the individual terms that form the trace of a product of random matrices. When dealing with generic $\mathcal{M}_N(\mathbb{C})$ -valued matrices, we additionally associate a directed graph with each word. When combined with the approach of Mua Cirone et al., this insight yields a more concise proof which eliminates certain Gaussian and independence assumptions as well as allowing for complex valued matrices. The relevant definitions on words and their associated graphs are found in Section 3.0.3.

Section 3.0.4 provides a background on matrix models, non-commutative laws and Schwinger-Dyson equations. We present one of our main results [Theorem 3.2.13](#) in [Section 3.2](#), where we prove that a universal limit exists for random $U(N; \mathbb{C})$ -valued path developments driven by Wigner matrices. The existence of such a limit is a relatively straightforward consequence of [\[AGZ10, Theorem 5.4.2\]](#) and the techniques from [Section 3.1](#). Our main contribution is showing that this limit solves an explicit quadratic functional equation, which has a unique solution, see [Lemma 3.2.15](#). The proof of uniqueness is non-trivial due to the quadratic nature of the functional equation. Our proof relies on Dyck words and (to the best of our knowledge) a new notion of the “generation” of a Dyck word. The generation of a Dyck word is not its length, but the number of steps taken to generate the word using a certain algorithm; see [Section 3.2.1](#) for details.

In [Section 3.3](#), we outline an explicit numerical scheme to solve the functional equation. We prove convergence along with rates in [Theorem 3.3.2](#) through an object known as the iterated sums signature and corresponding results from [\[KO19\]](#). We end by comparing this numerical scheme with the Monte Carlo approach of averaging multiple random $U(N; \mathbb{C})$ path developments with large N . In particular, we analyse how the computational time for each method scales in the length of the input path, the dimension of the input path and the dimension N in the case of the randomised approach.

Although we do not address the use of quantum algorithms to overcome the challenges of simulating trajectories for large N , we present in [Section 3.4](#) the random matrix theory underpinning the algorithm of [\[Cre+25\]](#).

We conclude this chapter with a straightforward extension of the results in [Section 3.2](#), where we show in [Section 3.5](#) that the Schwinger-Dyson equations for perturbations of the Gaussian Unitary Ensemble also allow us to derive a limiting functional equation in this setting.

3.0.2 The signature kernel

The signature kernel, originally introduced by Király and Oberhauser in [\[KO19\]](#), is defined in the following way.

Definition 3.0.1. *The signature kernel between two paths $\gamma, \sigma \in \mathcal{X}$ is defined by*

$$K_{sig}^{\gamma, \sigma}(s, t) := \langle \mathcal{S}_{0,s}(\gamma), \mathcal{S}_{0,t}(\sigma) \rangle := \sum_{m=0}^{\infty} \langle \mathcal{S}_{0,s}(\gamma)^m, \mathcal{S}_{0,t}(\sigma)^m \rangle_{HS}, \quad (3.10)$$

where for each m , $\langle \cdot, \cdot \rangle_{HS}$ denotes the Hilbert-Schmidt inner product on $V^{\otimes m}$.

Efficient computation of the signature kernel has seen much interest lately. The authors of [KO19] provided a kernel trick for a truncated version of $K_{\text{sig}}(s, t)$. In [Sal+21] it is shown that K_{sig} satisfies a Goursat PDE in the case that γ and σ are continuously differentiable. Finally, in [MLS23] a connection was established between the signature kernel and N -dimensional randomised CDEs in the limit as $N \rightarrow \infty$.

3.0.3 Letters, words and graphs

For ease of cross referencing we will base much of our notation on [AGZ10].

Definition 3.0.2 (Words). *For $d \in \mathbb{N}$ let $\mathcal{W}_d$ denote the set of words in the d -letters $\{1, \dots, d\}$ and let $\emptyset \in \mathcal{W}_d$ denote the empty-word. Define $\mathcal{A}_d$ to be the algebra over $\mathcal{W}_d$ with multiplication defined by*

$$\mathbf{w}\mathbf{u} := w_1 \dots w_n u_1 \dots u_m,$$

for $\mathbf{w}, \mathbf{u} \in \mathcal{W}_d$. Define also the map $|\cdot| : \mathcal{W}_d \rightarrow \mathbb{N}$ to be the length of a word. We also write $\mathcal{W}_d^k$ for set words in d letters of length k . We also define the set $\mathcal{W}_d^0 = \{\emptyset\}$, which contains the empty word.

We define the following operations on words.

Definition 3.0.3 (Operations on words). *For a word $\mathbf{w} \in \mathcal{W}_d$ we define*

- 1) *The support of $\mathbf{w}$, $\text{supp}(\mathbf{w}) \subseteq \{1, \dots, d\}$, as the set of distinct letters in $\mathbf{w}$;*
- 2) *Its weight, $\text{wt}(\mathbf{w})$, to be the number of distinct letters in $\mathbf{w}$;*
- 3) *Its transpose $\mathbf{w}^* := w_k \dots w_1$;*
- 4) *An undirected graph $G_{\mathbf{w}}^u = (V_{\mathbf{w}}, E_{\mathbf{w}}^u)$ associated with $\mathbf{w}$ with vertex set $V_{\mathbf{w}} = \text{supp}(\mathbf{w})$ and (undirected) edge set*

$$E_{\mathbf{w}}^u = \{\{w_i, w_{i+1}\} : i = 1, \dots, k-1\};$$

- 5) *A directed graph $G_{\mathbf{w}}^d = (V_{\mathbf{w}}, E_{\mathbf{w}}^d)$ associated with $\mathbf{w}$ with vertex set $V_{\mathbf{w}} = \text{supp}(\mathbf{w})$ and (directed) edge set*

$$E_{\mathbf{w}}^d = \{(w_i, w_{i+1}) : i = 1, \dots, k-1\}.$$

We recall that a walk on a graph $G = (E, V)$ is a sequence of edges connecting a sequence of vertices. A path is a walk in which all vertices are distinct. Finally,

a traversal is a walk that visits each vertex at least once. Any word w defines a traversal on its associated graph G_w^u and for any edge $e \in E_w^u$, we set N_w^e as the number of times e is traversed in the traversal defined by w . We call a word $w = w_1 \dots w_k$ *closed* if $w_1 = w_k$. We call a closed word $w \in \mathcal{W}_d^{k+1}$ a *Wigner word* if $\text{wt}(w) = \frac{k}{2} + 1$ and $N_w^e \geq 2$ for each $e \in E_w^u$. We note that by [AGZ10], this in fact implies $N_w^e = 2$ for any edge in the undirected graph associated with a Wigner word. Finally, two words $w, u \in \mathcal{W}_d$ are called *equivalent* if there is a bijection on $\{1, \dots, d\}$ that maps the words onto each other.

3.0.4 Non-commutative laws and Schwinger-Dyson equations

We now introduce the language of non-commutative probability with focus on non-commutative laws and Schwinger-Dyson equations.

Definition 3.0.4 (Involution). *A map $*$: $\mathcal{A} \rightarrow \mathcal{A}$ is an involution on a unital algebra $\mathcal{A}$ if*

- 1) $(a^*)^* = a$;
- 2) $(a + b)^* = a^* + b^*$;
- 3) $(ab)^* = b^*a^*$;
- 4) $(\lambda a)^* = \bar{\lambda}a^*$ for all $\lambda \in \mathbb{C}$;

Definition 3.0.5. *Let $X_1, \dots, X_d$ be non-commutative indeterminates. Define $\mathcal{C}_d := \mathbb{C}\langle X_1, \dots, X_d \rangle$ to be the space of polynomials in the indeterminates with complex coefficients. Let $*$ be an involution on $\mathcal{C}_d$, such that the indeterminates are self-adjoint with respect to $*$ i.e. $X_i^* = X_i$. Given a word $w \in \mathcal{W}_d$ we will write $X_w := X_{w_1} \dots X_{w_n}$, which may be extended by linearity to all $w \in \mathcal{A}_d$ with $X_\emptyset = 1$. This induces an algebra isomorphism $\mathcal{A}_d \rightarrow \mathcal{C}_d$.*

Definition 3.0.6 (Non-commutative law). *A non-commutative law $\tau \in \mathcal{C}_d^*$, the algebraic dual of $\mathcal{C}_d$, is a linear functional on $\mathbb{C}\langle X_1, \dots, X_d \rangle$ taking values in $\mathbb{C}$ such that*

- 1) *Positivity: for all $P \in \mathbb{C}\langle X_1, \dots, X_d \rangle$, $\tau(P^*) = \overline{\tau(P)}$ and $\tau(PP^*) \geq 0$;*
- 2) *Mass: $\tau(1) = 1$;*
- 3) *Trace: for all $P, Q \in \mathbb{C}\langle X_1, \dots, X_d \rangle$, $\tau(PQ) = \tau(QP)$;*
- 4) *Boundedness: there exists an $R > 0$ such that $|\tau(X_w)| < R^{|w|}$ for all $w \in \mathcal{W}_d$.*

A general non-commutative probability space is a pair $(\mathcal{A}, \varphi)$ of a unital C^* algebra $\mathcal{A}$ equipped with a state φ . Indeed, given a non-commutative law $\tau \in \mathcal{C}_d^*$, there exists a non-commutative probability space $(\mathcal{A}, \varphi)$ and self-adjoint elements

$a_1, \dots, a_d \in \mathcal{A}$ such that $\tau(P(X_1, \dots, X_d)) = \varphi(P(a_1, \dots, a_d))$ for any polynomial P . Restricting our analysis to the set of non-commutative laws [Definition 3.0.6](#) will be enough for our purposes.

Definition 3.0.7. A sequence $(\tau_n)_{n=1}^\infty$ of non-commutative laws is said to converge weakly to $\tau \in \mathcal{C}_d$ if

$$\lim_{n \rightarrow \infty} \tau_n(P) = \tau(P), \text{ for all } P \in \mathcal{C}_d.$$

It can readily be checked that the properties defining a non-commutative law are closed under weak convergence, and so any weak limit of non-commutative laws is itself a non-commutative law. It turns out that several important non-commutative laws are characterised by various integration by parts formulae. To motivate this perspective we mention a famous example from the commutative world: the Gaussian integration by parts formula. Consider a vector $(X_1, \dots, X_d) \sim \mathcal{N}(0, I_d)$ of commutative Gaussian random variables, and a sufficiently smooth function $f : \mathbb{R}^d \rightarrow \mathbb{R}$. Then it is known that

$$\mathbb{E}[X_i f(X_1, \dots, X_d)] = \mathbb{E}[\partial_i f(X_1, \dots, X_d)]. \quad (3.11)$$

Now, taking f to be some (commutative) monomial $X_i^n X_w$, where w is devoid of the letter i , then

$$\mathbb{E}[X_i^{n+1} X_w] = n \mathbb{E}[X_i^{n-1} X_w]. \quad (3.12)$$

This inductive relation is exactly the defining property of Wick's formula:

$$\mathbb{E}[X_w] = \sum_{\pi \in \mathcal{P}_2^{|w|}} \prod_{(i,j) \in \pi} \mathbb{E}[X_{w_i} X_{w_j}],$$

where $\mathcal{P}_2^n$ denotes the set of pair partitions on n letters. Indeed, the factor of n in (3.12) is exactly the number of ways of pairing the new X_i with one X_i from X_i^n . Thus, combinatorial characterisation of the Gaussian law, Wick's formula, and the integration by parts characterisation, (3.11), are equivalent. To extend this observation to non-commutative laws we first need suitable notions of non-commutative derivatives. In the following, $\mathcal{C}_d \otimes \mathcal{C}_d$ is the usual tensor product of vector spaces.

Definition 3.0.8 (Free Difference Quotient). For $d \in \mathbb{N}$ and $1 \leq i \leq d$ we define the free difference quotient $\partial_i : \mathcal{C}_d \rightarrow \mathcal{C}_d \otimes \mathcal{C}_d$ by

$$\partial_i PQ := \partial_i P \times (1 \otimes Q) + (P \otimes 1) \times \partial_i Q,$$

with $P \otimes Q \times R \otimes S := PR \otimes QS$ and $\partial_i X_j := (1 \otimes 1)\delta_{ij}$. In particular, for a word $w \in \mathcal{W}_d$ we have

$$\partial_i X_w = \sum_{w=uv} X_u \otimes X_v. \quad (3.13)$$

Definition 3.0.9 (Cyclic Derivative). Define $\mathfrak{m} : \mathcal{C}_d \otimes \mathcal{C}_d \rightarrow \mathcal{C}_d$ by $\mathfrak{m}(P \otimes Q) := QP$, then the cyclic derivative $D_i : \mathcal{C}_d \rightarrow \mathcal{C}_d$ is given by

$$D_i P := (\mathfrak{m} \circ \partial_i)P.$$

In particular, for a word $w \in \mathcal{W}_d$ we have

$$D_i X_w = \sum_{w=uv} X_v X_u.$$

Via the bijection $\mathcal{C}_d \rightarrow \mathcal{A}_d$, the operators ∂_i and D_i induce well-defined maps on $\mathcal{A}_d \rightarrow \mathcal{A}_d \otimes \mathcal{A}_d$ and $\mathcal{A}_d \rightarrow \mathcal{A}_d$ respectively. As such, we will often write $\partial_i X_w = X_{\partial_i w}$ or $D_i X_w = X_{D_i w}$. We are now ready to give the non-commutative version of an integration by parts rule.

Definition 3.0.10 (Schwinger-Dyson Equations). Let $\tau : \mathcal{C}_d \rightarrow \mathbb{C}$ be a non-commutative law. We say that τ satisfies a Schwinger-Dyson equation with conjugates $P_i \in \mathcal{C}_d$ if for all $1 \leq i \leq d$ and $Q \in \mathcal{C}_d$,

$$\tau \otimes \tau(\partial_i Q) = \tau(Q P_i), \quad (3.14)$$

where $\tau \otimes \tau(P \otimes Q) := \tau(P)\tau(Q)$.

Perhaps the most well-known and important example of a Schwinger-Dyson equation is the one satisfied by freely independent semicircular random variables. Free independence is a non-commutative notion of independence, where joint moments are determined by non-crossing partitions rather than the usual product rule for classically independent random variables. To illustrate this, suppose that X_1 and X_2 are freely independent with mean zero and law τ , then $\tau(X_1 X_2 X_1 X_2) = 0$. By contrast, in the commutative world with ordinary independence, it would be equal to $\mathbb{E}[X_1^2]\mathbb{E}[X_2^2] \neq 0$ as long as X_1 and X_2 are non-trivial.

Example 3.0.11 (Semicircular Variables). The law τ of d -free semicircular random variables satisfies the Schwinger-Dyson equation with conjugates X_i :

$$\tau \otimes \tau(\partial_i Q) = \tau(Q X_i).$$

And so for a word $w \in \mathcal{W}_d$,

$$\tau(X_w) = \sum_{w=uv} \tau(X_u)\tau(X_v).$$

Iterating this expression gives the analogue of Wick's formula for freely independent semicircular random variables:

$$\tau(X_w) = \sum_{\pi \in \text{NC}_2^{|w|}} \prod_{(i,j) \in \pi} \tau(X_{w_i} X_{w_j}),$$

where NC_2^n is the set of non-crossing pair partitions of n letters.

An important class of Schwinger-Dyson equations including [Example 3.0.11](#) are those where the conjugate polynomials P_i are given by the cyclic derivatives of some polynomial $P \in \mathcal{C}_d$. In this case, we have

$$\tau \otimes \tau(\partial_i Q) = \tau(Q D_i P). \quad (3.15)$$

The non-commutative laws satisfying the Schwinger-Dyson equation of type (3.15) arise as the asymptotic limit of the non-commutative laws of interacting random matrix ensembles. For a self-adjoint polynomial $P \in \mathcal{C}_d$, consider the law μ_P^N on $N \times N$ Hermitian matrices given by

$$d\mu_P^N(X_1^N, \dots, X_d^N) = \frac{1}{Z_P^N} \exp \left\{ -N \text{Tr}(P(X_1^N, \dots, X_d^N)) \right\} dX^N,$$

where

$$dX^N = \prod_{1 \leq j \leq k \leq N} d\Re(X^N(jk)) \prod_{1 \leq j < k \leq N} d\Im(X^N(jk)),$$

and Z_P^N is a normalising constant. In order to ensure the partition function Z_P^N is finite for all N , some restrictions on P are required. Assume that P takes the form $P = \frac{1}{2} \sum_{i=1}^d X_i^2 + W$ for a self-adjoint polynomial $W = \sum_{i=1}^m g_i X_{w_i}$. We say that W is c -convex if the map

$$(X_\nu^N(jk))_{\substack{1 \leq j \leq k \leq N \\ 1 \leq \nu \leq d}} \mapsto \text{Tr}(W(X^N)) + \frac{1-c}{2} \sum_{i=1}^d \text{Tr}(X_i^2) \quad (3.16)$$

is convex for all N when viewed as a function of $(\mathbb{R}^{N^2})^d$. Denote by U_W the set of $\mathbf{g} = (g_1, \dots, g_d) \subseteq \mathbb{C}^d$ for which [Eq. \(3.16\)](#) holds. For $\varepsilon > 0$, we also write B_ε for the set of coefficients $\mathbf{g}$ for which $\|\mathbf{g}\|_\infty \leq \varepsilon$. Recalling that $\text{tr}(\cdot) := \frac{1}{N} \text{Tr}(\cdot)$ defines the

normalised trace, we now summarise a collection of results from [GM06; GM07], see also [Gui19, Chapter 7].

Theorem 3.0.12 ([GM06; GM07]). *For every self-adjoint c -convex $W = \sum_{i=1}^m g_i X_{w_i}$ and $R > 2$ there exists $\varepsilon > 0$, such that for all $\mathbf{g} \in U_W \cap B_\varepsilon$*

$$\lim_{N \rightarrow \infty} \mathbb{E}_{\mu_P^N} [\text{tr}(X_{\mathbf{w}}^N)] = \tau_P(X_{\mathbf{w}}),$$

for every $\mathbf{w} \in \mathcal{W}_d$ where τ_P solves the Schwinger-Dyson equation Eq. (3.15). Moreover, there is exactly one solution $\tau_P \in \mathcal{C}_d^*$ to Eq. (3.15) and it satisfies

$$|\tau_P(X_{\mathbf{w}})| \leq R^{|\mathbf{w}|} \text{ for all } \mathbf{w} \in \mathcal{W}_d.$$

Note that when $W = 0$, the distribution μ_P^N is the Gaussian Unitary Ensemble (GUE) and the limiting law τ_P is the law of d free semicircular random variables. From now on, we will only consider those P satisfying the assumptions of Theorem 3.0.12. In the following let $\lambda_{\max}^N(A)$ be the maximum of the absolute values of the eigenvalues of an $N \times N$ matrix A , and let $\lambda_{\max}^N(\mathbf{A})$ be the maximum of $(\lambda_{\max}^N(A_i))_{i=1}^d$. We will need the following lemma to be able to exchange certain limits and expectations.

Lemma 3.0.13 (Sub-Gaussian tail of largest eigenvalue). *Let V be a potential satisfying the conditions of Theorem 3.0.12, and for each N suppose that $(A_i)_{i=1}^d \sim \mu_P^N$. Then there exists $\alpha > 0$ and $t_0 < \infty$ such that, for every $t \geq t_0$ and N ,*

$$\mathbb{P}_{\mu_P^N} [\lambda_{\max}^N(\mathbf{A}) \geq t] \leq e^{-\alpha N t^2}.$$

Proof. See Appendix A. □

3.1 The signature kernel is a universal limit

We consider first the case of $\text{GL}(N; \mathbb{R})$ -developments. In [MLS23], it was shown that the signature kernel between $\gamma, \sigma \in C^{1\text{-var}}([0, T], V)$ may be recovered from the scaling limit of these random developments driven by γ and σ . Indeed, if ξ_N is the Gaussian measure on $\text{Hom}(V, \mathcal{M}_N(\mathbb{R}))$ with variance $\frac{1}{N}$, then one obtains

$$K_{\text{sig}}^{\gamma, \sigma}(T, T) = \lim_{N \rightarrow \infty} \frac{1}{N} \mathbb{E}_{M \sim \xi_N} [\langle \mathcal{G}(\gamma; M), \mathcal{G}(\sigma; M) \rangle_{\text{HS}}]. \quad (3.17)$$

The proof of Eq. (3.17) in [MLS23] is specialised to the Gaussian setting and relies heavily on Isserli's Theorem. We identify a random matrix theory approach inspired by [AGZ10], where we generalise the preceding in the following ways.

- 1) We replace the space $\mathrm{GL}(N; \mathbb{R})$ with $\mathrm{GL}(N; \mathbb{C})$, whose Lie algebra is $\mathfrak{g}_N = \mathcal{M}_N(\mathbb{C})$;
- 2) We replace the Gaussian assumption on the entries of the matrices with a sub-Gaussian assumption on the operator norm of the matrices;
- 3) For fixed indices $(m, l) \in \{1, \dots, N\}^2$ we require that $A_i^N(m, l)$ and $A_j^N(m, l)$ are uncorrelated and not necessarily independent.

Moreover, our approach results in shorter proof than [MLS23] and allows for different limits to be derived for randomised CDEs with a different matrix structure, see Section 3.2. In the following we will use $\|\cdot\|_{\mathrm{op}}$ to denote the operator norm of matrices.

If one fixes a basis for V , then every $M \in \mathrm{Hom}(V, \mathfrak{g}_N)$ may be written in the form

$$M(v) = \sum_{j=1}^d A_j v^j, \quad \text{for } A_i \in \mathfrak{g}_N,$$

so that we have a canonical isomorphism between $\mathrm{Hom}(V, \mathfrak{g}_N)$ and $\mathfrak{g}_N^d$. In particular, the spaces of Borel probability measures $\mathcal{P}(\mathrm{Hom}(V, \mathfrak{g}_N))$ and $\mathcal{P}(\mathfrak{g}_N^d)$ are in bijection. Additionally, the development in Eq. (3.3) can then be written as

$$dZ_t = \sum_{i=1}^d Z_t A_i d\gamma_t^i, \quad Z_0 = I_N.$$

Theorem 3.1.1. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space supporting for each $N \in \mathbb{N}$ and $1 \leq i \leq d$, a random matrix $A_i^N : \Omega \rightarrow \mathfrak{g}_N$. Assume that for all $k \in \mathbb{N}$,*

$$\sup_{N \in \mathbb{N}} \sup_{1 \leq i \leq d} \sup_{1 \leq m, l \leq N} \mathbb{E}[|A_i^N(m, l)|^k] \leq c_A(k) < \infty, \quad (3.18)$$

and that $\mathbb{E}[A_i^N(m, l)] = 0$ and $\mathbb{E}[|A_i^N(m, l)|^2] = 1$. Suppose also that there exists some $\kappa > 0$ such that for every $p \in \mathbb{N}$

$$\sup_{N \in \mathbb{N}} \sup_{1 \leq i \leq d} \mathbb{E} \left[\frac{1}{N^{\frac{p}{2}}} \|A_i^N\|_{\mathrm{op}}^p \right] \leq \kappa^p \Gamma(\frac{p}{2} + 1), \quad (3.19)$$

where $\Gamma(\cdot)$ is the Gamma function. For each $N \in \mathbb{N}$, we assume that

$$\bigvee_{i=1}^d \sigma(A_i^N(m, l)), \quad \text{for } 1 \leq m, l \leq N, \quad (3.20)$$

are independent sigma algebras and that

$$\mathbb{E} \left[A_i^N(m, l) \overline{A_j^N(m, l)} \right] = 0, \quad \text{for } i \neq j. \quad (3.21)$$

For each N , let ξ_N be the law of this collection in $\mathcal{P}(\mathfrak{g}_N^d)$. Let $\gamma, \sigma \in C^{1-var}([0, T], V)$ and let $Z_\gamma^N(t), Z_\sigma^N(t)$ be the $GL(N; \mathbb{C})$ -valued solutions to

$$dZ_\gamma^N(t) = \frac{1}{\sqrt{N}} \sum_{i=1}^d Z_\gamma^N(t) A_i^N d\gamma_t^i, \quad Z_\gamma^N(0) = I_N \quad (3.22)$$

$$dZ_\sigma^N(t) = \frac{1}{\sqrt{N}} \sum_{i=1}^d Z_\sigma^N(t) A_i^N d\sigma_t^i, \quad Z_\sigma^N(0) = I_N. \quad (3.23)$$

Then for every $s, t \in [0, T]$

$$\lim_{N \rightarrow \infty} \frac{1}{N} \mathbb{E}_{\xi_N} [\langle Z_\gamma^N(s), Z_\sigma^N(t) \rangle_{HS}] = K_{sig}^{\gamma, \sigma}(s, t). \quad (3.24)$$

As in [MLS23], the proof relies on the expansion of Eq. (3.22) as a convergent infinite series of integrals, and a careful analysis of the trace of products of random matrices from the set $\{A_i^N : 1 \leq i \leq d\}$ in the limit as $N \rightarrow \infty$.

Remark 3.1.2. Special cases of the following lemma can be deduced from existing results in the literature. For example, both the complex and real Gaussian case may be derived from [DP21, Sections 2 and 4.4], while matrices with i.i.d. entries satisfying a fourth moment condition can also be found in [DP21, Section 4.1].

Notation 3.1.3. For $A \in \mathcal{M}_N(\mathbb{C})$, A^* will denote the conjugate transpose of A . The Hilbert-Schmidt inner product on $\mathcal{M}_N(\mathbb{C})$ is then given by

$$\langle A, B \rangle_{HS} = \text{Tr}(A^* B).$$

Lemma 3.1.4. Suppose that for each $N \in \mathbb{N}$, $\{A_i^N : 1 \leq i \leq d\}$ is a collection of d matrices satisfying the assumptions of Theorem 3.1.1. Let $n, m \geq 0$ be non-negative integers and let $\mathbf{u} = u_1 \dots u_n \in \mathcal{W}_d^n, \mathbf{v} = v_1 \dots v_m \in \mathcal{W}_d^m$, with corresponding matrix products

$$\begin{aligned} A_{\mathbf{u}}^N &:= A_{u_1}^N \dots A_{u_n}^N \\ A_{\mathbf{u}^* \mathbf{v}}^N &:= A_{\mathbf{u}}^{N*} A_{\mathbf{v}}^N \end{aligned}$$

Then, setting $k = n + m$,

$$\lim_{N \rightarrow \infty} \frac{1}{N^{\frac{k}{2}}} \mathbb{E} \left[\text{tr} (A_{\mathbf{u}^* \mathbf{v}}^N) \right] = \begin{cases} 1, & \text{if } \mathbf{u} = \mathbf{v}, \\ 0, & \text{otherwise.} \end{cases} \quad (3.25)$$

We use the convention that if $|\mathbf{u}| = 0$, then $A_{\mathbf{u}}^N = I_N$, the $N \times N$ identity matrix.

Proof. We start by expanding the LHS of Eq. (3.25) in terms of the individual elements composing the trace. We write

$$\frac{1}{N^{\frac{k}{2}}} \mathbb{E} \left[\text{tr} (A_{\mathbf{u}^* \mathbf{v}}^N) \right] = \frac{1}{N^{\frac{k}{2}}} \sum_{\mathbf{w}} \mathcal{C}_{\mathbf{w}}, \quad (3.26)$$

where the sum is over closed words $\mathbf{w} = w_1 w_2 \dots w_n w_{n+1} \dots w_k w_1 \in \mathcal{W}_N^{k+1}$ of length $k + 1 := n + m + 1$, and where

$$\mathcal{C}_{\mathbf{w}} := \mathbb{E} \left[A_{u_n}^{N^*}(w_1, w_2) A_{u_{n-1}}^{N^*}(w_2, w_3) \dots A_{u_1}^{N^*}(w_n, w_{n+1}) A_{v_1}^N(w_{n+1}, w_{n+2}) \dots A_{v_m}^N(w_k, w_1) \right]$$

Since $\mathbf{w}$ is a closed word, it defines a traversal on the undirected graph $G_{\mathbf{w}}^u$. As in [AGZ10][Theorem 5.4.2, Lemma 2.1.6], we can deduce that

- 1) $\mathcal{C}_{\mathbf{w}}$ vanishes unless $N_{\mathbf{w}}^e \geq 2$ for all $e \in E_{\mathbf{w}}^u$;
- 2) The number of N -words in the equivalence class of any given N -word $\mathbf{w}$ with $\text{wt}(\mathbf{w}) = t$, is $N(N-1) \dots (N-t+1) \leq N^t$. This can be seen by counting the size of the image of $\mathbf{w}$ under all bijections on $\{1, \dots, N\}$;
- 3) The number of equivalence classes of closed N -words $\mathbf{w}$ of length $k+1$ with $\text{wt}(\mathbf{w}) = t$ for which $N_{\mathbf{w}}^e \geq 2$ for every $e \in E_{\mathbf{w}}^u$ is bounded above by $t^k \leq k^k$, since there are at most t choices for each letter excluding the final one;
- 4) The weight of a closed word of length $k+1$ with $N_{\mathbf{w}}^e \geq 2$ for every $e \in E_{\mathbf{w}}^u$ is at most $\frac{k}{2} + 1$.

It follows that

$$\left| \sum_{\mathbf{w}: \text{wt}(\mathbf{w}) \leq \frac{k}{2}} \mathcal{C}_{\mathbf{w}} \right| \leq \sum_{t \leq \frac{k}{2}} N^t c_A(k) t^k \leq C(k) N^{\frac{k}{2}},$$

and combining this with Eq. (3.26) we see

$$\left| \frac{1}{N^{\frac{k}{2}}} \mathbb{E} \left[\text{tr} (A_{\mathbf{u}^* \mathbf{v}}^N) \right] - \frac{1}{N^{\frac{k}{2}+1}} \sum_{\mathbf{w}: \text{wt}(\mathbf{w}) = \frac{k}{2}+1} \mathcal{C}_{\mathbf{w}} \right| \leq \frac{C(k)}{N}.$$

Note that the set $\{w : \text{wt}(w) = \frac{k}{2} + 1\}$ is empty whenever k is odd. Hence if $n + m$ is odd, we have that

$$\lim_{N \rightarrow \infty} \frac{1}{N^{\frac{k}{2}}} \mathbb{E} \left[\text{tr}(A_{u^*v}^N) \right] = 0.$$

If k is even and w is a closed word of length $k + 1$ with $\text{wt}(w) = \frac{k}{2} + 1$, then G_w^u is a tree [AGZ10][Chapter 2.1.3] with $N_w^e = 2$ for every $e \in E_w^u$. As such, $\mathcal{C}_w$ is a product of covariances which will either be 1 or 0. There are two situations in which this product can vanish. Firstly, whenever the two traversals of an edge $e \in E_w^u$ correspond to different matrices. Second, if both traversals are from the same matrix, but one traversal corresponds to an entry above the diagonal and the other to an entry below the diagonal. It is the latter reason that differentiates our problem from the Hermitian setting.

Since G_w^u is a tree, the traversal defined by w traverses each edge once in each direction. Let $(w_r, w_{r+1}) \in E_w^d$ correspond to the first traversal of $\{w_r, w_{r+1}\}$. Then, for some $s \in \{r + 1, \dots, k\}$, $w_s = w_{r+1}$ and $w_{s+1} = w_r$. In other words $(w_s, w_{s+1}) = (w_{r+1}, w_r)$. If $s \leq n$, the corresponding component of $\mathcal{C}_w$ is given by

$$\mathbb{E} \left[A_{u_{n+1-r}}^{N^*}(w_r, w_{r+1}) A_{u_{n+1-s}}^{N^*}(w_s, w_{s+1}) \right] = \mathbb{E} \left[A_{u_{n+1-r}}^{N^*}(w_r, w_{r+1}) A_{u_{n+1-s}}^{N^*}(w_{r+1}, w_r) \right] = 0.$$

Similarly if $r \geq n + 1$, the term is given by

$$\mathbb{E} \left[A_{v_{r-n}}^N(w_r, w_{r+1}) A_{v_{s-n}}^N(w_s, w_{s+1}) \right] = \mathbb{E} \left[A_{v_{r-n}}^N(w_r, w_{r+1}) A_{v_{s-n}}^N(w_{r+1}, w_r) \right] = 0.$$

So it must be that $r \leq n$ and $s \geq n + 1$. Consequently for every edge, one traversal comes from u and one from v meaning that $|u| = |v|$. Furthermore, since G_w^u is a tree, there can be no repeated letters in $w_1 \dots w_{n+1}$ as otherwise there would be two traversals of an edge coming from u . Hence, the walk corresponding to the sequence of vertices $w_1, \dots, w_{n+1}$ is a path, as is the walk on the sequence $w_{n+1}, w_{n+2}, \dots, w_1$. Consider the vertex w_{n+1} ; we must traverse the edge (w_{n+1}, w_n) at some point on the path $w_{n+1}, w_{n+2}, \dots, w_1$. Since the path cannot visit w_{n+1} again, in order to traverse the edge (w_{n+1}, w_n) it must be that $w_{n+2} = w_n$. By induction, the only way for each edge to be traversed twice is when $w_{n+1-r} = w_{n+1+r}$ for $r = 1, \dots, n$. And so there is exactly one equivalence class of words for which $\mathcal{C}_w$ is non-zero: those where G_w^u is a tree with one branch. When this holds, the term for the traversals of an edge $\{w_{n+1-r}, w_{n+2-r}\}$ in $\mathcal{C}_w$ is

$$\begin{aligned} & \mathbb{E} \left[A_{u_{n+1-(n+1-r)}}^{N^*}(w_{n+1-r}, w_{n+1-(r-1)}) A_{v_{(n+1+(r-1))-n}}^N(w_{n+1+(r-1)}, w_{n+1+r}) \right] \\ &= \mathbb{E} \left[A_{u_r}^N(w_{n+1-(r-1)}, w_{n+1-r}) A_{v_r}^N(w_{n+1+(r-1)}, w_{n+1+r}) \right] = \delta_{u_r v_r}. \end{aligned}$$

This implies $u = v$ for the limit to be non-trivial. We conclude that

$$\begin{aligned} \lim_{N \rightarrow \infty} \frac{1}{N^{\frac{k}{2}}} \mathbb{E}[\text{tr}(A_{u^*v}^N)] &= \lim_{N \rightarrow \infty} \frac{1}{N^{\frac{k}{2}+1}} \sum_{w: \text{wt}(w) = \frac{k}{2}+1} \mathcal{C}_w \\ &= \lim_{N \rightarrow \infty} \frac{N(N-1) \dots (N - \frac{k}{2})}{N^{\frac{k}{2}+1}} \delta_{uv} = \delta_{uv}. \end{aligned}$$

□

Proof of Theorem 3.1.1. As in [MLS23], by considering the expansion of the ordered-exponential we see that

$$\lim_{N \rightarrow \infty} \frac{1}{N} \mathbb{E}_{\xi_N} [\langle Z_\gamma^N(s), Z_\sigma^N(t) \rangle_{\text{HS}}] = \lim_{N \rightarrow \infty} \mathbb{E}_{\xi_N} \left[\sum_{|u|, |v|=0}^{\infty} \frac{1}{N^{\frac{|u^*v|}{2}}} \text{tr}(A_{u^*v}^N) \mathcal{S}_{0,s}^u(\gamma) \mathcal{S}_{0,t}^v(\sigma) \right]. \quad (3.27)$$

In order to apply Lemma 3.1.4, we would like to exchange the limit and expectation and the double sum. By Fubini-Tonelli and the dominated convergence theorem, to justify the exchange it is enough to show that

$$\sum_{|u|, |v|=0}^{\infty} \frac{1}{N^{\frac{|u^*v|}{2}}} \mathbb{E}_{\xi_N} \left[\left| \text{tr}(A_{u^*v}^N) \right| \right] |\mathcal{S}_{0,s}^u(\gamma) \mathcal{S}_{0,t}^v(\sigma)| \quad (3.28)$$

is uniformly bounded in N . By the assumptions of Theorem 3.1.1,

$$\frac{1}{N^{\frac{|u^*v|}{2}}} \mathbb{E}_{\xi_N} \left[\left| \text{tr}(A_{u^*v}^N) \right| \right] \leq \frac{1}{N^{\frac{|u^*v|}{2}}} \mathbb{E}_{\xi_N} \left[\|A_{u^*v}^N\|_{\text{op}} \right] \leq \kappa^{|u^*v|} \Gamma\left(\frac{|u^*v|}{2} + 1\right).$$

Then, by the factorial decay of the signature and the fact that $\sqrt{d}^n$ times the L_2 norm is at least as large as the L_1 norm on $V^{\otimes n}$, there exists some $\omega > 0$ for which Eq. (3.28) is bounded by

$$\begin{aligned} \sum_{n,m=0}^{\infty} \frac{\Gamma(\frac{n+m}{2} + 1) (\omega \kappa)^{m+n}}{\Gamma(n+1) \Gamma(m+1)} &\leq \sum_{n,m=0}^{\infty} \frac{\sqrt{\Gamma(n+1) \Gamma(m+1)} (\omega \kappa)^{n+m}}{\Gamma(n+1) \Gamma(m+1)} \\ &= \left(\sum_{n=0}^{\infty} \frac{(\omega \kappa)^n}{\sqrt{\Gamma(n+1)}} \right)^2 < \infty. \end{aligned}$$

The inequality follows from logarithmic convexity and Jensen's inequality. Hence by an exchange of limits and an application of [Lemma 3.1.4](#), we see that

$$\begin{aligned}
 \lim_{N \rightarrow \infty} \frac{1}{N} \mathbb{E}_{\xi_N} [\langle Z_\gamma^N(s), Z_\sigma^N(t) \rangle_{\text{HS}}] &= \sum_{|u|, |v|=0}^{\infty} \lim_{N \rightarrow \infty} \frac{1}{N^{\frac{|u^*v|}{2}}} \mathbb{E}_{\xi_N} [\text{tr}(A_{u^*v}^N)] \mathcal{S}_{0,s}^u(\gamma) \mathcal{S}_{0,t}^v(\sigma) \\
 &= \sum_{|u|=0}^{\infty} \mathcal{S}_{0,s}^u(\gamma) \mathcal{S}_{0,t}^u(\sigma) \\
 &= \langle \mathcal{S}_{0,s}(\gamma), \mathcal{S}_{0,t}(\sigma) \rangle_{\text{HS}} \\
 &= K_{\text{sig}}^{\gamma, \sigma}(s, t).
 \end{aligned}$$

□

As in [\[MLS23\]](#) we can also prove that convergence occurs in L^2 and not just in expectation.

Theorem 3.1.5. *Let $(\Omega, \mathcal{F}, \mathbb{P})$, A_i^N , γ, σ and Z_γ^N, Z_σ^N be as in [Theorem 3.1.1](#), then*

$$\frac{1}{N} \langle Z_\gamma^N(s), Z_\sigma^N(t) \rangle_{\text{HS}} \xrightarrow{L_2} K_{\text{sig}}^{\gamma, \sigma}(s, t). \quad (3.29)$$

Proof. By expanding

$$\mathbb{E}_{\xi_N} \left[\left(\frac{1}{N} \langle Z_\gamma^N(s), Z_\sigma^N(t) \rangle_{\text{HS}} - \langle \mathcal{S}_{0,s}(\gamma), \mathcal{S}_{0,t}(\sigma) \rangle_{\text{HS}} \right)^2 \right]$$

in the fashion of [Eq. \(3.27\)](#), we see that the objective is to prove

$$\begin{aligned}
 \lim_{N \rightarrow \infty} \mathbb{E}_{\xi_N} \left[\sum_{|u|, |v|, |k|, |l|=0}^{\infty} \left(\frac{1}{N^{\frac{|u^*v|}{2}}} \text{tr}(A_{u^*v}^N) - \delta_{uv} \right) \right. \\
 \left. \cdot \left(\frac{1}{N^{\frac{|l^*k|}{2}}} \text{tr}(A_{k^*l}^N) - \delta_{kl} \right) \mathcal{S}_{0,s}^u(\gamma) \mathcal{S}_{0,t}^v(\sigma) \mathcal{S}_{0,s}^k(\gamma) \mathcal{S}_{0,t}^l(\sigma) \right] = 0. \quad (3.30)
 \end{aligned}$$

As in the proof of [Theorem 3.1.1](#), we can show that exchanging the limit and expectation with the quadruple sum is justified. Multiplying out the brackets in [Eq. \(3.30\)](#) and taking limits of terms already dealt with in [Theorem 3.1.1](#) reduces the problem to showing

$$\begin{aligned}
 \sum_{|u|, |v|, |k|, |l|=0}^{\infty} \left(\lim_{N \rightarrow \infty} \frac{1}{N^{\frac{|u^*v| + |k^*l|}{2}}} \mathbb{E}_{\xi_N} [\text{tr}(A_{u^*v}^N) \text{tr}(A_{k^*l}^N)] \right. \\
 \left. - \delta_{uv} \delta_{kl} \right) \mathcal{S}_{0,s}^u(\gamma) \mathcal{S}_{0,t}^v(\sigma) \mathcal{S}_{0,s}^k(\gamma) \mathcal{S}_{0,t}^l(\sigma) = 0. \quad (3.31)
 \end{aligned}$$

Thus it remains to show that

$$\lim_{N \rightarrow \infty} \frac{1}{N^{\frac{|u^*v|+|k^*l|}{2}}} \mathbb{E}_{\xi_N} \left[\text{tr}(A_{u^*v}^N) \text{tr}(A_{k^*l}^N) \right] = \delta_{uv} \delta_{kl}.$$

Given closed words $w = w_1 \dots w_n w_1$ and $z = z_1 \dots z_m z_1$, we define the graphs $G_{w,z}^u := G_w^u \cup G_z^u$. Now we can write

$$\frac{1}{N^{\frac{|u^*v|+|k^*l|}{2}+2}} \mathbb{E}_{\xi_N} \left[\text{tr}(A_{u^*v}^N) \text{tr}(A_{k^*l}^N) \right] = \frac{1}{N^{\frac{|u^*v|+|k^*l|}{2}+2}} \sum_{w,z} \mathcal{C}_{w,z}, \quad (3.32)$$

where $w = w_1 w_2 \dots w_r w_{r+1} \dots w_n w_1$ and $u = u_1 u_2 \dots u_s u_{s+1} \dots u_m u_1$ and

$$\begin{aligned} \mathcal{C}_{w,z} := \mathbb{E}_{\xi_N} \left[A_{u_{|u|}}^{N^*}(w_1, w_2) \dots A_{u_1}^{N^*}(w_r, w_{r+1}) A_{v_1}^N(w_{r+1}, w_{r+2}) \dots A_{j_{|v|}}^N(w_n, w_1) \right. \\ \left. A_{k_{|k|}}^{N^*}(z_1, z_2) \dots A_{k_1}^{N^*}(z_s, z_{s+1}) A_{l_1}^N(z_{s+1}, z_{s+2}) \dots A_{l_{|l|}}^N(z_m, z_1) \right]. \end{aligned}$$

As in Lemma 3.1.4, $N_{w,z}^e$ must be at least 2 for every $e \in E_{w,z}^u$ in order for $\mathcal{C}_{w,z}$ not to vanish. This implies that $|E_{w,z}^u| \leq \frac{n+m}{2}$, and since $G_{w,z}^u$ has at most two connected components, the restriction $\text{wt}(wz) = |V_{w,z}^u| \leq \frac{n+m}{2} + 2$ is imposed. Moreover, $\text{wt}(wz) = \frac{n+m}{2} + 2$ if and only if $G_{w,z}^u$ has two connected components, one with $\frac{n}{2} + 1$ vertices and the other with $\frac{m}{2} + 1$ vertices. In other words, G_w^u and G_z^u are disjoint trees. In this case

$$\mathcal{C}_{w,z} = \mathcal{C}_w(A_{u^*v}^N) \mathcal{C}_z(A_{k^*l}^N),$$

since the words are disjoint. This reduces the problem to the one handled in Lemma 3.1.4. This product will be non-zero if and only if $u = v$ and $k = l$. Since there are $N(N-1) \dots (N - \frac{|u^*v|+|k^*l|}{2} - 1)$ members of the equivalence class of wz , it holds that

$$\begin{aligned} \lim_{N \rightarrow \infty} \frac{\mathbb{E}_{\xi_N} \left[\text{tr}(A_{u^*v}^N) \text{tr}(A_{k^*l}^N) \right]}{N^{\frac{|u^*v|+|k^*l|}{2}}} &= \lim_{N \rightarrow \infty} \frac{N(N-1) \dots (N - \frac{|u^*v|+|k^*l|}{2} - 1)}{N^{\frac{|u^*v|+|k^*l|}{2}+2}} \delta_{uv} \delta_{kl} \\ &= \delta_{uv} \delta_{kl}. \end{aligned}$$

□

Remark 3.1.6. The assumptions of [Theorems 3.1.1](#) and [3.1.5](#) are satisfied in the case of independent entries of sub-Gaussian random variables under the condition

$$\sup_{N \in \mathbb{N}} \sup_{1 \leq i \leq d} \sup_{1 \leq m, l \leq N} \left\| A_i^N(m, l) \right\|_{\psi} < \infty,$$

where $\|\cdot\|_{\psi}$ denotes the sub-Gaussian norm. In this case, it is clear that [Eq. \(3.18\)](#) holds and it can be verified from [\[Ver18, Theorem 4.4.5\]](#) that [Eq. \(3.19\)](#) holds.

Remark 3.1.7. In [\[MLS23\]](#) the authors in fact do not consider path developments directly (they also consider non-linear equations), rather randomised $\mathbb{R}^N$ -valued CDEs. Consider the equations

$$\begin{aligned} dX_{\gamma}^N(t) &= \frac{1}{\sqrt{N}} \sum_{i=1}^d A_i^N X_{\gamma}^N(t) d\gamma_t^i, \quad X_{\gamma}^N(0) = \eta^N, \\ dX_{\sigma}^N(t) &= \frac{1}{\sqrt{N}} \sum_{i=1}^d A_i^N X_{\sigma}^N(t) d\sigma_t^i, \quad X_{\sigma}^N(0) = \eta^N, \end{aligned}$$

where both the components of η^N and the pair $\{(A_i^N)_{i=1}^d, \eta^N\}$ are independent. As long as the components of η^N have zero mean and unit variance, then it is straightforward to adjust the proof of [Theorem 3.1.1](#) to show that

$$\lim_{N \rightarrow \infty} \mathbb{E} \left[\frac{1}{N} \langle X_{\gamma}^N(s), X_{\sigma}^N(t) \rangle \right] = K_{sig}^{\gamma, \sigma}(s, t).$$

If one further assumes that

$$\sup_{N \in \mathbb{N}} \sup_{1 \leq m \leq N} \mathbb{E} \left[|\eta^N(m)|^4 \right] < \infty,$$

then the convergence also happens in L^2 by adapting the proof of [Theorem 3.1.5](#). We note that the corresponding [Eq. \(3.31\)](#) will involve additional terms of the form

$$\lim_{N \rightarrow \infty} \frac{1}{N^{\frac{|u^* v| + |k^* l|}{2}}} \mathbb{E} \left[\text{tr}(A_{u^* v}^N A_{k^* l}^N) \right].$$

Due to the scaling factor outside the equation, one can use the techniques from [Theorem 3.1.1](#) to show that the limit of such terms is zero.

3.2 Universal limit of Wigner unitary developments

It was proposed recently in [\[LLN23\]](#) to use the unitary developments of a path as the basis for a discriminator in generative models. This can be seen as an extension

of the characteristic function discriminator for vector valued data proposed by [Li+20a; FSS20]. Consider the unitary Lie algebra and group of degree N defined by

$$\mathfrak{u}_N = \{A \in \mathcal{M}_N(\mathbb{C}) : A^* = -A\}, \quad U(N; \mathbb{C}) = \{A \in \mathcal{M}_N(\mathbb{C}) : A^*A = I_N\}. \quad (3.33)$$

Note that $\mathfrak{u}_N = i\mathfrak{h}_N$, where $\mathfrak{h}_N$ is the set of $N \times N$ Hermitian matrices. For any path $\gamma \in C^{1\text{-var}}([0, T], V)$ and $M \in \text{Hom}(V, \mathfrak{u}_N)$, the unitary feature of γ under M is the time- T solution to the CDE

$$dZ_t = Z_t \cdot M(d\gamma_t), \quad Z_0 = I_N. \quad (3.34)$$

The solution $\mathcal{U}(\gamma; M) := Z_T$ is a special case of the Cartan development of γ , taking values in the Lie group $U(N; \mathbb{C})$. Let μ be a Borel probability measure on $C^{1\text{-var}}([0, T], V)$ and $\gamma \sim \mu$, then the path characteristic function (PCF) of μ evaluated at M is defined as

$$\Phi_\mu(M) := \mathbb{E}_{\gamma \sim \mu} [\mathcal{U}(\gamma; M)]. \quad (3.35)$$

Definition 3.2.1 (PCFD [LLN23]). Let μ, ν be probability measures on $C^{1\text{-var}}([0, T], V)$ with $\gamma \sim \mu$ and $\sigma \sim \nu$, and let ξ be a probability measure on $\text{Hom}(V, \mathfrak{u}_N)$ with $M \sim \xi$. The path characteristic function distance (PCFD) between μ and ν with respect to ξ is defined as

$$d_\xi^2(\mu, \nu) := \mathbb{E}_{M \sim \xi} [\|\Phi_\mu(M) - \Phi_\nu(M)\|_{\text{HS}}^2]. \quad (3.36)$$

For fundamental results on PCFD that allow for its use as a loss function in applications as well as certain theoretical properties, we refer the reader to [LLN23]. In [LLN23, Proposition B.10] it was shown that $d_\xi(\mu, \nu)$ is a maximum mean discrepancy (MMD) distance on the space $\mathcal{P}(C^{1\text{-var}}([0, T], V))$, the set of Borel probability measures on $C^{1\text{-var}}([0, T], V)$.

Definition 3.2.2. Given a kernel function $\kappa : C^{1\text{-var}}([0, T], V) \times C^{1\text{-var}}([0, T], V) \rightarrow \mathbb{R}$, then the MMD distance associated to κ is a map

$$\text{MMD}_\kappa : \mathcal{P}(C^{1\text{-var}}([0, T], V)) \times \mathcal{P}(C^{1\text{-var}}([0, T], V)) \rightarrow \mathbb{R}^+,$$

which is defined by

$$\text{MMD}_\kappa^2(\mu, \nu) := \mathbb{E}_{\gamma, \gamma' \stackrel{\text{iid}}{\sim} \mu} [\kappa(\gamma, \gamma')] + \mathbb{E}_{\sigma, \sigma' \stackrel{\text{iid}}{\sim} \nu} [\kappa(\sigma, \sigma')] - 2\mathbb{E}_{\gamma \sim \mu, \sigma \sim \nu} [\kappa(\gamma, \sigma)]. \quad (3.37)$$

Proposition 3.2.3 (PCFD as an MMD: Proposition B.10 [LLN23]). *Let $\xi \in \mathcal{P}(\mathfrak{u}_N^d)$, $\mu, \nu \in \mathcal{P}(C^{1-var}([0, T], V))$, then $d_\xi(\mu, \nu) = \text{MMD}_\kappa(\mu, \nu)$, where*

$$\kappa(\gamma, \sigma) := \mathbb{E}_{M \sim \xi} \left[\text{Tr}(\mathcal{U}(\gamma \star \overleftarrow{\sigma}; M)) \right]. \quad (3.38)$$

One of the difficulties in the use of PCFD for the learning of a distribution on high-dimensional or complex paths is the computational cost of computing $\mathcal{U}(\gamma; M)$, $M \in \text{Hom}(V, \mathfrak{u}_n)$ for large N , especially for a fully supported measure on $\text{Hom}(V, \mathfrak{u}_n)$. The authors of [LLN23] propose two solutions: firstly they employ a reciprocal learning approach where they embed high dimensional paths into a lower dimensional space where the embedding should act as a type of pseudo-inverse of the original generator; secondly, they adversarially learn a finitely supported empirical measure on $\text{Hom}(V, \mathfrak{u}_n)$. We propose an alternative approach, where we take the dimension $N \rightarrow \infty$ for a class of measures supported on the whole space. The limiting kernel then solves a functional equation driven by $\gamma \star \overleftarrow{\sigma}$. Before we state and prove our result, we introduce non-crossing pair partitions and a new associated notion of generations.

Remark 3.2.4. *For any fixed $N < \infty$, the authors of [LLN23] remark that directly computing Eq. (3.36) is linear in sample size, whereas computing it as an MMD through Eq. (3.38) is quadratic in sample size.*

3.2.1 Non-crossing partitions, Dyck words and trees

From this point on, we will consider the alphabet A_n to be the ordered set $\{1, \dots, n\}$. However, all the definitions may be carried over to any ordered set of n letters. An element of a partition of A_n will be called a “part”.

Definition 3.2.5. *A partition π of A_n is called crossing if there exist $1 \leq a < b < c < d \leq n$ for which a, c belongs to one part of π and b, d to another part. A partition that is not crossing is called non-crossing.*

Of particular importance to us is the class of non-crossing pair partitions.

Definition 3.2.6. *A pair partition of A_n is one where each part is a two element set. We denote by $\text{NC}_2(n)$ the set of non-crossing pair partitions of A_n .*

For a non-crossing pair partition π and $\{i, j\} \in \pi$, we define $\pi_{\{i, j\}} = \pi_i = \pi_j := \min\{i, j\}$. A convenient method of encoding non-crossing pair partitions is through Dyck words. We recall their definition from [Tao12, Section 2].

Definition 3.2.7 (Dyck Words). *A Dyck word is a sequence of left and right parentheses, “(” and “)”, such that at every stage, when read from left-to-right, there are at least as many left parentheses as right parentheses and exactly the same number of each at the end of the word. We denote by $\mathcal{D}_n$ the set of Dyck words of length n .*

Note that $\mathcal{D}_n = \emptyset$ if n is odd. There is a well-established bijection $\text{NC}_2(n) \rightarrow \mathcal{D}_n$ where one reads the letters $1 \dots n$ from left to right and for each letter, if it is the first element in a pair, one writes a left parenthesis “(”, and if it is the second letter in a pair one writes a right parenthesis “)”. For a proof that this map indeed defines a bijection see [Tao12, Lemma 2.3.15].

It is also helpful to note the bijection between non-crossing pair partitions $\text{NC}_2(n)$ and the set of non-planar rooted trees $\mathcal{T}_n$ with vertices $\{1, \dots, \frac{n}{2} + 1\}$ and $\frac{n}{2}$ edges, where a left-to-right traversal of the tree visits the vertices in lexicographical order. Recall that a left-to-right traversal starts at the root node and proceeds with the condition that for any two subtrees of any given node, all vertices of the left subtree must be visited before the vertices of the right subtree.

Once again, we refer the reader to [Tao12, Section 2] and [AGZ10, Chapter 2] for full details on these bijections. An example of these bijections for a non-crossing pair-partition of A_8 is illustrated below.

$$\{\{1, 6\}, \{2, 3\}, \{4, 5\}, \{7, 8\}\} \equiv \begin{array}{c} \bullet \quad \bullet \quad \bullet \quad \bullet \quad \bullet \quad \bullet \quad \bullet \quad \bullet \\ 1 \quad 2 \quad 3 \quad 4 \quad 5 \quad 6 \quad 7 \quad 8 \end{array} \mapsto ((\))(\)) \mapsto \begin{array}{c} \bullet \quad \bullet \quad \bullet \quad \bullet \quad \bullet \\ 3 \quad 2 \quad 4 \quad 5 \quad 1 \end{array}.$$

Generations

We are interested in finding a unique sequence of steps to generate any non-crossing pair partition π starting from the partition $\{\{1, 2\}\}$ of $\{1, 2\}$, or equivalently any Dyck word from $()$ or any tree from $\bullet_1^2$. We define the generation of a pair of parentheses in a Dyck word recursively by:

- 1) The pair whose right parenthesis, “)”, is the final parenthesis in the Dyck word has generation 1;
- 2) A pair of parentheses is of generation $k + 1$ if the parenthesis in the Dyck word following the right-parenthesis of the pair is part of a pair of generation k .

Alternatively we can define the generation of an edge of a tree via

- 1) The right-most edge attached to the root has generation 1;

- 2) An edge is of generation $k + 1$ if it is the right-most edge attached above an edge of generation k , or it is the right-most edge attached to the left of an edge of generation k .

Example 3.2.8. Consider the partition $\pi \in \text{NC}_2(12)$

$$\pi = \begin{array}{cccccccccccc} \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \end{array} .$$

The Dyck word and tree corresponding to π are given by

$$())((())()), \text{ and } \begin{array}{c} \bullet 7 \\ | \\ \bullet 6 \\ / \quad \backslash \\ \bullet 5 \quad \bullet 4 \\ / \quad \backslash \\ \bullet 2 \quad \bullet 3 \\ \backslash \quad / \\ \bullet 1 \end{array} . \quad (3.39)$$

Here, generation one parentheses or edges have been drawn in red, generation two have been drawn in blue and generation three have been drawn in green.

The generation of a Dyck word is the maximum of the generations of its constituent pairs of parentheses. Likewise, the generation of a tree is the maximum of the generations of its edges. The generation of a pair in a non-crossing pair partition is the generation of the corresponding pair of parentheses of its Dyck word. Similarly, the generation of the partition is defined as the generation of the corresponding Dyck word. For each $k \in \mathbb{N}$, define G_k to be the set of non-crossing pair partitions (Dyck words or trees) of generation k . For any non-crossing pair partition π , we let $g(\pi)$ be the set of parts (pairs or edges) whose generation is equal to the generation of π . In the case $k = 0$, we also define $G_0 := \{\emptyset\}$.

Remark 3.2.9. *It can be shown that the notion of generation and its properties are preserved by the aforementioned bijection between Dyck words and trees.*

For any Dyck word d , we let $g(d)$ be the set of pairs of parentheses with maximal generation. Additionally, we define $G(d)$ to be the set of Dyck words obtained by inserting a pair “ $()$ ” immediately to the left of at least one parenthesis that is part of a pair in $g(d)$.

Lemma 3.2.10. *The following facts about generations hold.*

- 1) For any non-crossing pair partition π , any $\{i, j\} \in g(\pi)$ is of the form $\{i, i + 1\}$;
- 2) Let $d_1 \neq d_2$ be Dyck words of generation k , then $G(d_1) \cap G(d_2) = \emptyset$;
- 3) We have that $G_{n+1} = \bigsqcup_{\pi \in G_n} G(\pi)$.

Proof. If any pair of parentheses in a Dyck word is not consecutive, then they cannot have the highest generation: if they are not consecutive, they must enclose a pair that has a higher generation. Noting that consecutively paired parentheses in a Dyck word correspond to consecutively paired letters in a non-crossing pair partition proves the first point. Since insertion only happens immediately to the **left** of a parenthesis of generation k , the generation of an inserted pair must be of generation $k + 1$, and moreover any pair of generation $k + 1$ must have been inserted. Hence by deleting the pairs of generation $k + 1$ in $d \in G(d_1) \cap G(d_2)$ we should obtain both d_1 and d_2 , from which we deduce $d_1 = d_2$. The third point is an immediate corollary of the second. $\square$

3.2.2 A limiting functional equation

Recall that $u_N = i\mathfrak{h}_N$ so that as in Section 3.1, Eq. (3.34) can be written in the form

$$dZ_t = i \sum_{j=1}^d Z_t A_j d\gamma_t^j, \quad Z_0 = I_N,$$

for a collection of d matrices in $\mathfrak{h}_N$. Just as we required a version of the derivative for functions in non-commutative indeterminates, we will also need the notion of derivatives of functionals on path space. There are many such choices, but we will use the Dupire or vertical derivative [DT23; CGP25]. This notion of derivative has been popular in deriving Taylor-like expansions of functions on path space.

Definition 3.2.11 (Vertical derivative). *Let $f : C^{1-var}([0, T], V) \rightarrow \mathbb{R}$ be a suitably regular functional, $1 \leq i \leq d$, and $h^i \in C^{1-var}([0, T], V)$ the path defined by $t \mapsto \frac{ht}{T}e^i$ with $\{e^i\}_{i=1}^d$ basis vectors in V . The vertical derivative of f at a path γ along the direction i is defined by*

$$\nabla^i f(\gamma) := \lim_{h \rightarrow 0} \frac{f(\gamma \star h^i) - f(\gamma)}{h}.$$

For a word $w = w_1 \cdots w_k \in \mathcal{W}_d$ of length k , define

$$\nabla^w := \nabla^{w_1} \cdots \nabla^{w_k} f(\gamma).$$

Proposition 3.2.12 (Proposition 3.8 [DT23]). *For any path $\gamma \in C^{1-var}([0, T], V)$ and $v, w \in \mathcal{W}_d$, the vertical derivative satisfies*

$$\nabla^v \mathcal{S}_{s,t}^w(\gamma) = \begin{cases} \mathcal{S}_{s,t}^u(\gamma), & w = uv, \\ 0, & \text{otherwise.} \end{cases}$$

We are now ready to state the main result of this section.

Theorem 3.2.13. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space supporting for each $N \in \mathbb{N}$ and $1 \leq i \leq d$, a Hermitian random matrix $A_i^N : \Omega \rightarrow \mathfrak{h}_N$. Assume that for all $k \in \mathbb{N}$,*

$$\sup_{N \in \mathbb{N}} \sup_{1 \leq i \leq d} \sup_{1 \leq m, l \leq N} \mathbb{E}[|A_i^N(m, l)|^k] \leq c_A(k) < \infty, \quad (3.40)$$

and that $\mathbb{E}[A_i^N(m, l)] = 0$ and $\mathbb{E}[|A_i^N(m, l)|^2] = 1$. Suppose also that there exists some $\kappa > 0$ such that for every $p \in \mathbb{N}$

$$\sup_{N \in \mathbb{N}} \sup_{1 \leq i \leq d} \mathbb{E}\left[\frac{1}{N^{\frac{p}{2}}} \|A_i^N\|_{\text{op}}^p\right] \leq \kappa^p \Gamma(\frac{p}{2} + 1), \quad (3.41)$$

where $\Gamma(\cdot)$ is the Gamma function. For each $N \in \mathbb{N}$ we assume that

$$\bigvee_{i=1}^d \sigma(A_i^N(m, l)) \quad \text{for } 1 \leq m \leq l \leq N, \quad (3.42)$$

are independent sigma algebras and that

$$\mathbb{E}[A_i^N(m, l) \overline{A_j^N(m, l)}] = 0, \quad \text{for } i \neq j. \quad (3.43)$$

For each N , let ξ_N be the law of the collection in $\mathcal{P}(\mathfrak{h}_N^d)$. Let $\gamma \in C^{1\text{-var}}([0, T], V)$ and let $Z_\gamma^N(s, t)$ be the $U(N; \mathbb{C})$ -valued solution to

$$dZ_\gamma^N(s, t) = \frac{i}{\sqrt{N}} \sum_{j=1}^d Z_\gamma^N(s, t) A_j^N d\gamma_t^j, \quad Z_\gamma^N(s, s) = I_N. \quad (3.44)$$

Then for $0 \leq s \leq t \leq T$, $\langle \gamma_{s,t} \rangle_{\mu_G^\infty} := \lim_{N \rightarrow \infty} \mathbb{E}_{\xi_N} [\text{tr}(Z_\gamma^N(s, t))]$ exists and solves the functional equation

$$\langle \gamma_{s,t} \rangle_{\mu_G^\infty} = 1 - \int_s^t \int_s^r \langle \gamma_{s,u} \rangle_{\mu_G^\infty} \langle \gamma_{u,r} \rangle_{\mu_G^\infty} \langle d\gamma_u, d\gamma_r \rangle, \quad \langle \gamma_{u,u} \rangle_{\mu_G^\infty} = 1 \text{ for all } u \in [0, T]. \quad (3.45)$$

Remark 3.2.14. *The use of the notation $\langle \gamma_{\cdot, \cdot} \rangle_{\mu_G^\infty}$ will be justified in [Section 3.5](#).*

Proof. As in the proof of [Theorem 3.1.1](#), we may expand the solution $Z_\gamma(s, t)$ in terms of the signature of γ . Indeed, we may write

$$\text{tr}(Z_\gamma(s, t)) = \sum_{|u|=0}^{\infty} \frac{i^{|u|}}{N^{\frac{|u|}{2}}} \text{tr}(A_u^N) \mathcal{S}_{s,t}^u(\gamma).$$

The same techniques in the proof of [Theorem 3.1.1](#) show that the limits, expectations and sums may be interchanged. Meanwhile, [[AGZ10](#), Theorem 5.4.2] gives

$$\lim_{N \rightarrow \infty} \frac{1}{N^{\frac{|u|}{2}}} \mathbb{E}_{\xi_N} \left[\text{tr}(A_{\mathbf{u}}^N) \right] = \tau(X_{\mathbf{u}}),$$

where τ is the law of d free semi-circular random variables. Hence

$$\langle \gamma_{s,t} \rangle_{\mu_G^\infty} = \lim_{N \rightarrow \infty} \mathbb{E}_{\xi_N} \left[\text{tr}(Z_\gamma^N(s, t)) \right] = \sum_{|\mathbf{u}|=0}^{\infty} i^{|\mathbf{u}|} \tau(X_{\mathbf{u}}) \mathcal{S}_{s,t}^{\mathbf{u}}(\gamma). \quad (3.46)$$

We recall from [Section 3.0.4](#) that the non-commutative law τ of d -free semi-circular random variables is uniquely characterised by the Schwinger-Dyson equation

$$\tau \otimes \tau(\partial_i P) = \tau(P X_i). \quad (3.47)$$

Consider first the following

$$\begin{aligned} \sum_{\mathbf{w} \in \mathcal{W}_d} i^{|\mathbf{w}|+1} \tau \otimes \tau(\partial_i X_{\mathbf{w}}) \mathcal{S}_{s,t}^{\mathbf{w}}(\gamma) &= \sum_{i=1}^d \sum_{\mathbf{u}, \mathbf{v} \in \mathcal{W}_d} i^{|\mathbf{u}|+|\mathbf{v}|+2} \tau(X_{\mathbf{u}}) \tau(X_{\mathbf{v}}) \mathcal{S}_{s,t}^{uiv}(\gamma) \\ &= - \sum_{\mathbf{u}, \mathbf{v} \in \mathcal{W}_d} i^{|\mathbf{u}|+|\mathbf{v}|} \tau(X_{\mathbf{u}}) \tau(X_{\mathbf{v}}) \int_s^t \mathcal{S}_{s,u}^{\mathbf{u}}(\gamma) \mathcal{S}_{u,t}^{\mathbf{v}}(\gamma) d\gamma_u^i \\ &= - \int_s^t \left(\sum_{\mathbf{u} \in \mathcal{W}_d} i^{|\mathbf{u}|} \tau(X_{\mathbf{u}}) \mathcal{S}_{s,u}^{\mathbf{u}}(\gamma) \right) \left(\sum_{\mathbf{v} \in \mathcal{W}_d} i^{|\mathbf{v}|} \tau(X_{\mathbf{v}}) \mathcal{S}_{u,t}^{\mathbf{v}}(\gamma) \right) d\gamma_u^i \\ &= - \int_s^t \langle \gamma_{s,u} \rangle_{\mu_G^\infty} \langle \gamma_{u,t} \rangle_{\mu_G^\infty} d\gamma_u^i. \end{aligned}$$

By applying the right-hand side of the Schwinger Dyson equation [Eq. \(3.47\)](#) we deduce that

$$\begin{aligned} \sum_{\mathbf{w} \in \mathcal{W}_d} i^{|\mathbf{w}|+1} \tau \otimes \tau(\partial_i X_{\mathbf{w}}) \mathcal{S}_{s,t}^{\mathbf{w}}(\gamma) &= \sum_{\mathbf{w} \in \mathcal{W}_d} i^{|\mathbf{w}|+1} \tau(X_{\mathbf{w}} X_i) \mathcal{S}_{s,t}^{\mathbf{w}}(\gamma) \\ &= \sum_{\mathbf{w} \in \mathcal{W}_d} i^{|\mathbf{w}|} \tau(X_{\mathbf{w}}) \nabla^i \mathcal{S}_{s,t}^{\mathbf{w}}(\gamma) \\ &= \nabla^i \sum_{\mathbf{w} \in \mathcal{W}_d} i^{|\mathbf{w}|} \tau(X_{\mathbf{w}}) \mathcal{S}_{s,t}^{\mathbf{w}}(\gamma) \\ &= \nabla^i \langle \gamma_{s,t} \rangle_{\mu_G^\infty}. \end{aligned}$$

Combining the two derivations gives

$$\nabla^i \langle \gamma_{s,t} \rangle_{\mu_G^\infty} = - \int_s^t \langle \gamma_{s,u} \rangle_{\mu_G^\infty} \langle \gamma_{u,t} \rangle_{\mu_G^\infty} d\gamma_u^i,$$

and integrating both sides with respect $d\gamma^i$ yields

$$\nabla^i \int_s^t \langle \gamma_{s,r} \rangle_{\mu_G^\infty} d\gamma_r^i = - \int_s^t \int_s^r \langle \gamma \rangle_{s,u} \langle \gamma \rangle_{u,r} d\gamma_u^i d\gamma_r^i.$$

Since

$$\langle \gamma_{s,t} \rangle_{\mu_G^\infty} = 1 + \sum_i \nabla^i \int_s^t \langle \gamma_{s,r} \rangle_{\mu_G^\infty} d\gamma_r^i,$$

we conclude that

$$\langle \gamma_{s,t} \rangle_{\mu_G^\infty} = 1 - \int_s^t \int_s^r \langle \gamma_{s,u} \rangle_{\mu_G^\infty} \langle \gamma_{u,r} \rangle_{\mu_G^\infty} \langle d\gamma_u d\gamma_r \rangle.$$

□

Lemma 3.2.15. *For any $\gamma \in C^{1-var}([0, T], V)$, the functional equation*

$$K(s, t) = 1 - \int_s^t \int_s^r K(s, u) K(u, r) \langle d\gamma_u, d\gamma_r \rangle, \quad K(u, u) = 1 \text{ for all } u \in [0, T], \quad (3.48)$$

has a unique solution $\langle \gamma_{s,t} \rangle_{\mu_G^\infty}$ for $0 \leq s \leq t \leq T$.

Due to the quadratic nature of the equation, we adopt the approach of iteratively expanding the equation and proving convergence to the series in Eq. (3.46).

Proof. Existence of a solution is settled by Theorem 3.2.13. To prove uniqueness, we show that any solution K must agree with the solution K_γ described in Theorem 3.2.13. Fix a solution K to Eq. (3.48), then it must be uniformly bounded on the simplex by the structure of Eq. (3.48). Define $\kappa = \max\{1, \|K\|_\infty\}$. We now introduce some notation. For any $\pi \in \text{NC}_2(2k)$, recall that π_i is the minimum value of the pair to which i belongs. We then define following.

- 1) The time-ordered simplex $\Delta^\pi[s, t] := \{s = t_0 \leq t_1 \leq \dots \leq t_{2k} \leq t\}$;
- 2) The “differential” $d\gamma^\pi := (-1)^k \sum_{j_{\pi_1}, \dots, j_{\pi_{2k}}=1}^d d\gamma_{t_1}^{j_{\pi_1}} d\gamma_{t_2}^{j_{\pi_2}} \dots d\gamma_{t_{2k}}^{j_{\pi_{2k}}}$. Note that the sum is over k free indices since $j_{\pi_i} = j_{\pi_j}$ if $\{i, j\} \in \pi$;
- 3) The integrand

$$K^\pi := \prod_{\{i, i+1\} \in g(\pi)} K(t_{i-1}, t_i) K(t_i, t_{i+1}).$$

Hence, for any even k and $\pi \in \text{NC}_2(k)$, the integral

$$I(\pi) := \int_{\Delta^\pi[s, t]} K^\pi d\gamma^\pi$$

is well defined. We also define

$$\mathcal{S}_{s,t}^\pi(\gamma) := \int_{\Delta^\pi[s,t]} d\gamma^\pi.$$

As usual, we set $\mathcal{S}_{s,t}^\emptyset(\gamma) \equiv 1$. We recall that G_k is the set of non-crossing pair-partitions of generation k . Using this language, K satisfies the equation

$$K(s, t) = \sum_{\pi \in G_0} \mathcal{S}_{s,t}^\pi(\gamma) + \sum_{\pi \in G_1} I(\pi), \quad (3.49)$$

since $G_0 = \{\emptyset\}$ and $G_1 = \{\{1, 2\}\}$. We now claim that

$$K(s, t) = \sum_{\pi \in \bigcup_{i=0}^{n-1} G_i} \mathcal{S}_{s,t}^\pi(\gamma) + \sum_{\pi \in G_n} I(\pi). \quad (3.50)$$

To see this, fix $\pi \in G_n$ and for each $\{i, i+1\} \in g(\pi)$, the terms in the product K_γ^π associated with $\{i, i+1\}$ are given by

$$K(t_{i-1}, t_i)K(t_i, t_{i+1}) = \left(1 - \int_{t_{i-1}}^{t_i} \int_{t_{i-1}}^{r_1} K(t_{i-1}, u_1)K(u_1, r_1) \sum_{l_1=1}^d d\gamma_{u_1}^{l_1} d\gamma_{r_1}^{l_1} \right) \\ \left(1 - \int_{t_i}^{t_{i+1}} \int_{t_i}^{r_2} K(t_i, u_2)K(u_2, r_2) \sum_{l_2=1}^d d\gamma_{u_2}^{l_2} d\gamma_{r_2}^{l_2} \right).$$

From each bracket we may now either select the constant 1 or the double integral. We see that the four possible selections correspond to replacing the pair $()$, corresponding to $\{i, i+1\}$ in the Dyck word associated with π , with $()$, $()()$, $(())$ or $()(())$. By applying these selections to each pair of brackets in K^π , we see that

$$I(\pi) = \mathcal{S}_{s,t}^\pi(\gamma) + \sum_{\pi' \in G(\pi)} I(\pi').$$

Now

$$\begin{aligned}
 K(s, t) &= \sum_{\pi \in \bigcup_{i=0}^n G_i} \mathcal{S}_{s,t}^\pi(\gamma) + \sum_{\pi \in G_{n+1}} I(\pi) \\
 &= \sum_{|\pi| \leq 2n} \mathcal{S}_{s,t}^\pi(\gamma) + \sum_{\substack{\pi \in \bigcup_{i=0}^n G_i \\ |\pi| > 2n}} \mathcal{S}_{s,t}^\pi(\gamma) + \sum_{\pi \in G_{n+1}} I(\pi) \\
 &= \sum_{|u|=0}^{2n} (-1)^{\frac{|u|}{2}} \mathcal{S}_{s,t}^u(\gamma) \sum_{\pi \in \text{NC}_2(|u|)} \prod_{\{i,j\} \in \pi} \mathbb{1}_{u_i=u_j} + \sum_{\substack{\pi \in \bigcup_{i=0}^n G_i \\ |\pi| > 2n}} \mathcal{S}_{s,t}^\pi(\gamma) + \sum_{\pi \in G_{n+1}} I(\pi) \\
 &= \sum_{|u|=0}^{2n} (-1)^{\frac{|u|}{2}} \tau(X_u) \mathcal{S}_{s,t}^u(\gamma) + \sum_{\substack{\pi \in \bigcup_{i=0}^n G_i \\ |\pi| > 2n}} \mathcal{S}_{s,t}^\pi(\gamma) + \sum_{\pi \in G_{n+1}} I(\pi),
 \end{aligned}$$

where the final equality follows from another characterisation of the moments of d -free semicircular random variables, see for example, [AGZ10, Chapter 5 pp.380-381]. We observe that the first term converges to $\langle \gamma_{s,t} \rangle_{\mu_G^\infty}$, so it remains to show that the second two sums converge to 0 as $n \rightarrow \infty$. Here we note that by the uniqueness of the construction of Dyck words, for a word v with $|v| > 2n$, the differential d^v appears at most $\tau(X_v)$ times in the second two sums. And so we have that

$$\begin{aligned}
 \left| \sum_{\substack{\pi \in \bigcup_{i=0}^n G_i \\ |\pi| > 2n}} \mathcal{S}_{s,t}^\pi(\gamma) + \sum_{\pi \in G_{n+1}} I(\pi) \right| &\leq \sum_{|v| > 2n} \tau(X_v) \kappa^{|J|} \left| \mathcal{S}_{s,t}^v(\gamma) \right| \\
 &\leq \sum_{|v| > 2n} C_{\frac{|v|}{2}} \kappa^{|v|} \left| \mathcal{S}_{s,t}^v(\gamma) \right| \\
 &\rightarrow 0,
 \end{aligned}$$

as $n \rightarrow \infty$, where C_k is the k^{th} Catalan number. The convergence is due to the factorial decay of the signature and the well known fact that $C_k = \mathcal{O}(4^k)$. And so the solution is unique. $\square$

The results of this section show that the limit of $d_{\xi_N}^2(\mu, \nu)$ exists under mild conditions on the sequence $(\xi_N)_{N \in \mathbb{N}}$. The limit is an MMD, with kernel given by

$$K_{\text{SD}}(\gamma, \sigma) := \langle y_{0,T} \rangle_{\mu_G^\infty}, \quad \text{with } y := \gamma * \overleftarrow{\sigma}. \quad (3.51)$$

We leave it to future work to investigate the discriminative properties of K_{SD} and how it compares with the adversarially learnt d_ξ proposed by [LLN23]. We

note that replacing d_ξ with the MMD based on K_{SD} is akin to the non-adversarial training of neural-SDEs proposed by [Iss+23] using the MMD based on K_{sig} .

3.3 Numerical Schemes

In order for K_{SD} to have practical application, we require some method of approximating the solution to Eq. (3.48) without the need for computing the signature $\mathcal{S}(\gamma)$. We propose explicit and implicit numerical schemes based on discriminations.

3.3.1 Explicit Scheme

We propose a straightforward explicit numerical scheme to solve Eq. (3.48). Indeed, suppose that γ is absolutely continuous, then we may use the Lebesgue-Differentiation theorem and a mixed endpoint approximation to the resulting Riemann-Stieltjes to arrive at the following discretisation

$$\begin{aligned} K_{s,t+\Delta} &= 1 - \int_s^{t+\Delta} \int_s^r K_{s,u} K_{u,r} \langle d\gamma_u, d\gamma_r \rangle \\ &= K_{s,t} - \int_t^{t+\Delta} \int_s^r K_{s,u} K_{u,r} \langle d\gamma_u, d\gamma_r \rangle \\ &\approx K_{s,t} - \int_s^t K_{s,u} K_{u,t} \langle d\gamma_u, \gamma_{t,t+\Delta} \rangle \\ &\approx K_{s,t} - \sum_{[u,v] \in \mathcal{Q}} K_{s,u} K_{v,t} \langle \delta\gamma_{u,v}, \delta\gamma_{t,t+\Delta} \rangle du, \end{aligned}$$

where $\delta\gamma_{u,v}$ denotes the increment $\gamma_v - \gamma_u$ and $\mathcal{Q}$ is a partition of $[s, t]$. Let $\mathcal{P} = \{0 = t_0 < t_1 < \dots < t_\ell = T\}$ be a partition of the interval $[0, T]$ and for $0 < n \leq \ell$ define

$$\delta_n \gamma = \gamma_{t_n} - \gamma_{t_{n-1}}.$$

The above discretisation naturally leads to the recursion

$$K_{n,m+1} = K_{n,m} - \sum_{n < r < m+1} K_{n,r-1} K_{r,m} \langle \delta_r \gamma_r, \delta_{m+1} \gamma \rangle, \quad (3.52)$$

for $0 \leq n < m < \ell$, with the boundary condition $K_{n,n+1} = 1$ for all n .

Remark 3.3.1. The use of a mixed endpoint approximation to the integrand $\langle \gamma_{s,u} \rangle_{\mu_G^\infty} \langle \gamma_{u,r} \rangle_{\mu_G^\infty}$ is motivated by a natural connection to the iterated sums signature (ISS) of [KO19; DET20b;

[DET20a](#)]. In fact the recursion [Eq. \(3.52\)](#) is exactly the recursion found for the functional

$$\kappa_{n,m} = \sum_{w \in \mathcal{W}_d} i^{|w|} \tau(X_w) \text{ISS}_{n,m}^w(\gamma),$$

where $\text{ISS}_{n,m}(\gamma)$ is the iterated sums signature of γ on the partition $\mathcal{P}$ between times n and m . This connection allows one to use established results concerning the ISS to prove convergence of $K_{0,\ell}$ to $\langle \gamma_{0,T} \rangle_{\mu_{\mathcal{G}}}$ as the mesh-size of $\mathcal{P}$ vanishes, see the proof of [Theorem 3.3.2](#) for details.

Naïvely solving the recursion defined by [Eq. \(3.52\)](#) for $K_{0,\ell}$ incurs a time complexity of $\mathcal{O}(\ell^3)$ in the length of the partition ℓ . Indeed there are $\mathcal{O}(\ell^2)$ points in the solution grid $\Pi = \{0 \leq n < m \leq \ell\}$, and the sum on the right hand side of [Eq. \(3.52\)](#) contributes the final factor of ℓ . However, just as for the finite difference scheme proposed in [\[Sal+21\]](#) for the ordinary signature kernel, rather than solving the grid in row and column order we can update a whole anti-diagonal of the solution grid Π at once, since there is no shared dependency between points on the same anti-diagonal. It is possible to exploit this parallelisation using GPU architecture, provided that the length $n + 1$ of γ does not exceed the total number of threads on the GPU. The overall time taken is then $\mathcal{O}(d\ell^2)$ to populate Π , where d is the dimension of the underlying path space. The following result guarantees that the preceding scheme converges to the true solution as the mesh of the partition $\mathcal{P}$ shrinks to zero.

Theorem 3.3.2. *Let $\gamma \in C^{1\text{-var}}([0, T], V)$ and let $\mathcal{P} = \{0 = t_0 < \dots < t_\ell = T\}$, with $K_{n,m}$ as defined by [Eq. \(3.52\)](#). Then*

$$\left| \langle \gamma_{0,T} \rangle_{\mu_{\mathcal{G}}} - K_{0,\ell} \right| \leq 16 \|\gamma\|_{1,[0,T]} \exp(4 \|\gamma\|_{1,[0,T]}) \max_{1 \leq i \leq \ell} \|\gamma\|_{1,[t_i, t_{i-1}]} . \quad (3.53)$$

We follow the method of [\[KO19, Appendix A.3\]](#), wherein the authors prove convergence of numerical schemes for the ordinary signature kernel based on a discretisation of the terms in the signature. Throughout, let $\gamma \in C^{1\text{-var}}([0, T], V)$ and let $\mathcal{P} = \{0 = t_0 < \dots < t_\ell = T\}$ be a partition of $[0, T]$. We first show that [Eq. \(3.52\)](#) may be derived by applying the Schwinger-Dyson equations to a function of the iterated-sums-signature. Given a partition $\mathcal{P}$, the iterated sums signature is defined by

$$\text{ISS}_{n,m}^w(\gamma) = \sum_{n < i_1 < \dots < i_k \leq m} (\delta_{i_1} \gamma)^{w_1} \dots (\delta_{i_k} \gamma)^{w_k}, \quad (3.54)$$

for any word $w = w_1 \dots w_k$ in d -letters and $0 \leq n < m \leq \ell$. As usual, we define $\text{ISS}_{n,m}^\emptyset = 1$. Note that $\text{ISS}_{n,m}^u = 0$ whenever $|u| > n - m$. Consider now the

discretisation

$$\kappa_{n,m} := \sum_{|\mathbf{w}|=0}^{\infty} i^{|\mathbf{w}|} \tau(X_{\mathbf{w}}) \text{ISS}_{n,m}^{\mathbf{w}}(\gamma). \quad (3.55)$$

Then, as in the proof of [Theorem 3.2.13](#) we may deduce that

$$\begin{aligned} \kappa_{n,m} &= \sum_{|\mathbf{w}|=0}^{\infty} i^{|\mathbf{w}|} \tau(X_{\mathbf{w}}) \text{ISS}_{n,m}^{\mathbf{w}}(\gamma) \\ &= 1 + \sum_{j=1}^d \sum_{|\mathbf{w}|=0}^{\infty} i^{|\mathbf{w}|+1} \tau(X_{\mathbf{w}j}) \text{ISS}_{n,m}^{\mathbf{w}j}(\gamma) \\ &= 1 - \sum_{j=1}^d \sum_{|\mathbf{u}|, |\mathbf{v}|=0}^{\infty} i^{|\mathbf{u}|+1} \tau(X_{\mathbf{u}}) \tau(X_{\mathbf{v}}) \text{ISS}_{n,m}^{\mathbf{u}j\mathbf{v}j}(\gamma) \\ &= 1 - \sum_{j=1}^d \sum_{|\mathbf{u}|, |\mathbf{v}|=0}^{\infty} i^{|\mathbf{u}|+|\mathbf{v}|} \tau(X_{\mathbf{u}}) \tau(X_{\mathbf{v}}) \text{ISS}_{n,m}^{\mathbf{u}j\mathbf{v}j}(\gamma). \end{aligned}$$

Now for $\mathbf{u} = u_1 \dots u_l$, $\mathbf{v} = v_1, \dots, v_k$ and $\mathbf{i} := i_1 < \dots < i_l$ and $\mathbf{j} := j_1 < \dots < j_k$, we see that

$$\begin{aligned} \text{ISS}_{n,m}^{\mathbf{u}j\mathbf{v}j}(\gamma) &= \sum_{n < \mathbf{i} < r < \mathbf{j} < s \leq m} (\delta_{i_1} \gamma)^{u_1} \dots (\delta_{i_l} \gamma)^{u_l} (\delta_r \gamma)^j (\delta_{j_1} \gamma)^{v_1} \dots (\delta_{j_k} \gamma)^{v_k} (\delta_s \gamma)^j \\ &= \sum_{n < r < s \leq m} \text{ISS}_{n,r-1}^{\mathbf{u}} \text{ISS}_{r,s-1}^{\mathbf{v}} (\delta_r \gamma)^j (\delta_s \gamma)^j, \end{aligned}$$

so that

$$\begin{aligned} \kappa_{n,m} &= 1 - \sum_{j=1}^d \sum_{\mathbf{u}, \mathbf{v} \in \mathcal{W}_d} i^{|\mathbf{u}|+|\mathbf{v}|} \tau(\mathbf{u}) \tau(\mathbf{v}) \sum_{n < r < s \leq m} \text{ISS}_{n,r-1}^{\mathbf{u}} \text{ISS}_{r,s-1}^{\mathbf{v}} (\delta_r \gamma)^j (\delta_s \gamma)^j \\ &= 1 - \sum_{n < r < s \leq m} \left(\sum_{|\mathbf{u}|=0}^{\infty} i^{|\mathbf{u}|} \tau(\mathbf{u}) \text{ISS}_{n,r-1}^{\mathbf{u}} \right) \left(\sum_{|\mathbf{v}|=0}^{\infty} i^{|\mathbf{v}|} \tau(\mathbf{v}) \text{ISS}_{r,s-1}^{\mathbf{v}} \right) \langle \delta_r \gamma, \delta_s \gamma \rangle \\ &= 1 - \sum_{n < r < s \leq m} \kappa_{n,r-1} \kappa_{r,s-1} \langle \delta_r \gamma, \delta_s \gamma \rangle \\ &= 1 - \sum_{n < r < s \leq m-1} \kappa_{n,r-1} \kappa_{r,s-1} \langle \delta_r \gamma, \delta_s \gamma \rangle - \sum_{n < r < m} \kappa_{n,r-1} \kappa_{r,m-1} \langle \delta_r \gamma, \delta_m \gamma \rangle \\ &= \kappa_{n,m-1} - \sum_{n < r < m} \kappa_{n,r-1} \kappa_{r,m-1} \langle \delta_r \gamma, \delta_m \gamma \rangle, \end{aligned}$$

which is exactly the recursion defined in [Eq. \(3.52\)](#). It remains then to show that $\text{ISS}_{0,T}^{\mathbf{w}}(\gamma)$ is a suitable approximation to $\mathcal{S}_{0,T}^{\mathbf{w}}(\gamma)$ for any γ . To this end, define the

m^{th} term of the iterated sums signature by

$$\text{ISS}_{a,b}(\gamma)^m = \sum_{a < i_1 < \dots < i_k \leq b} (\delta_{i_1} \gamma) \otimes \dots \otimes (\delta_{i_k} \gamma).$$

The following bound on the difference between the m^{th} term of the signature of γ and its iterated sums signature may be deduced from a careful reading of [KO19, Lemma A.4 and Proposition A.5].

Lemma 3.3.3. *For all $m \geq 2$, it holds that*

$$\left\| \mathcal{S}_{0,T}(\gamma)^m - \text{ISS}_{0,\ell}(\gamma)^m \right\| \leq \frac{\|\gamma\|_{1,[0,T]}^{m-1}}{(m-2)!} \max_{1 \leq i \leq \ell} \|\gamma\|_{1,[t_{i-1}, t_i]}. \quad (3.56)$$

Sketch of Proof. By [KO19, Lemma A.4], the difference $\mathcal{S}_{0,T}(\gamma)^m - \text{ISS}_{0,\ell}(\gamma)^m$ is the m^{th} iterated integral of γ over the subset of the m -simplex where the projections of at least two coordinates lie in the same interval $[t_{i-1}, t_i] \in \mathcal{P}$. By dividing this subset into disjoint regions, the authors of [KO19] obtain a bound on the difference in terms of the 1-variation norm of γ over each sub-interval $[t_{i-1}, t_i]$. By writing $\|\gamma\|_{1,[0,T]} = \sum_{i=1}^{\ell} \|\gamma\|_{1,[t_{i-1}, t_i]}$ it is then shown in [KO19, Proposition A.5] that this bound on the difference is smaller than $\|\gamma\|_{1,[0,T]} \max_{1 \leq i \leq \ell} \|\gamma\|_{1,[t_{i-1}, t_i]}$ multiplied by the $(m-2)^{\text{th}}$ term in the Taylor expansion of $\exp(\|\gamma\|_{1,[0,T]})$. $\square$

Corollary 3.3.4. *For any $\gamma \in C^{1\text{-var}}([0, T], V)$ and $\mathcal{P}$, we have the bound*

$$\left| \langle \gamma_{0,T} \rangle_{\mu_G^\infty} - K_{0,\ell} \right| \leq 16 \|\gamma\|_{1,[0,T]} \exp(4 \|\gamma\|_{1,[0,T]}) \max_{1 \leq i \leq \ell} \|\gamma\|_{1,[t_{i-1}, t_i]}.$$

Proof. By the triangle inequality and Lemma 3.3.3, we see that

$$\begin{aligned} |K_\gamma(0, T) - K_{0,\ell}| &= \left| \sum_{|w|=0}^{\infty} (-1)^{\frac{|w|}{2}} \tau(X_w)(\gamma^\pi) (\mathcal{S}_{0,T}^w(\gamma) - \text{ISS}_{0,\ell}^w(\gamma)) \right| \\ &\leq \sum_{|w|=2}^{\infty} \tau(X_w) \left| \mathcal{S}_{0,T}^w(\gamma) - \text{ISS}_{0,\ell}^w(\gamma) \right| \\ &\leq \sum_{m=2}^{\infty} C_m \left\| \mathcal{S}_{0,T}(\gamma)^m - \text{ISS}_{0,\ell}(\gamma)^m \right\| \\ &\leq \sum_{m=2}^{\infty} 4^m \frac{\|\gamma\|_{1,[0,T]}^{m-1}}{(m-2)!} \max_{1 \leq i \leq n} \|\gamma\|_{1,[t_i, t_{i+1}]} \\ &\leq 16 \|\gamma\|_{1,[0,T]} \max_{1 \leq i \leq \ell} \|\gamma\|_{1,[t_{i-1}, t_i]} \sum_{m=0}^{\infty} \frac{(4 \|\gamma\|_{1,[0,T]})^m}{m!} \\ &= 16 \|\gamma\|_{1,[0,T]} \exp(4 \|\gamma\|_{1,[0,T]}) \max_{1 \leq i \leq \ell} \|\gamma\|_{1,[t_{i-1}, t_i]}. \end{aligned}$$

□

Remark 3.3.5. Note that for any path $\gamma \in C^{1-var}([0, T], V)$ and any sequence of partitions π_n with $|\pi_n| \rightarrow 0$, it holds that $\lim_{n \rightarrow \infty} \max_{1 \leq i \leq n} \|\gamma\|_{1;[t_i, t_{i+1}]} = 0$, so that [Theorem 3.3.2](#) guarantees convergence along any sequence of partitions with vanishing mesh size.

Remark 3.3.6 (Exact Solution for Straight Lines and $d = 1$). We fix $T = 1$ without loss of generality, since the signature is invariant under time reparameterisation. There are two situations for which the solution to [Eq. \(3.48\)](#) is explicit. Firstly, assume that γ is a straight line. That is, for $t \in [0, 1]$

$$\gamma_t = tv, \quad \text{for some } v \in \mathbb{R}^d,$$

so that $\gamma'_t \equiv v$ and $\langle d\gamma_s, d\gamma_t \rangle = \|v\|^2 ds dt$ for all $s, t \in [0, 1]$. In this case [Eq. \(3.48\)](#) reduces to

$$\langle \gamma_{s,t} \rangle_{\mu_G^\infty} = 1 - \int_s^t \int_s^r \langle \gamma_{s,u} \rangle_{\mu_G^\infty} \langle \gamma_{u,r} \rangle_{\mu_G^\infty} \|v\|^2 du dr,$$

which depends only on $\|v\|$. Hence, the resulting solution will be the same as for the $\mathbb{R}$ -valued path $\sigma_t = t\|v\|$. Recalling the expansion [Eq. \(3.46\)](#), we see that

$$\langle \gamma_{s,t} \rangle_{\mu_G^\infty} = K_\sigma(s, t) = \sum_{k=0}^{\infty} i^{2k} C_{2k} \frac{\|(t-s)v\|^{2k}}{(2k)!} = \mathbb{E} \left[\exp(iR\|(t-s)v\|) \right],$$

for a semi-circular random variable R with radius 2. The characteristic function of a semi-circular random variable is well known to be given by $\frac{J_1(2t)}{t}$, where $J_1(\cdot)$ is a Bessel function of the first kind. It is possible to use this exact solution to populate the solution grid Π at any point (n, m) for which a path γ is a straight line on the interval $[t_n, t_m]$. Finally, the solution is also explicit for any path in $d = 1$, since the signature then depends only on the increment $\gamma_1 - \gamma_0$.

3.3.2 Implicit Scheme

We also propose an implicit scheme derived from a discretisation of [Eq. \(3.48\)](#) given by

$$\begin{aligned}
 K_{s,t+\Delta} &= 1 - \int_s^{t+\Delta} \int_s^r K_{s,u} K_{u,r} \langle d\gamma_u, d\gamma_r \rangle \\
 &= K_{s,t} - \int_t^{t+\Delta} \int_s^r K_{s,u} K_{u,r} \langle d\gamma_u, d\gamma_r \rangle \\
 &\approx K_{s,t} - \int_s^{t+\Delta} K_{s,u} K_{u,t+\Delta} \langle d\gamma_u, \gamma_{t,t+\Delta} \rangle \\
 &\approx K_{s,t} - \sum_{[u,v] \in \mathcal{Q}} K_{s,v} K_{v,t} \langle \delta\gamma_{u,v}, \delta\gamma_{t,t+\Delta} \rangle du,
 \end{aligned}$$

where $\mathcal{Q}$ is again some partition of $[s, t]$. This motivates the recursion

$$K_{n,m+1}^+ = K_{n,m}^+ - \sum_{n < r \leq m+1} K_{n,r}^+ K_{r,m+1}^+ \langle \delta_r \gamma, \delta_{m+1} \gamma \rangle.$$

Rearranging the above yields

$$K_{n,m+1}^+ = \frac{K_{n,m}^+ - \sum_{n < r \leq m} K_{n,r}^+ K_{r,m+1}^+ \langle \delta_r \gamma, \delta_{m+1} \gamma \rangle}{1 + \|\delta_{m+1} \gamma\|^2}. \quad (3.57)$$

Remark 3.3.7. While we do not prove convergence of the implicit scheme, the experiments in the subsequent section utilise the implicit scheme rather than the explicit scheme as it appears more stable in practice. Indeed, from the form of Eq. (3.57) one may see that the first update $K_{n,n+1}^+$ remains bounded in $[-1, 1]$, while this is not true for the explicit scheme. We also remark that there are multiple variations on the implicit and explicit schemes to estimate the double integral in Eq. (3.48). For example, one could implement a quadrature style scheme where the integrand is averaged at both endpoints of the interval.

Remark 3.3.8. Since the limiting equation satisfied by $\langle \gamma, \cdot \rangle_{\mu_\infty}$ is derived as the large- N limit of randomised path developments, it is natural to consider a Monte-Carlo estimator for the solution to Theorem 3.2.13. That is, we average the time- T solution of multiple trajectories of the equation

$$dZ_\gamma^N(t) = \frac{i}{\sqrt{N}} \sum_{j=1}^d Z_\gamma^N(t) A_j^N d\gamma_t^j, \quad Z_\gamma^N(s, s) = I_N, \quad (3.58)$$

driven by independently sampled random matrices A_j . Let $\hat{\gamma}$ be a piecewise linear interpolation of γ on the partition $\mathcal{P} = \{0 = t_0 < t_1 < \dots < t_\ell = T\}$. For each $j = 1, \dots, M$ let $Z_{\hat{\gamma}}^{k,N}$ be independent solutions to Eq. (3.58), i.e. the driving matrices are

sampled independently for each k . We then define the Monte Carlo estimator

$$\text{MC}_{\hat{\gamma}}^{M,N} := \frac{1}{M} \sum_{k=1}^M \frac{1}{N} \text{tr}(Z_{\hat{\gamma}}^{k,N}(T)). \quad (3.59)$$

By [Theorem 3.2.13](#), we have the convergence

$$\lim_{N \rightarrow \infty} \lim_{M \rightarrow \infty} \text{MC}_{\hat{\gamma}}^{M,N} = K_{\hat{\gamma}}(0, T) \xrightarrow{|\mathcal{P}| \rightarrow 0} K_{\gamma}(0, T).$$

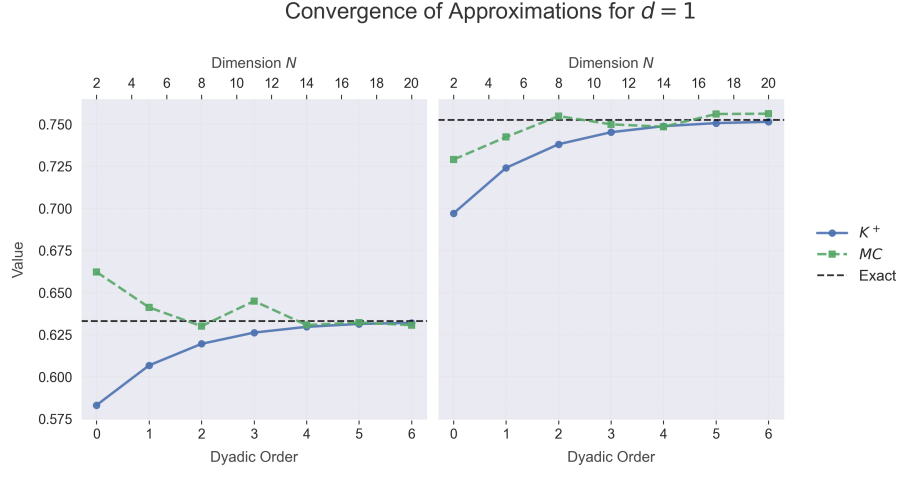
Since $\hat{\gamma}$ is a piecewise-linear interpolation of γ , the solutions $Z_{\hat{\gamma}}^{k,N}$ are explicit and given as the product of matrix exponentials involving only the driving matrices and the increments of $\hat{\gamma}$, see [\[LLN23, Section 2.2\]](#) for details.

3.3.3 Comparing Numerical Solvers

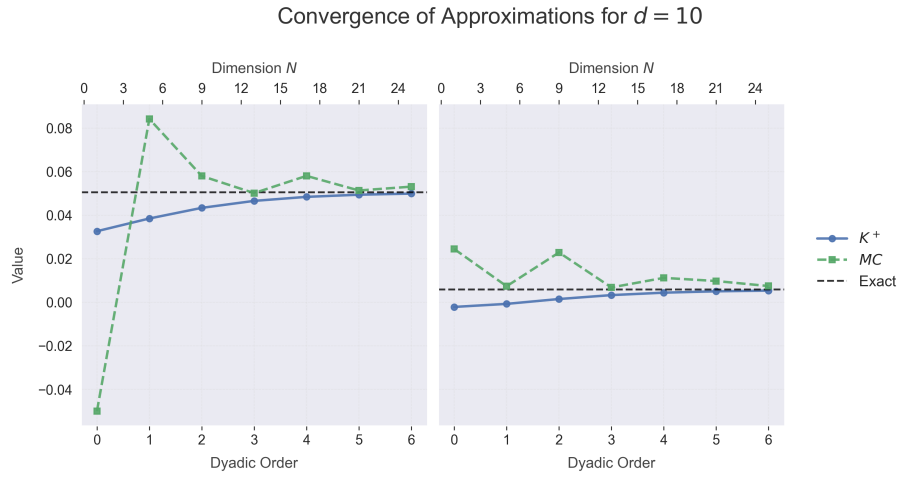
Given the stark contrast between the numerical approaches outlined in [Sections 3.3.1](#) and [3.3.2](#) and the Monte-Carlo approach of [Remark 3.3.8](#), it is natural to compare their relative benefits and limitations. Of particular interest is the scaling behaviour of the computational time in the length of a path, its dimension, and the hyperparameters of $\text{MC}_{\hat{\gamma}}^{M,N}(s, t)$. All experiments were performed on an NVIDIA GeForce RTX 2060 with Max-Q Design.

Remark 3.3.9. *This section is merely intended to compare the scaling of computational costs, and does not claim superiority of one method over the other.*

In the following, the true underlying path γ will be a piecewise-linear interpolation of a fractional Brownian motion with Hurst parameter $H = 0.75$ on some uniform partition $\mathcal{P}$ of $[0, 1]$. For the implicit numerical scheme from [Section 3.3.2](#), we examine the convergence of K^+ to $\langle \gamma_{0,T} \rangle_{\mu_G^\infty}$ for a sequence of dyadic refinements $\mathcal{P}(\lambda)$ of order $\lambda \in \mathbb{N}$ as $\lambda \rightarrow \infty$. For $\text{MC}_{\gamma}^{M,N}$, we fix $M = 50$ and vary the dimension N of the random matrices. All random matrices will be generated according to the Gaussian Unitary Ensemble (GUE).



(a) Convergence in $d = 1$.



(b) Convergence in $d = 10$.

Figure 3.1: Convergence for the implicit numerical scheme K^+ on the dyadic partition $\mathcal{P}(\lambda)$ and the Monte Carlo $\text{MC}_{\gamma}^{M,N}$ for $d = 1$ and $d = 10$. For each dimension, we generate two examples of a piecewise-linear interpolation of fractional Brownian motion with Hurst parameter $H = 0.75$ on a uniform partition π of $[0, 1]$ of size 15. The blue line corresponds to the lower x -axis (indexed by the dyadic order) and the red line corresponds to the upper x -axis (indexed by the dimension of the driving matrices in $\text{MC}_{\gamma}^{M,N}$). The number of Monte-Carlo samples is fixed at $M = 100$. The dotted black line is the true value $\langle \gamma_{0,T} \rangle_{\mu_G^{\infty}}$, computed in $d = 1$ by Remark 3.3.6 and by $\text{MC}^{125,450}$ in $d = 20$.

From Section 3.3.2, we know that the computational time for K^+ scales as $\mathcal{O}(\ell^2)$ in the size of $\mathcal{P}$ if computed on a GPU. In contrast, the randomised approach will scale as $\mathcal{O}(\ell)$. However, due to the increased computational costs involved (matrix exponentiation, matrix multiplication and Monte-Carlo sampling), it can be more costly to compute $\text{MC}_{\gamma}^{M,N}$. Furthermore, both K^+ and $\text{MC}_{\gamma}^{M,N}$ scale as $\mathcal{O}(d)$ in the dimension of the underlying state space of γ , although the constants involved are different. For K^+ , ℓ^2 dot products of increments in $\mathbb{R}^d$ need to be computed. And for $\text{MC}_{\gamma}^{M,N}$, there are $\ell \times d$ scalar-matrix multiplications $\mathbb{R} \times \mathbb{R}^{N \times N} \rightarrow \mathbb{R}$.

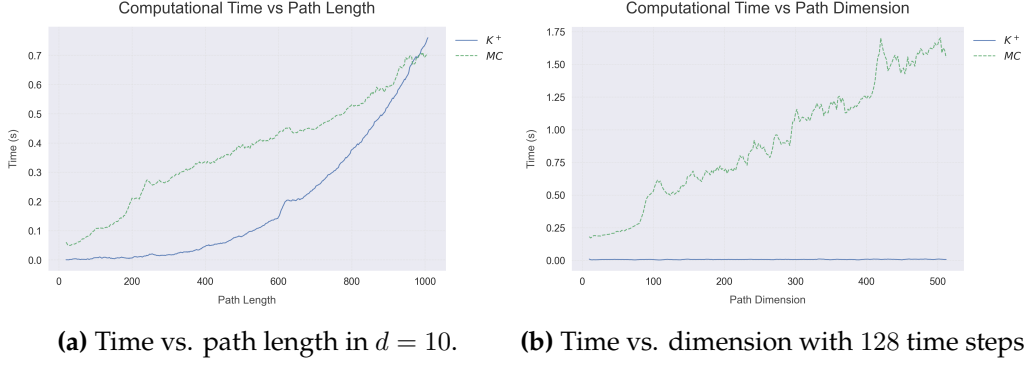


Figure 3.2: A comparison of the computational time scaling for the implicit numerical scheme and the randomised scheme when path length and path dimension vary. The hyper-parameters of $MC_{\gamma}^{M,N}$ are fixed at $N = 20$ and $M = 10$.

Finally, we note the impact of the choice of hyper-parameters M and N on the computational time for $MC_{\gamma}^{M,N}$. The computational time is linear in the number of Monte-Carlo samples (without additional parallelisation) but more than quadratic in the dimension N , even when leveraging efficient algorithms that exploit GPU parallelisation.

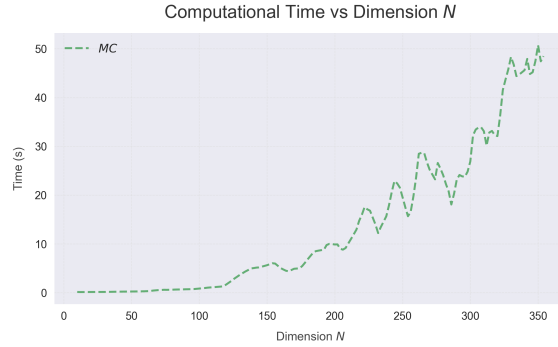


Figure 3.3: Computational time vs. N for a path with 128 time steps in $d = 10$. The remaining hyper-parameter of $MC_{\gamma}^{M,\cdot}$ is fixed at $M = 2$.

3.4 Pauli string path developments

Recall again the differential equation underlying [Eq. \(3.45\)](#)

$$dZ_t^{\gamma} = i \sum_{j=1}^d Z_t^{\gamma} A_j d\gamma_t^j, \quad Z_0^{\gamma} = I_N,$$

and for convenience suppose that each $A_j \in \mathbb{R}^{N \times N}$ is a random matrix with i.i.d. $\mathcal{N}(0, \frac{1}{N})$ entries and N is large. Supposing that $\gamma = \gamma_1 \star \dots \star \gamma_m$ is a piecewise linear path valued in $\mathbb{R}^d$ with linear increments γ_i for $i = 1, \dots, m$, then Z_T^{γ} is given

exactly by the matrix product

$$Z_T^\gamma = \exp\left(i \sum_{j=1}^d A_j \gamma_1^j\right) \cdots \exp\left(i \sum_{j=1}^d A_j \gamma_m^j\right). \quad (3.60)$$

Clearly, this exact solution becomes costly to compute when N is large. An approach to circumvent this issue might be to consider sparse approximations of the A_j . However, even for sparse A_j , each matrix exponential is typically dense and so the products are still an obstacle. However if we restrict ourselves to the case $N = 2^n$, then there exists a sparse - but still high-rank - random matrix ensemble, whose eigenvalue distribution converges to the free semicircular random variables as $n \rightarrow \infty$. Moreover, this representation allows one to embed [Eq. \(3.60\)](#) efficiently in a quantum circuit.

The random matrix ensemble we utilise is called the random Pauli string ensemble, which is a known example of a non-trivial random Hamiltonian that is “quantumly easy” [[Che+24b](#)].

Definition 3.4.1 (Pauli strings). *A set of Hermitian operators on $\mathbb{C}^2$ given by*

$$\sigma_I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \sigma_X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (3.61)$$

A (real) linear basis for $\mathfrak{u}_N$ is then given by tensor products of the Pauli matrices

$$\sigma_{\mathbf{w}} = \sigma_{w_1} \otimes \cdots \otimes \sigma_{w_n}, \quad (3.62)$$

where $\mathbf{w} = w_1 \dots w_n$ is a word with n letters chosen from $\{I, X, Y, Z\}$ and $\otimes$ denotes the tensor product. The operator $\sigma_{\mathbf{w}}$ is referred to as a Pauli string. We write the set of words on Pauli matrices as $\mathcal{P}$.

Definition 3.4.2 (Pauli matrices). *We denote by $\mathcal{P} = \{\sigma_I, \sigma_X, \sigma_Y, \sigma_Z\}$ the set of Pauli matrices, introduced in [Definition 3.4.1](#) and write $\sigma \sim \mathcal{P}$ for a random variable σ uniformly distributed over $\mathcal{P}$.*

We are now ready to introduce the random Pauli string ensemble.

Definition 3.4.3. *Let $N = 2^n$. We say that a matrix A belongs to the random Pauli ensemble with m Pauli strings if*

$$A = \frac{1}{\sqrt{m}} \sum_{i=1}^m r^i \sigma^i,$$

where $r_i \stackrel{i.i.d.}{\sim} \text{Rad}(1/2)$ and $\sigma^i = \sigma(1)^i \otimes \cdots \otimes \sigma(n)^i$ with $\sigma(j)^i \stackrel{i.i.d.}{\sim} \mathcal{P}$.

The remainder of this section is dedicated to proving the following result.

Theorem 3.4.4. *For each n , suppose that $A_1^n, \dots, A_d^n$ be i.i.d. matrices from the random Pauli string ensemble of dimension $N = 2^n$ with m_n strings each. Let Z_γ^n be the solution to*

$$dZ_\gamma^n(s, t) = i \sum_{j=1}^d Z_\gamma^n(s, t) A_j^n d\gamma_t^j, \quad Z_\gamma^n(s, s) = I.$$

Then there exists $C > 0$ depending only on the length of γ and the dimension of the underlying space V for which

$$\mathbb{E} \left[\left(\text{tr}(Z_\gamma^n(s, t)) - \langle \gamma_{s,t} \rangle_{\mu_G^\infty} \right)^2 \right] < C(m_n^{-1} + 2^{-2n}). \quad (3.63)$$

The core of the argument is a (quantitative) proof that for $A_1, \dots, A_d$ from the random Pauli string ensemble, their joint eigenvalue distribution converges to the law of d -free semicircular random variables as $n \rightarrow \infty$. We note that this is a generalisation of the result found in [Che+24b] for $d = 1$, although the method of proof is different.

We start by recalling and proving some basic properties concerning the Pauli ensemble. Firstly, products of Pauli matrices follow the following multiplication table.

$\times$	σ_I	σ_X	σ_Y	σ_Z
σ_I	σ_I	σ_X	σ_Y	σ_Z
σ_X	σ_X	σ_I	$i\sigma_Z$	$-i\sigma_Y$
σ_Y	σ_Y	$-i\sigma_Z$	σ_I	$i\sigma_X$
σ_Z	σ_Z	$i\sigma_Y$	$-i\sigma_X$	σ_I

We may then consider the group $G = \langle \mathcal{P} \rangle$ generated by Pauli matrices under matrix multiplication. By the above multiplication rules, it is clear that every $g \in G$ may be uniquely written as

$$g = \phi(g)m(g), \quad \phi(g) \in \{1, i, -1, -i\} \quad \text{and} \quad m(g) \in \mathcal{P}. \quad (3.64)$$

We refer to $\phi(g)$ as the phase of g . Note the rules of commutation of two elements of $g, h \in G$: $gh = hg$ whenever

- 1) $m(g) = m(h)$; or
- 2) $m(g) = I$ or $m(h) = I$.

The matrices g and h anti-commute ($gh = -hg$) if $m(g) \neq m(h)$ and neither is the identity. Moreover, since $m(g) \in \mathcal{P}$ for any $g \in G$, it follows that $|\text{tr}(g)| \leq 1$ for all $g \in G$.

Lemma 3.4.5. *Suppose that X and Y are independent random variables with values in G such that $m(X)$ and $m(Y)$ are uniformly distributed on $\mathcal{P}$, then $m(XY) \sim \mathcal{P}$.*

Proof. This is clear from the multiplication table and the fact that $m(XY) = m(X)m(Y)$. $\square$

By induction on the above lemma, we obtain the following.

Corollary 3.4.6. *Let $X_1, \dots, X_n \stackrel{i.i.d.}{\sim} \mathcal{P}$, then $m(X_1 \cdots X_n) \sim \mathcal{P}$.*

We also require the following result.

Lemma 3.4.7. *Let $g \in G$ and $X \sim \mathcal{P}$, then $\mathbb{E}[XgX] = \mathbb{1}_{\{m(g)=\sigma_I\}}\phi(g)I$.*

Proof. If $m(g) = I$, then $\mathbb{E}[XgX] = \phi(g)\mathbb{E}[XX] = \phi(g)I$. If $m(g) \neq I$, then

$$XgX = \begin{cases} g, & X = I \text{ or } X = m(g); \\ -g, & X \neq I \text{ and } X \neq m(g). \end{cases}$$

Since both cases have probability $\frac{1}{2}$, $\mathbb{E}[XgX] = 0$. $\square$

Combining the above allows us to bound the expected trace of products of random Pauli matrices.

Proposition 3.4.8. *Let $X_1, \dots, X_p \stackrel{i.i.d.}{\sim} \mathcal{P}$ and let $\mathbf{w}$ be a word in p letters of length $2p$ such that each letter $\{1, \dots, p\}$ appears exactly twice. If pairing equal letters results in a non-crossing pair partition, then $\mathbb{E}[\text{tr}(X_{w_1} \cdots X_{w_{2p}})] = 1$, otherwise*

$$\left| \mathbb{E}[\text{tr}(X_{w_1} \cdots X_{w_{2p}})] \right| \leq \frac{1}{4}. \quad (3.65)$$

In the following, we will write $[m]$ for an alphabet in the letters $\{1, \dots, m\}$.

Proof. If $\mathbf{w}$ defines a non-crossing pair partition, then any consecutive pairs of letters $X_j X_j$ may be collapsed to the identity. By the non-crossing property there must be at least one such pair, which results in a shorter word of $2p - 2$ letters, defining a non-crossing pair partition in $p - 1$ -distinct letters. By induction the entire word may be collapsed to the identity, from which the result follows.

Assume now that $\mathbf{w}$ does not define a non-crossing pair partition. First note that if there are any consecutive letters in $\mathbf{w}$, i.e. the appearance of $X_j X_j$ for some $j \in \{1, \dots, p\}$, then the pair may be collapsed to the identity. If this collapse results in a new consecutive pair appearing, this may also be collapsed. Since the pairing of equal letters does not result in a non-crossing pair partition, the entire product cannot be collapsed to the identity. As such, we may assume without loss of generality that $\mathbf{w}$ contains no consecutive pairs. We claim that there must be a subword of $\mathbf{w}$ of the form $X_j X_l X_j$ for some non-empty word l that contains no repeat letters. This can be checked by induction on the length of the word with the conclusion that if no such subword exists, then there must be a consecutive pair. We may then write $\mathbf{w} = \mathbf{k} j l j \mathbf{m}$, where $\mathbf{k}, \mathbf{m}$ do not contain j . Hence

$$\begin{aligned} \mathbb{E}[\text{tr}(X_{\mathbf{w}})] &= \text{tr}(\mathbb{E}[X_{\mathbf{w}}]) \\ &= \text{tr}(\mathbb{E}[X_{\mathbf{k}} X_j X_l X_j X_{\mathbf{m}}]) \\ &= \text{tr}(\mathbb{E}[X_{\mathbf{k}} \mathbb{E}[X_j X_l X_j \mid X_i : i \in [p] \setminus j] X_{\mathbf{m}}]) \\ &= \text{tr}(\mathbb{E}[X_{\mathbf{k}} \mathbb{E}[X_j X_l X_j \mid X_l] X_{\mathbf{m}}]) \\ &= \text{tr}(\mathbb{E}[X_{\mathbf{k}} \phi(X_l) \mathbb{1}_{\{m(X_l)=l\}} X_{\mathbf{m}}]) \\ &= \mathbb{E}[\phi(X_l) \mathbb{1}_{\{m(X_l)=l\}} \text{tr}(X_{\mathbf{k}} X_{\mathbf{m}})]. \end{aligned}$$

The fifth equality used [Lemma 3.4.7](#). Therefore

$$\left| \mathbb{E}[\text{tr}(X_{\mathbf{w}})] \right| \leq \mathbb{E} \left[\mathbb{1}_{\{m(X_l)=l\}} |\phi(X_l)| |\text{tr}(X_{\mathbf{k}} X_{\mathbf{m}})| \right] \leq \mathbb{P}(\{m(X_l) = l\}) = \frac{1}{4},$$

since $|\phi(g)| = 1$ and $|\text{tr}(g)| \leq 1$ for all $g \in G$, with the final equality using [Corollary 3.4.6](#). $\square$

Lemma 3.4.9. *Let $X_1, \dots, X_p \stackrel{i.i.d.}{\sim} \mathcal{P}$ and let $\mathbf{u}, \mathbf{v}$ be words in the letters $[p]$ such that*

- 1) $|\mathbf{uv}| = 2p$;
- 2) *Each letter appears exactly twice in $\mathbf{uv}$;*
- 3) *There is at least one letter that appears in both $\mathbf{u}$ and $\mathbf{v}$.*

Then

$$\mathbb{E}[\text{tr}(X_{\mathbf{u}}) \text{tr}(X_{\mathbf{v}})] \leq \frac{1}{4}.$$

Proof. For any word $\mathbf{w}$ we have that $\text{tr}(X_{\mathbf{w}}) = \phi(X_{\mathbf{w}}) \mathbb{1}_{\{m(X_{\mathbf{w}})=\sigma_I\}}$. We know there exists a letter $j \in [p]$ that appears once in both $\mathbf{u}$ and $\mathbf{v}$, therefore we can decompose the two words as $\mathbf{u} = \mathbf{u}_1 j \mathbf{u}_2$ and $\mathbf{v} = \mathbf{v}_1 j \mathbf{v}_2$, where none of $\mathbf{u}_1, \mathbf{u}_2, \mathbf{v}_1, \mathbf{v}_2$ contain

the letter j . Thus, we have

$$|\text{tr}(X_{\mathbf{u}_1} X_j X_{\mathbf{u}_2}) \text{tr}(X_{\mathbf{v}_1} X_j X_{\mathbf{v}_2})| = \mathbb{1}_{\{m(X_{\mathbf{u}_1} X_j X_{\mathbf{u}_2}) = \sigma_I\}} \mathbb{1}_{\{m(X_{\mathbf{v}_1} X_j X_{\mathbf{v}_2}) = \sigma_I\}}.$$

The two indicator functions are simultaneously equal to 1 if and only if

$$m(X_j) = \alpha^{-1} \quad \text{and} \quad m(X_j) = \beta^{-1} \quad (3.66)$$

where $\alpha = m(X_{\mathbf{u}_1})m(X_{\mathbf{u}_2})$ and $\beta = m(X_{\mathbf{v}_1})m(X_{\mathbf{v}_2})$.

Now, consider the σ -algebra $\mathcal{F} = \sigma(X_i : i \neq j)$ generated by all Paulis X_i except X_j . Conditioning on $\mathcal{F}$, the random matrix X_j is uniformly distributed on $\mathcal{P}$. Hence,

$$\mathbb{E} \left[|\text{tr}(X_{\mathbf{u}_1} X_j X_{\mathbf{u}_2}) \text{tr}(X_{\mathbf{v}_1} X_j X_{\mathbf{v}_2})| \mid \mathcal{F} \right] = \mathbb{P}((3.66) \text{ holds} \mid \mathcal{F}) = \begin{cases} \frac{1}{4}, & \text{if } \alpha = \beta, \\ 0, & \text{if } \alpha \neq \beta. \end{cases}$$

In either case, the conditional expectation does not exceed $\frac{1}{4}$. Thus, taking expectation and using the fact that $|\phi(g)| = 1$ for all $g \in G$ yields the result. $\square$

Recall that the tensor product $\sigma_{\mathbf{w}} = \sigma_{w_1} \otimes \cdots \otimes \sigma_{w_n}$, where $\sigma_{w_i} \in \mathcal{P}$, is a $2^n \times 2^n$ matrix such that

$$\text{tr}(\sigma_{\mathbf{w}}) = \text{tr}(\sigma_{w_1}) \cdots \text{tr}(\sigma_{w_n}). \quad (3.67)$$

Definition 3.4.10. For any alphabet size $m \geq 1$ (denoted by $[m]$) and integer $p \geq 1$ define

- 1) $W(m, p)$: number of words of length $2p$ where distinct letters appear an even number of times;
- 2) $N(m, p)$: number of words of length $2p$ where distinct letters appear exactly twice.

Lemma 3.4.11. For any $m \geq 1$ and $p \geq 1$,

$$W(m, p) = \frac{1}{2^m} \sum_{k=0}^m \binom{m}{k} (m-2k)^{2p} = \sum_{\substack{k_1 + \cdots + k_m = p \\ k_i \geq 0}} \frac{(2p)!}{(2k_1)! \cdots (2k_m)!}. \quad (3.68)$$

Proof. For a given word $\mathbf{w} = w_1 \cdots w_{2p}$ define the vector $c(\mathbf{w}) = (c_1(\mathbf{w}), \dots, c_m(\mathbf{w}))$, where $c_i(\mathbf{w})$ denotes the number of times letter i appears in $\mathbf{w}$. Write

$$\delta(c_i(\mathbf{w})) = \frac{1 + (-1)^{c_i(\mathbf{w})}}{2},$$

for the indicator that the count of letter i is even. Then $\mathbf{w}$ contributes to $W(m, p)$ if and only if

$$1 = \prod_{i=1}^m \frac{1 + (-1)^{c_i(\mathbf{w})}}{2} = \frac{1}{2^m} \sum_{\epsilon \in \{0,1\}^m} (-1)^{\sum_{i=1}^m \epsilon_i c_i(\mathbf{w})}.$$

Hence, by summing over all words in the alphabet $[m]$ of length $2p$, we have

$$W(m, p) = \sum_{\mathbf{w}} \frac{1}{2^m} \sum_{\epsilon \in \{0,1\}^m} (-1)^{\sum_{i=1}^m \epsilon_i c_i(\mathbf{w})} = \frac{1}{2^m} \sum_{\epsilon \in \{0,1\}^m} \sum_{\mathbf{w}} (-1)^{\sum_{i=1}^m \epsilon_i c_i(\mathbf{w})}.$$

For fixed ϵ , we can compute the inner sum as

$$\sum_{\mathbf{w}} (-1)^{\sum_{i=1}^m \epsilon_i c_i(\mathbf{w})} = \sum_{\mathbf{w}} \prod_{j=1}^{2p} (-1)^{\epsilon_{w_j}} = \left(\sum_{i=1}^m (-1)^{\epsilon_i} \right)^{2p},$$

where $\epsilon_{w_j} = \epsilon_i$ whenever $w_j = i$. Noticing that

$$\sum_{i=1}^m (-1)^{\epsilon_i} = m - 2t, \quad \text{where } t = \#\{i : \epsilon_i = 1\},$$

we can combine the above to get

$$W(m, p) = \frac{1}{2^m} \sum_{\epsilon \in \{0,1\}^m} \left(\sum_{i=1}^m (-1)^{\epsilon_i} \right)^{2p} = \frac{1}{2^m} \sum_{t=0}^m \binom{m}{t} (m - 2t)^{2p}.$$

For the second equality, from the conditions defining $W(m, p)$ we have $c_i(\mathbf{w}) = 2k_i(\mathbf{w})$, for some $k_i(\mathbf{w}) \geq 0$, and $\sum_{i=1}^m k_i(\mathbf{w}) = p$. Using the multinomial theorem we know that the letters of $\mathbf{w}$ can be arranged in $\frac{(2p)!}{(2k_1(\mathbf{w}))! \cdots (2k_m(\mathbf{w}))!}$ different ways. Putting these two observations together yields the result. $\square$

Lemma 3.4.12. For any $m \geq 1$ and $p \geq 1$,

$$N(m, p) = \binom{m}{p} \frac{(2p)!}{2^p} = \sum_{\substack{k_1 + \cdots + k_m = p \\ k_i \in \{0,1\}}} \frac{(2p)!}{(2k_1)! \cdots (2k_m)!}. \quad (3.69)$$

Proof. We have to choose which p letters out of the m possible letters will appear and then arrange them in a word of length $2p$ so that they appear twice each. There are $\binom{m}{p}$ ways of choosing p letters from the alphabet $[m]$. For the chosen p letters, we must place two copies of each letter among the $2p$ positions. For any set of the form $\{w_1, w_1, w_2, w_2, \dots\}$, the number of distinct orderings is the multinomial count $\frac{(2p)!}{2^p}$, which yields the first equality. The second equality follows from the

observation that each multi-index $(k_1, \dots, k_m)$ with $k_i \in \{0, 1\}$ and $\sum_i k_i = p$ has exactly p of the k_i 's equal to 1 and the remaining $m - p$ equal to 0. Hence

$$\frac{(2p)!}{(2k_1)!(2k_2)! \cdots (2k_m)!} = \frac{(2p)!}{\underbrace{2! \cdots 2!}_{p \text{ factors}} \underbrace{0! \cdots 0!}_{m-p \text{ factors}}} = \frac{(2p)!}{(2!)^p} = \frac{(2p)!}{2^p}.$$

There are $\binom{m}{p}$ ways to choose which p indices satisfy $k_i = 1$, so the total sum is $\binom{m}{p} \frac{(2p)!}{2^p}$ as claimed. $\square$

Lemma 3.4.13. *For any $m \geq 1$ and $p \geq 1$, the inequality $W(m, p) \leq 2^p m^p p!$ holds.*

Proof. For any integer $k \geq 0$, $(2k)! = 2^k k! (1 \cdot 3 \cdots (2k-1)) \geq 2^k k!$. Hence from (3.68) we obtain

$$W(m, p) \leq \frac{(2p)!}{2^p} \sum_{\substack{k_1 + \cdots + k_m = p \\ k_i \geq 0}} \frac{1}{k_1! \cdots k_m!} = \frac{(2p)!}{2^p p!} m^p,$$

where we used the fact that

$$\sum_{\substack{k_1 + \cdots + k_m = p \\ k_i \geq 0}} \frac{1}{k_1! \cdots k_m!} = \frac{m^p}{p!}.$$

Finally the lemma follows since $(2p)! = 2^p p! (1 \cdot 3 \cdots (2p-1)) \leq 2^p p! (2 \cdot 4 \cdots 2p) = (2^p p!)^2$. $\square$

Lemma 3.4.14. *For any $m \geq 1$ and $p \geq 1$ we have*

$$W(m, p) - N(m, p) \leq (4e)^p m^{p-1} p! \quad (3.70)$$

Proof. Using (3.68) and (3.69) we obtain

$$W(m, p) - N(m, p) = \sum_{\substack{k_1 + \cdots + k_m = p \\ \text{at least one } k_i \geq 2}} \frac{(2p)!}{(2k_1)! \cdots (2k_m)!}.$$

Using similar arguments as in the proof of [Lemma 3.4.13](#),

$$\begin{aligned}
 W(m, p) - N(m, p) &\leq \frac{(2p)!}{2^p} \sum_{\substack{k_1 + \dots + k_m = p \\ \text{at least one } k_i \geq 2}} \frac{1}{k_1! \dots k_m!} \\
 &= \frac{(2p)!}{2^p} \left(\sum_{\substack{k_1 + \dots + k_m = p \\ k_i \geq 0}} \frac{1}{k_1! \dots k_m!} - \sum_{\substack{k_1 + \dots + k_m = p \\ k_i \in \{0, 1\}}} \frac{1}{k_1! \dots k_m!} \right) \\
 &= \frac{(2p)!}{2^p} \left(\frac{m^p}{p!} - \binom{m}{p} \right) \\
 &= \frac{(2p)!}{2^p p!} \left(m^p - \frac{m!}{(m-p)!} \right) \\
 &= \frac{(2p)!}{2^p p!} m^{p-1} \left(m - m \left(1 - \frac{1}{m} \right) \left(1 - \frac{2}{m} \right) \dots \left(1 - \frac{p-1}{m} \right) \right).
 \end{aligned}$$

Now, for any $j = 1, \dots, p-1$ then $0 < \frac{j}{m} < 1$ and $(1 - \alpha) \geq \exp(-\frac{\alpha}{1-\alpha})$ for any $\alpha \in (0, 1)$. Hence,

$$\begin{aligned}
 W(m, p) - N(m, p) &\leq \frac{(2p)!}{2^p p!} m^{p-1} \left(m - m \prod_{j=1}^{p-1} \exp \left(-\frac{j/m}{1 - j/m} \right) \right) \\
 &= \frac{(2p)!}{2^p p!} m^{p-1} \left(m - m \exp \left(-\frac{1}{m} \sum_{j=1}^{p-1} \frac{j}{1 - j/m} \right) \right) \\
 &\leq \frac{(2p)!}{2^p p!} m^{p-1} \left(m - m \exp \left(-\frac{p(p-1)}{2m} \right) \right) \\
 &\leq \frac{(2p)!}{2^p p!} \frac{p(p-1)}{2} m^{p-1} \leq (2p)^p \frac{p^2}{2} m^{p-1}
 \end{aligned}$$

where in the penultimate equality we used the simple fact that $1 - e^{-x} \leq x$ for any $x \geq 0$ and set $x = \frac{p(p-1)}{2m}$. The classical Stirling bound $p! \geq (p/e)^p$ easily yields $(2p)^p \frac{p^2}{2} \leq (4e)^p p!$, from which the result follows. $\square$

We are now ready to prove that the random Pauli string ensemble converges to the law of d -free semicircular random variables.

Proposition 3.4.15. *For each n , suppose that $A_1^n, \dots, A_d^n$ be i.i.d. matrices from the random Pauli string ensemble of dimension $N = 2^n$ with m_n strings each. If $\lim_{n \rightarrow \infty} m_n = \infty$, then*

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[\text{tr}(A_{\mathbf{w}}^n) \right] = \tau(X_{\mathbf{w}})$$

for all $\mathbf{w} \in \mathcal{W}_d$, where τ is the law of d -free semicircular random variables.

Proof. First note that that

$$\mathbb{E}[\text{tr}(A_w^n)] = m_n^{-|w|/2} \sum_{\mathbf{u}} \mathbb{E}[r_w^{\mathbf{u}}] \mathbb{E}[\text{tr}(\sigma_w^{\mathbf{u}})],$$

where the sum runs over words $\mathbf{u}$ in m_n -letters of length $|w|$. Since $\mathbb{E}[r_w^{\mathbf{u}}]$ disappears unless each letter in $\mathbf{u}$ appears an even number of times, we need only to consider the case where $|w| = 2p$ for $p \in \mathbb{N}$. We will still require for each letter in $\mathbf{u}$ to appear at least twice, so that the total number of distinct letters in $\mathbf{u}$ is at most p . We will write $\text{wt}(\mathbf{u})$ for the number of distinct letters in $\mathbf{u}$. We say that two words $\mathbf{u}$ and $\mathbf{v}$ are equivalent if there is a bijection on $[m_n]$ that takes $\mathbf{u}$ to $\mathbf{v}$. We have the following facts.

- 1) The number of words in an equivalence class of a given m -word with weight t is exactly

$$m_n(m_n - 1) \cdots (m_n - t + 1) \leq m_n^t;$$

- 2) The number of equivalence classes of words with weight t is bounded by t^{2p} .

By (3.67) and the property of traces of the group G , $|\mathbb{E}[\text{tr}(\sigma_w^{\mathbf{u}})]| \leq 1$. Combining these facts yields

$$\sum_{\mathbf{u}: \text{wt}(\mathbf{u}) \leq p-1} \left| \mathbb{E}[r_w^{\mathbf{u}}] \mathbb{E}[\text{tr}(\sigma_w^{\mathbf{u}})] \right| \leq \sum_{t=1}^{p-1} m_n^t t^{2p} \leq m_n^{p-1} p^{2p+1}. \quad (3.71)$$

We turn our attention to words of weight p . First, we see that

$$\mathbb{E}[\text{tr}(\sigma_w^{\mathbf{u}})] = \mathbb{E}[\text{tr}(\sigma(1)_w^{\mathbf{u}}) \cdots \text{tr}(\sigma(n)_w^{\mathbf{u}})] = \mathbb{E}[\text{tr}(\sigma(1)_w^{\mathbf{u}})]^n,$$

by the i.i.d. construction of the Pauli strings. Consider first the case where pairing equal letters in $\mathbf{u}$ defines a crossing pair partition, now by Proposition 3.4.8 we know that

$$\left| \mathbb{E}[\text{tr}(\sigma(1)_w^{\mathbf{u}})]^n \right| \leq 2^{-2n}.$$

The number of words of length $2p$ with weight p where distinct letters appear exactly twice may be bounded by $m_n^p \frac{p^{2p}}{p!}$. And so the total contribution of these terms is bounded by $m_n^p \frac{p^{2p}}{p!} 2^{-2n}$. Finally, consider those words $\mathbf{u}$ where pairing equal letters results in a non-crossing pair partition. Then

$$\mathbb{E}[r_w^{\mathbf{u}}] \mathbb{E}[\text{tr}(\sigma_w^{\mathbf{u}})] = \prod_{(a,b) \in \pi_{\mathbf{u}}} \mathbb{1}_{w_a=w_b},$$

where $\pi_{\mathbf{u}}$ is the pair partition of the letters $\{1, \dots, 2p\}$ defined by the pairing of the letters in $\mathbf{u}$. It is easy to see that two valid words $\mathbf{u}$ and $\mathbf{v}$ are equivalent if and only if the non-crossing pair partitions they each generate are equal. It remains then to count the number of non-crossing pair partitions of $\{1, \dots, 2p\}$ for which

$$\prod_{(a,b) \in \pi_{\mathbf{u}}} \mathbb{1}_{w_a=w_b} = 1.$$

Recalling the discussion in [Section 3.0.4](#), this is one of the defining properties of $\tau(X_{\mathbf{w}})$. The number of words in each equivalence class is $m_n(m_n - 1) \cdots (m_n - p + 1)$. It follows that

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{E} \left[\text{tr}(A_{\mathbf{w}}^n) \right] &= \lim_{n \rightarrow \infty} \left\{ \tau(X_{\mathbf{w}}) \frac{m_n(m_n - 1) \cdots (m_n - p + 1)}{m_n^p} + \mathcal{O}(2^{-2n}) + \mathcal{O}(m_n^{-1}) \right\} \\ &= \tau(X_{\mathbf{w}}). \end{aligned}$$

Here, the first term is the contribution of the non-crossing pair partitions, the second is the crossing pair partitions and the third is the contribution of terms with weight less than p . $\square$

We now derive a quantitative version of the previous result.

Proposition 3.4.16. *For each n , suppose that $A_1^n, \dots, A_d^n$ are i.i.d. matrices from the random Pauli string ensemble of dimension $N = 2^n$ with m_n strings each. Then*

$$\left| \mathbb{E} \left[\text{tr}(A_{\mathbf{w}}^n) \right] - \tau(X_{\mathbf{w}}) \right| \leq (4e)^p p! (m_n^{-1} + 2^{-2n}),$$

where τ is the law of d -free semicircular random variables and $\mathbf{w} \in \mathcal{W}_d$ has length $2p$ for some $p \in \mathbb{N}$.

Proof. As in the previous proposition we note that

$$\mathbb{E} \left[\text{tr}(A_{\mathbf{w}}^n) \right] = m_n^{-|\mathbf{w}|/2} \sum_{\mathbf{u}} \mathbb{E}[r_{\mathbf{w}}^{\mathbf{u}}] \mathbb{E}[\text{tr}(\sigma_{\mathbf{w}}^{\mathbf{u}})],$$

where the sum runs over words $\mathbf{u}$ in m -letters of length $|\mathbf{w}|$. Again, it suffices to consider words $\mathbf{w}$ of even length $|\mathbf{w}| = 2p$, with $p \geq 1$. We write $\text{wt}(\mathbf{u})$ for the number of distinct letters in $\mathbf{u}$. As in the previous proposition, we split this sum into three parts.

(i) (Words of weight $t < p$) It is easy to see that

$$W(m, p) - N(m, p) = \#\{\mathbf{u} : \text{wt}(\mathbf{u}) < p\}.$$

Thus, using [Lemma 3.4.14](#) we obtain directly

$$m_n^{-p} \sum_{\mathbf{u}: \text{wt}(\mathbf{u}) < p} \left| \mathbb{E}[r_{\mathbf{w}}^{\mathbf{u}}] \mathbb{E}[\text{tr}(\sigma_{\mathbf{w}}^{\mathbf{u}})] \right| = m_n^{-p} (W(m_n, p) - N(m_n, p)) \leq (4e)^p p! \frac{1}{m_n} \quad (3.72)$$

(ii) (Words of weight $t = p$) As before, we need to distinguish two sub-cases:

Case 1: where pairing equal letters in $\mathbf{u}$ defines a crossing pair partition. By [Proposition 3.4.8](#)

$$\mathbb{E}[\text{tr}(\sigma_{\mathbf{w}}^{\mathbf{u}})] = \mathbb{E}[\text{tr}(\sigma(1)_{\mathbf{w}}^{\mathbf{u}})]^n \leq 2^{-2n}.$$

The count of such words is at most $N(m_n, p) = \binom{m_n}{p} \frac{(2p)!}{2^p} \leq m_n^p \frac{(2p)!}{2^p p!}$. Hence,

$$m_n^{-p} \sum_{\substack{\mathbf{u}: \text{wt}(\mathbf{u}) = p \\ \text{Case 1}}} \left| \mathbb{E}[r_{\mathbf{w}}^{\mathbf{u}}] \mathbb{E}[\text{tr}(\sigma_{\mathbf{w}}^{\mathbf{u}})] \right| \leq \frac{(2p)!}{2^p p!} 2^{-2n} \leq (4e)^p p! 2^{-2n}, \quad (3.73)$$

where we used the Stirling bound $(2p)!/(2^p p!) \leq (4e)^p p!$.

Case 2: where pairing equal letters in $\mathbf{u}$ results in a non-crossing pair partition. Using the same counting arguments as in the proof of the previous proposition, we have

$$m_n^{-p} \sum_{\substack{\mathbf{u}: \text{wt}(\mathbf{u}) = p \\ \text{Case 2}}} \mathbb{E}[r_{\mathbf{w}}^{\mathbf{u}}] \mathbb{E}[\text{tr}(\sigma_{\mathbf{w}}^{\mathbf{u}})] = \tau(X_{\mathbf{w}}) \frac{m_n(m_n - 1) \cdots (m_n - p + 1)}{m_n^p}.$$

Using the fact that

$$\left| \frac{m(m-1) \cdots (m-p+1)}{m^p} - 1 \right| \leq \frac{p^2}{2m}, \quad (3.74)$$

yields

$$\left| m_n^{-p} \sum_{\substack{\mathbf{u}: \text{wt}(\mathbf{u}) = p \\ \text{Case 2}}} \mathbb{E}[r_{\mathbf{w}}^{\mathbf{u}}] \mathbb{E}[\text{tr}(\sigma_{\mathbf{w}}^{\mathbf{u}})] - \tau(X_{\mathbf{w}}) \right| \leq \frac{p^2}{2m_n} |\tau(X_{\mathbf{w}})| \leq \frac{(4e)^p p!}{m_n}, \quad (3.75)$$

where we used the fact that $|\tau(X_{\mathbf{w}})| \leq 4^p$.

Combining the bounds (3.72), (3.73) and (3.75) yields

$$\left| \mathbb{E} \left[\text{tr}(A_w^n) \right] - \tau(X_w) \right| \leq (4e)^p p! (2m_n^{-1} + 2^{-2n}).$$

□

Proposition 3.4.17. *For each n , suppose that $A_1^n, \dots, A_d^n$ are i.i.d. matrices from the random Pauli string ensemble of dimension $N = 2^n$ with m_n strings each. Then*

$$\left| \mathbb{E} \left[\text{tr}(A_u^n) \text{tr}(A_v^n) \right] - \tau(X_u) \tau(X_v) \right| \leq (4e)^p p! (2m_n^{-1} + 2^{-2n})$$

for all words $u, v \in \mathcal{W}_d$ such that $|u| + |v| = 2p$ for some $p \in \mathbb{N}$ (with the left-hand-side being zero otherwise).

Proof. We first write

$$\mathbb{E} \left[\text{tr}(A_u^n) \text{tr}(A_v^n) \right] = m_n^{-(|u|+|v|)/2} \sum_{\mathbf{a}, \mathbf{b}} \mathbb{E}[r_u^{\mathbf{a}} r_v^{\mathbf{b}}] \mathbb{E}[\text{tr}(\sigma_u^{\mathbf{a}}) \text{tr}(\sigma_v^{\mathbf{b}})],$$

where the sum runs over words $\mathbf{a}$ and $\mathbf{b}$ in the letters $[m]$ of length $|u|$ and $|v|$ respectively. Due to the Rademacher random variables, we need only consider the case where $|u| + |v| = 2p$ for some $p \in \mathbb{N}$. Moreover, we require each letter in the word $\mathbf{ab}$ to appear at least twice for the expectation to be non-zero. As before, we then split the sum into three parts.

(i) (Words of weight $t < p$) Using Proposition 3.4.8 we directly compute

$$\begin{aligned} m_n^{-p} \sum_{\mathbf{a}, \mathbf{b}: \text{wt}(\mathbf{ab}) < p} \left| \mathbb{E}[r_u^{\mathbf{a}} r_v^{\mathbf{b}}] \mathbb{E}[\text{tr}(\sigma_u^{\mathbf{a}}) \text{tr}(\sigma_v^{\mathbf{b}})] \right| &= m_n^{-p} (W(m_n, p) - N(m_n, p)) \\ &\leq (4e)^p p! \frac{1}{m_n}. \end{aligned} \tag{3.76}$$

(ii) (Words of weight $t = p$) We now consider two subcases

Case 1: where at least one letter appears in both $\mathbf{a}$ and $\mathbf{b}$ or where pairing equal letters in $\mathbf{a}$ and $\mathbf{b}$ results in a crossing. By noting that

$$\begin{aligned} \mathbb{E}[\text{tr}(\sigma_u^{\mathbf{a}}) \text{tr}(\sigma_v^{\mathbf{b}})] &= \mathbb{E}[\text{tr}(\sigma(1)_u^{\mathbf{a}}) \text{tr}(\sigma(1)_v^{\mathbf{b}}) \cdots \text{tr}(\sigma(n)_u^{\mathbf{a}}) \text{tr}(\sigma(n)_v^{\mathbf{b}})] \\ &= \mathbb{E}[\text{tr}(\sigma(1)_u^{\mathbf{a}}) \text{tr}(\sigma(1)_v^{\mathbf{b}})]^n, \end{aligned}$$

we may apply [Proposition 3.4.8](#) and [Lemma 3.4.9](#) to see that

$$m_n^{-p} \sum_{\substack{\mathbf{a}, \mathbf{b}: \text{wt}(\mathbf{ab})=p \\ \text{Case 1}}} \left| \mathbb{E}[r_u^{\mathbf{a}} r_v^{\mathbf{b}}] \mathbb{E}[\text{tr}(\sigma_u^{\mathbf{a}}) \text{tr}(\sigma_v^{\mathbf{b}})] \right| \leq \frac{(2p)!}{2^p p!} 2^{-2n} \leq (4e)^p p! 2^{-2n}, \quad (3.77)$$

where we have used the fact that the total number of terms in the sum is bounded $N(m_n, p)$.

Case 2: where the words $\mathbf{a}$ and $\mathbf{b}$ are distinct and pairing equal letters in $\mathbf{ab}$ results in a non-crossing pair partition. In this case

$$\mathbb{E}[\text{tr}(\sigma_u^{\mathbf{a}}) \text{tr}(\sigma_v^{\mathbf{b}})] = \mathbb{E}[\text{tr}(\sigma(1)_u^{\mathbf{a}})]^n \mathbb{E}[\text{tr}(\sigma(1)_v^{\mathbf{b}})]^n = 1.$$

We then see that

$$\mathbb{E}[r_u^{\mathbf{a}} r_v^{\mathbf{b}}] \mathbb{E}[\text{tr}(\sigma_u^{\mathbf{a}}) \text{tr}(\sigma_v^{\mathbf{b}})] = \prod_{(a,b) \in \pi_{\mathbf{a}}} \mathbb{1}_{u_a = u_b} \prod_{(c,d) \in \pi_{\mathbf{b}}} \mathbb{1}_{v_c = v_d},$$

where $\pi_{\mathbf{a}}$ is the non-crossing pair partition of the letters $\{1, \dots, |\mathbf{w}|\}$ defined by the pairing of the letters in $\mathbf{a}$, with $\pi_{\mathbf{b}}$ defined similarly. Words $\mathbf{ab}$ and $\mathbf{cd}$ in Case 2 are then equivalent if and only if $\pi_{\mathbf{a}} = \pi_{\mathbf{c}}$ and $\pi_{\mathbf{b}} = \pi_{\mathbf{d}}$. It follows that the total number of equivalence classes in Case 2 is given by $\tau(X_u) \tau(X_v)$. Using the counting arguments as in the previous proposition, we have that

$$\begin{aligned} m_n^{-p} \sum_{\substack{\mathbf{a}, \mathbf{b}: \text{wt}(\mathbf{ab})=p \\ \text{Case 2}}} \mathbb{E}[r_u^{\mathbf{a}} r_v^{\mathbf{b}}] \mathbb{E}[\text{tr}(\sigma_u^{\mathbf{a}}) \text{tr}(\sigma_v^{\mathbf{b}})] \\ = \tau(X_u) \tau(X_v) \frac{m_n(m_n - 1) \cdots (m_n - p + 1)}{m_n^p}. \end{aligned}$$

Using the fact

$$\left| \frac{m_n(m_n - 1) \cdots (m_n - p + 1)}{m_n^p} - 1 \right| \leq \frac{p^2}{2m},$$

we conclude that

$$\left| m_n^{-p} \sum_{\substack{\mathbf{a}, \mathbf{b}: \text{wt}(\mathbf{ab})=p \\ \text{Case 2}}} \mathbb{E}[r_u^{\mathbf{a}} r_v^{\mathbf{b}}] \mathbb{E}[\text{tr}(\sigma_u^{\mathbf{a}}) \text{tr}(\sigma_v^{\mathbf{b}})] - \tau(X_u) \tau(X_v) \right| \leq (4e)^p p! \frac{1}{m_n}. \quad (3.78)$$

Combining (3.76)-(3.77)-(3.78) gives

$$\left| m_n^{-p} \sum_{\mathbf{a}, \mathbf{b}} \mathbb{E} \left[r_{\mathbf{u}}^{\mathbf{a}} r_{\mathbf{v}}^{\mathbf{b}} \right] \mathbb{E} \left[\text{tr}(\sigma_{\mathbf{u}}^{\mathbf{a}}) \text{tr}(\sigma_{\mathbf{v}}^{\mathbf{b}}) \right] - \tau(X_{\mathbf{u}}) \tau(X_{\mathbf{v}}) \right| \leq (4e)^p p! (2m_n^{-1} + 2^{-2n}).$$

□

As a corollary of all the results, we can now prove the main result of this section.

Theorem 3.4.18 (Theorem 3.4.4). *For each n , suppose that $A_1^n, \dots, A_d^n$ are i.i.d. matrices from the random Pauli string ensemble of dimension $N = 2^n$ with m_n strings each. Let Z_γ^n be the solution to*

$$dZ_\gamma^n(s, t) = i \sum_{j=1}^d Z_\gamma^n(s, t) A_j^n d\gamma_t^j, \quad Z_\gamma^n(s, s) = I.$$

Then there exists $C > 0$ depending only on the length of γ and the dimension of the underlying space V for which

$$\mathbb{E} \left[\left(\text{tr}(Z_\gamma^n(s, t)) - \langle \gamma_{s,t} \rangle_{\mu_G^\infty} \right)^2 \right] < C(m_n^{-1} + 2^{-2n}). \quad (3.79)$$

Proof. We start by expanding

$$\begin{aligned} \left(\text{tr}(Z_\gamma^n(s, t)) - \langle \gamma_{s,t} \rangle_{\mu_P^\infty} \right)^2 &= \sum_{\mathbf{u}, \mathbf{v} \in \mathcal{W}_d} i^{|\mathbf{u}|+|\mathbf{v}|} \left(\text{tr}(A_{\mathbf{u}}) \text{tr}(A_{\mathbf{v}}) - \tau(X_{\mathbf{u}}) \tau(X_{\mathbf{v}}) \right. \\ &\quad \left. + \tau(X_{\mathbf{u}}) \tau(X_{\mathbf{v}}) - \text{tr}(A_{\mathbf{u}}) \tau(X_{\mathbf{v}}) \right. \\ &\quad \left. - \tau(X_{\mathbf{u}}) \text{tr}(A_{\mathbf{v}}) + \tau(X_{\mathbf{u}}) \tau(X_{\mathbf{v}}) \right) \mathcal{S}_{s,t}^{\mathbf{u}}(\gamma) \mathcal{S}_{s,t}^{\mathbf{v}}(\gamma). \end{aligned}$$

Let us deal with the first term.

$$\begin{aligned}
 & \sum_{\mathbf{u}, \mathbf{v} \in \mathcal{W}_d} \left| \mathbb{E}[\text{tr}(A_{\mathbf{u}}) \text{tr}(A_{\mathbf{v}})] - \tau(X_{\mathbf{u}}) \tau(X_{\mathbf{v}}) \right| \\
 & \leq \sum_{p=0}^{\infty} \sum_{|\mathbf{u}|=0}^{2p} \sum_{|\mathbf{v}|=2p-|\mathbf{u}|} \left| \mathbb{E}[\text{tr}(A_{\mathbf{u}}) \text{tr}(A_{\mathbf{v}})] - \tau(X_{\mathbf{u}}) \tau(X_{\mathbf{v}}) \right| \left| \mathcal{S}_{s,t}^{\mathbf{u}}(\gamma) \mathcal{S}_{s,t}^{\mathbf{v}}(\gamma) \right| \\
 & \leq (2m_n^{-1} + 2^{-2n}) \sum_{p=0}^{\infty} \sum_{k=0}^{2p} \sum_{\substack{|\mathbf{u}|=k \\ |\mathbf{v}|=2p-k}} (4e)^p p! \cdot \left| \mathcal{S}_{s,t}^{\mathbf{u}}(\gamma) \mathcal{S}_{s,t}^{\mathbf{v}}(\gamma) \right| \\
 & \leq (2m_n^{-1} + 2^{-2n}) \sum_{p=0}^{\infty} \sum_{k=0}^{2p} \frac{(4e)^p p! \cdot d^{2p} \|\gamma\|_{1\text{-var}}^{2p}}{k! (2p-k)!} \\
 & \leq (2m_n^{-1} + 2^{-2n}) \sum_{p=0}^{\infty} \frac{(2p+1)(4e)^p \cdot d^{2p} \|\gamma\|_{1\text{-var}}^{2p}}{p!} \\
 & \leq C_1(\|\gamma\|_{1\text{-var}}, d) (2m_n^{-1} + 2^{-2n}).
 \end{aligned}$$

The remaining two terms may each be bounded by

$$\begin{aligned}
 & \sum_{\mathbf{u}, \mathbf{v} \in \mathcal{W}_d} \left| \mathbb{E}[\text{tr}(A_{\mathbf{u}})] - \tau(X_{\mathbf{u}}) \right| \tau(X_{\mathbf{v}}) \left| \mathcal{S}_{s,t}^{\mathbf{u}}(\gamma) \mathcal{S}_{s,t}^{\mathbf{v}}(\gamma) \right| \\
 & \leq \sum_{\mathbf{u} \in \mathcal{W}_d} \left| \mathbb{E}[\text{tr}(A_{\mathbf{u}})] - \tau(X_{\mathbf{u}}) \right| \left| \mathcal{S}_{s,t}^{\mathbf{u}}(\gamma) \right| \sum_{\mathbf{v} \in \mathcal{W}_d} \tau(X_{\mathbf{v}}) \left| \mathcal{S}_{s,t}^{\mathbf{v}}(\gamma) \right| \\
 & \leq (2m_n^{-1} + 2^{-2n}) \sum_{p=0}^{\infty} \frac{(4e)^p d^{2p} \|\gamma\|_{1\text{-var}}^{2p}}{p!} \sum_{q=0}^{\infty} \frac{2^q d^{2q} \|\gamma\|_{1\text{-var}}^{2q}}{(2q)!} \\
 & \leq C_2(\|\gamma\|_{1\text{-var}}, d) (2m_n^{-1} + 2^{-2n}).
 \end{aligned}$$

Taking $C = 3 \max\{C_1, C_2\}$ yields the result. $\square$

To see how the results of this section can be used to derive an efficient quantum algorithm for [Eq. \(3.60\)](#), we refer the reader to [\[Cre+25\]](#).

3.5 Functional equation for interacting matrix models

We conclude this section by extending [Theorem 3.1.1](#) to the case of perturbations of the Gaussian Unitary Ensemble (GUE); we will again leverage the Schwinger-Dyson equations to derive the corresponding limiting functional equations in this setting. To this end consider $P \in \mathcal{C}_d$ given by

$$P = \frac{1}{2} \sum_{i=1}^d X_i^2 + W = \frac{1}{2} \sum_{i=1}^d X_i^2 + \sum_{j=1}^m g_j X_{v_j},$$

satisfying the conditions of [Theorem 3.0.12](#) with associated non-commutative law τ_P . Also associate to W the path-dependent differential operators

$$\mathcal{D}_W^k := - \sum_{j=1}^m (-i)^{|v_j|} g_j \nabla^{D_k v_j}, \quad (3.80)$$

where D_k is the cyclic derivative from [Definition 3.0.9](#). We define the following functionals on $C^{1\text{-var}}([0, T], V)$:

$$\begin{aligned} \langle \gamma_{a,b} \rangle_{\mu_P^N} &:= \sum_{\mathbf{w} \in \mathcal{W}_d} i^{|\mathbf{w}|} \mathbb{E}_{\mu_P^N} [\text{tr}(A_{\mathbf{w}})] \mathcal{S}_{a,b}^{\mathbf{w}}(\gamma), \\ \langle \gamma_{a,b} \rangle_{\mu_P^\infty} &:= \sum_{\mathbf{w} \in \mathcal{W}_d} i^{|\mathbf{w}|} \lim_{N \rightarrow \infty} \mathbb{E}_{\mu_P^N} [\text{tr}(A_{\mathbf{w}})] \mathcal{S}_{a,b}^{\mathbf{w}}(\gamma). \end{aligned}$$

Proposition 3.5.1. *Suppose that P satisfies the assumptions of [Theorem 3.0.12](#) and for each $s \in [0, T]$ let $U_{s,u}$ be the U_N -valued solution to the controlled differential equation*

$$dU_{s,u} = i \sum_{j=1}^d U_{s,u} A_j d\gamma_u^j, \quad U_{s,s} = I_N,$$

where $(A_1^N, \dots, A_d^N) \sim \mu_P^N$. Then

$$\mathbb{E}_{\mu_P^N} [\text{tr}(U_{s,t})] = \langle \gamma_{s,t} \rangle_{\mu_P^N} < \infty \quad \text{and} \quad \lim_{N \rightarrow \infty} \mathbb{E}_{\mu_P^N} [\text{tr}(U_{s,t})] = \langle \gamma_{s,t} \rangle_{\mu_P^\infty} < \infty. \quad (3.81)$$

The content of the above is simply a statement that both limits and expectations may be exchanged with the sum over words.

Proof. The expansion of the path ordered exponential in terms of its signature provides that

$$\text{tr}(U_{s,t}^N) = \sum_{\mathbf{w} \in \mathcal{W}_d} i^{|\mathbf{w}|} \text{tr}(A_{\mathbf{w}}^N) \mathcal{S}_{s,t}^{\mathbf{w}}(\gamma).$$

To attain both parts of (3.81), we simply need to justify the interchange of the limits and expectation with the sum over words $\mathbf{w}$. To do this, it is enough to show that

$$\sum_{\mathbf{w} \in \mathcal{W}_d} \mathbb{E}_{\mu_P^N} \left[\left| \text{tr}(A_{\mathbf{w}}^N) \right| \right] |\mathcal{S}_{s,t}^{\mathbf{w}}(\gamma)|$$

is uniformly bounded in N . The sub-Gaussian nature of the tail behaviour of $\lambda_{\max}^N(\mathbf{A})$ (see [Lemma 3.0.13](#)) yields the following bound

$$\mathbb{E}_{\mu_P^N} \left[\left| \text{tr}(A_{\mathbf{w}}^N) \right| \right] \leq \mathbb{E}_{\mu_P^N} \left[\lambda_{\max}^N(\mathbf{A})^{|\mathbf{w}|} \right] \leq \kappa^{|\mathbf{w}|} \Gamma\left(\frac{|\mathbf{w}|}{2} + 1\right) \quad (3.82)$$

for some $\kappa > 0$, where Γ denotes the usual Gamma function. The second inequality uses a common equivalent definition of sub-Gaussian random variables, see for example [Ver18, Proposition 2.5.2]. Then, by the factorial decay of the signature

$$\sum_{\mathbf{w} \in \mathcal{W}_d} \mathbb{E}_{\mu_P^N} \left[|\text{tr}(A_{\mathbf{w}}^N)| \right] |\mathcal{S}_{s,t}^{\mathbf{w}}(\gamma)| \leq \sum_{\mathbf{w} \in \mathcal{W}_d} \kappa^{|\mathbf{w}|} \Gamma\left(\frac{|\mathbf{w}|}{2} + 1\right) |\mathcal{S}_{s,t}^{\mathbf{w}}(\gamma)| \leq \sum_{n=0}^{\infty} \frac{\tilde{\kappa}^n \Gamma\left(\frac{n}{2} + 1\right)}{\Gamma(n+1)} < \infty,$$

for some $\tilde{\kappa} > 0$. Then Fubini-Tonelli and the dominated convergence theorem allows for the interchange of limits and expectation with the sum of words. $\square$

We now turn to the main result of this section.

Theorem 3.5.2. *The functional $\langle \gamma_{s,t} \rangle_{\mu_P^\infty}$ satisfies the path-dependent integro-differential equation*

$$\langle \gamma_{s,t} \rangle_{\mu_P^\infty} = 1 - \int_{s \leq u \leq v \leq t} \langle \gamma_{s,u} \rangle_{\mu_P^\infty} \langle \gamma_{u,v} \rangle_{\mu_P^\infty} \langle d\gamma_u, d\gamma_v \rangle - \sum_{k=1}^d \int_s^t \mathcal{D}_W^k \langle \gamma_{s,u} \rangle_{\mu_P^\infty} d\gamma_u^k, \quad (3.83)$$

with the boundary condition $\langle \gamma_{s,s} \rangle_{\mu_P^\infty} = 1$ for all $s \in [0, T]$.

The proof follows the same method as in the proof of Theorem 3.1.1, while accounting for the additional terms coming from the perturbation W .

Proof. Recall from Theorem 3.0.12 that

$$\lim_{N \rightarrow \infty} \mathbb{E}_{\mu_P^N} [\text{tr}(A_{\mathbf{w}})] = \tau_P(X_{\mathbf{w}}),$$

where τ_P is the unique non-commutative law associated with P solving the corresponding Schwinger-Dyson equation. Consider now the following expansion where we use (3.13) for the first equality.

$$\begin{aligned} \sum_{\mathbf{w} \in \mathcal{W}_d} i^{|\mathbf{w}|+1} \tau_P \otimes \tau_P(\partial_k X_{\mathbf{w}}) \mathcal{S}_{s,t}^{\mathbf{w}}(\gamma) &= \sum_{\mathbf{u}, \mathbf{v} \in \mathcal{W}_d} i^{|\mathbf{u}|+|\mathbf{v}|+2} \tau_P(X_{\mathbf{u}}) \tau_P(X_{\mathbf{v}}) \mathcal{S}_{s,t}^{u k v}(\gamma) \\ &= - \sum_{\mathbf{u}, \mathbf{v} \in \mathcal{W}_d} i^{|\mathbf{u}|+|\mathbf{v}|} \tau_P(X_{\mathbf{u}}) \tau_P(X_{\mathbf{v}}) \int_s^t \mathcal{S}_{s,u}^{\mathbf{u}}(\gamma) \mathcal{S}_{u,t}^{\mathbf{v}}(\gamma) d\gamma_u^k \\ &= - \int_s^t \left(\sum_{\mathbf{u} \in \mathcal{W}_d} i^{|\mathbf{u}|} \tau_P(X_{\mathbf{u}}) \mathcal{S}_{s,u}^{\mathbf{u}}(\gamma) \right) \left(\sum_{\mathbf{v} \in \mathcal{W}_d} i^{|\mathbf{v}|} \tau_P(X_{\mathbf{v}})(\gamma) \mathcal{S}_{u,t}^{\mathbf{v}} \right) d\gamma_u^k \\ &= - \int_s^t \langle \gamma_{s,u} \rangle_{\mu_P^\infty} \langle \gamma_{u,t} \rangle_{\mu_P^\infty} d\gamma_u^k. \end{aligned}$$

We now expand the left-hand-side using the Schwinger-Dyson equation

$$\begin{aligned}
 \sum_{\mathbf{w} \in \mathcal{W}_d} i^{|\mathbf{w}|+1} \tau_P \otimes \tau_P(\partial_k X_{\mathbf{w}}) \mathcal{S}_{s,t}^{\mathbf{w}}(\gamma) &= \sum_{\mathbf{w} \in \mathcal{W}_d} i^{|\mathbf{w}|+1} \tau_P(X_{\mathbf{w}} D_k V) \mathcal{S}_{s,t}^{\mathbf{w}}(\gamma) \\
 &= \sum_{\mathbf{w} \in \mathcal{W}_d} i^{|\mathbf{w}|+1} \left(\tau(X_{\mathbf{w}} X_k) + \sum_{j=1}^m g_j \tau_P(X_{\mathbf{w}} X_{D_k v_j}) \right) \mathcal{S}_{s,t}^{\mathbf{w}}(\gamma) \\
 &= \sum_{\mathbf{w} \in \mathcal{W}_d} i^{|\mathbf{w}|} \tau(X_{\mathbf{w}}) \nabla^k \mathcal{S}_{s,t}^{\mathbf{w}}(\gamma) + \sum_{j=1}^m (-i)^{|v_j|-2} g_j \sum_{\mathbf{w} \in \mathcal{W}_d} \tau_P(X_{\mathbf{w}}) i^{|\mathbf{w}|} \nabla^{D_k v_j} \mathcal{S}_{s,t}^{\mathbf{w}}(\gamma) \\
 &= \nabla^k \sum_{\mathbf{w} \in \mathcal{W}_d} i^{|\mathbf{w}|} \tau(X_{\mathbf{w}}) \mathcal{S}_{s,t}^{\mathbf{w}}(\gamma) + \sum_{j=1}^m (-i)^{|v_j|-2} g_j \nabla^{D_k v_j} \sum_{\mathbf{w} \in \mathcal{W}_d} \tau_P(X_{\mathbf{w}}) \mathcal{S}_{s,t}^{\mathbf{w}}(\gamma) \\
 &= \nabla^k \langle \gamma_{s,t} \rangle_{\mu_P^\infty} + \sum_{j=1}^m (-i)^{|v_j|-2} g_j \nabla^{D_k v_j} \langle \gamma_{s,t} \rangle_{\mu_P^\infty} \\
 &= (\nabla^k + \mathcal{D}_W^k) \langle \gamma_{s,t} \rangle_{\mu_P^\infty}.
 \end{aligned}$$

Equating these identities, integrating both sides with respect to $d\gamma^k$ and then summing over k results in Eq. (3.83)

$$\langle \gamma_{s,t} \rangle_{\mu_P^\infty} = 1 - \int_{s \leq u \leq v \leq t} \langle \gamma_{s,u} \rangle_{\mu_P^\infty} \langle \gamma_{u,v} \rangle_{\mu_P^\infty} \langle d\gamma_u, d\gamma_v \rangle - \sum_{k=1}^d \int_s^t \mathcal{D}_W^k \langle \gamma_{s,u} \rangle_{\mu_P^\infty} d\gamma_u^k. \quad (3.84)$$

□

Example 3.5.3 (GUE). Consider the case of the quadratic potential $G = \frac{1}{2} \sum_{i=1}^d X_i^2$, for which μ_G^N corresponds to the GUE. Then each $\mathcal{D}_W^k \equiv 0$, and we recover Eq. (3.45)

$$\langle \gamma_{s,t} \rangle_{\mu_G^\infty} = 1 - \int_{s \leq u \leq v \leq t} \langle \gamma_{s,u} \rangle_{\mu_G^\infty} \langle \gamma_{u,v} \rangle_{\mu_G^\infty} \langle d\gamma_u, d\gamma_v \rangle.$$

To conclude this subsection, we consider the uniqueness of solutions to (3.83). A complete theory for uniqueness in a general space of functions on path space is a topic for future research. Instead, we consider infinite linear functionals on the signatures.

Proposition 3.5.4. Consider the set of infinite linear functionals $\ell : C^{1-var}([0, T], V) \rightarrow \mathbb{C}$ on the signature

$$\ell(\gamma_{s,t}) = \sum_{\mathbf{w} \in \mathcal{W}_d} i^{|\mathbf{w}|} a_{\mathbf{w}} \mathcal{S}_{s,t}^{\mathbf{w}}(\gamma),$$

for which $|a_w| \leq C^{|w|}$ for some $C > 0$. Suppose that

$$\ell(\gamma_{s,t}) = 1 - \int_{s \leq u \leq v \leq t} \ell(\gamma_{s,u}) \ell(\gamma_{u,v}) \langle d\gamma_u, d\gamma_v \rangle - \sum_{k=1}^d \int_s^t \mathcal{D}_W^k \ell(\gamma_{s,u}) d\gamma_u^k \quad (3.85)$$

holds for all $\gamma \in \mathcal{X}$, then $a_w = \tau_P(X_w)$ for all $w \in \mathcal{W}_d$.

Proof. Expanding (3.85) gives

$$\begin{aligned} \sum_{w \in \mathcal{W}_d} i^{|w|} a_w \mathcal{S}_{s,t}^w(\gamma) &= 1 + \sum_{k=1}^d \sum_{u,v \in \mathcal{W}_d} i^{|u|+|v|+2} a_u a_v \mathcal{S}_{s,t}^{ukvk}(\gamma) \\ &\quad + \sum_{k=1}^d \sum_{u \in \mathcal{W}_d} \sum_{j=1}^m i^{|u|+|v_j|-2-|v_j|} g_j a_u D_k v_j \mathcal{S}_{s,t}^{uk}(\gamma). \end{aligned}$$

Noting that the left-hand-side may be expanded as

$$\sum_{w \in \mathcal{W}_d} i^{|w|} a_w \mathcal{S}_{s,t}^w(\gamma) = 1 + \sum_{k=1}^d \sum_{w \in \mathcal{W}_d} i^{|w|+1} a_{wk} \mathcal{S}_{s,t}^{wk}(\gamma),$$

meaning that we may equate the coefficients of $\mathcal{S}^{wk}$ to obtain the system of equations

$$a_{wk} = \sum_{u=ukv} a_u a_v - \sum_{j=1}^m g_j a_w D_k v_j,$$

which after rearranging yields

$$\sum_{w=ukv} a_u a_v = a_{wk} + \sum_{j=1}^m g_j a_w D_k v_j.$$

This implies that the extension of ℓ to a linear functional $\tau_a : \mathcal{C}_d \rightarrow \mathbb{C}$ satisfies the Schwinger-Dyson equation $\tau_a \otimes \tau_a(\partial_k X_w) = \tau_a(X_w D_k P)$. By the uniqueness of solutions to the Schwinger-Dyson equation ([Theorem 3.0.12](#)) we must have $\tau_a = \tau_P$. $\square$

CHAPTER 4

SIGNATURE ASYMPTOTICS AND WASSERSTEIN DISTANCES

Our next example of a function on unparameterised path space is the asymptotic limit of suitably scaled norms of signature terms. While it is easy to compute the signature given a representative of an equivalence class of unparameterised paths, it is a non-trivial task to recover information about representatives given their signature. While certain information such as the increment or area is straightforward, it is significantly more challenging to recover the tree-reduced length or total curvature. The study of this problem is connected to the aforementioned asymptotic limit and is the subject of several open conjectures. We will see that under a suitable choice of norm, the leading order term in the asymptotic expansion is the tree-reduced length and the second order term is a weighted curvature of the tree-reduced path. This chapter extends work by Hambly & Lyons [HL10] and is based on the paper [CMT23].

Some terms in the signature are explicitly relatable to properties of the original path, for example the increment and the area can be recovered from the terms of order 1 and order 2 respectively. Recovering more granular information demands a more sophisticated approach. A rich stream of recent work has tackled the explicit reconstruction of the tree-reduced representative from its signature. For example in the probabilistic setting, the work of Le Jan & Qian [LQ13] showed how to construct a Brownian sample path given its Stratonovich signature using polygonal approximations. This has since been generalised to hypoelliptic diffusions [GQ16], Gaussian processes [BG15] and even to the deterministic setting [Gen17]. However, the approaches presented in the aforementioned papers are not implementable in practice. In the more realistic setting of piecewise linear paths, the work in

[LX17] showed how the hyperbolic development can be used to find the length and direction of the last linear piece. Iterating this approach allows one to reconstruct the entire trajectory. On the other hand, [LX18] proposed a symmetrisation procedure on the signature which leads to a reconstruction algorithm in some cases. Both of these latter approaches have the advantage of being implementable.

An alternative branch of investigation is the recovery of broad features of a path, such as its length or curvature properties, using the asymptotics of its signature. This problem has been an active area of research over the course of the preceding decade and a half. Indeed, Hambly & Lyons initiated this direction in [HL10], with the papers [BG19b; CLN18b; BG19a; BGS20] all contributing to this direction. Recall that if γ is absolutely continuous, that is $\gamma \in C^{0,1\text{-var}}([0, T], V)$, its length is given by

$$\|\gamma\|_{1\text{-var};[0,t]} = \int_0^t |\gamma'_s| ds, \quad (4.1)$$

and denoted by $l := \|\gamma\|_{1\text{-var};[0,T]}$. Such a curve admits a unit-speed reparametrisation. That is, the existence of a path $\tilde{\gamma}$ and reparameterisation $\alpha : [0, T] \rightarrow [0, l]$ such that

$$\gamma = \tilde{\gamma} \circ \alpha$$

where $|\tilde{\gamma}'| \equiv 1$.

Remark 4.0.1. *The unit speed reparameterisation is closely related to the constant speed reparameterisation we leveraged in Chapter 2. Indeed, a constant-speed parameterised curve γ satisfies $|\tilde{\gamma}'| \equiv \frac{1}{T} \|\gamma\|_{1\text{-var};[0,T]}$ on $[0, T]$, while the unit speed parameterisation satisfies $|\tilde{\gamma}'| \equiv 1$ on $[0, \|\gamma\|_{1\text{-var};[0,T]}]$.*

Hambly and Lyons [HL10] initially showed that the arc-length of a unit-speed path can be recovered from the asymptotics of the norm of terms in the signature under a broad class of norms. Concretely, they proved that if $\gamma : [0, l] \rightarrow V$ is a continuously differentiable unit-speed curve whose derivative satisfies a modulus of continuity condition, then it holds that

$$\lim_{n \rightarrow \infty} \frac{n! \|\mathcal{S}(\gamma)^n\|}{l^n} = 1 \quad (4.2)$$

for a class of norms $\|\cdot\|$ on the tensor product $V^{\otimes n}$ including the projective tensor norm but excluding the Hilbert-Schmidt norm in the case where V is endowed with an inner product. This result has since been extended by Lyons and Xu [LX17] to all unit speed C^1 curves. For the same class of norms, this also implies the weaker statement

$$\lim_{n \rightarrow \infty} \left(n! \|\mathcal{S}(\gamma)^n\| \right)^{1/n} = l. \quad (4.3)$$

A natural question is whether other properties of the original path can also be recovered in a similar fashion, such as (a functional of) its curvature.

When the Hilbert-Schmidt norm is considered instead, [HL10] also showed that Eq. (4.3) holds, but that statement Eq. (4.2) fails to be true in general. Under the assumption that γ is three times continuously differentiable it was proved that¹

$$\lim_{n \rightarrow \infty} \frac{n! \|\mathcal{S}(\gamma)^n\|}{l^n} = c(\gamma) := \mathbb{E} \left[\exp \left(\int_0^l (B_{0,s})^2 \langle \gamma'_s, \gamma_s''' \rangle ds \right) \right]^{1/2}, \quad (4.4)$$

where $(B_{0,s})_{s \in [0,l]}$ is a Brownian bridge which returns to zero at time l . It can be seen that $c(\gamma) < 1$ unless γ is a straight line, and so Eq. (4.3) indeed holds. More recent articles have focused on proving statements similar to Eq. (4.3) under the fewest possible assumptions on γ and on the tensor norms. For example, in the paper [CLN18b] it was proved that if γ is continuous and of finite 1-variation, then under any reasonable tensor norm we have that

$$\lim_{n \rightarrow \infty} \left(n! \|\mathcal{S}(\gamma)^n\| \right)^{1/n} = \sup_{n \geq 1} \left(n! \|\mathcal{S}(\gamma)^n\| \right)^{1/n} > 0 \quad (4.5)$$

provided that the sequence $\{\|\mathcal{S}(\gamma)^n\| : n = 1, 2, \dots, \infty\}$ does not contain an infinite subsequence of zeros. Boedihardjo and Geng [BG19b] strengthened this result by proving that the existence of such a subsequence is equivalent to the underlying path being tree-like, see also [CLN18a]. Taken together, these papers prove that the identity in Eq. (4.5) holds true for a wide class of continuous bounded variation paths. It is conjectured that for a tree-reduced path γ the limit is exactly the length of γ , see [CLN18a]. This question remains open at the time of writing.

4.0.1 Contributions

In this chapter, we contribute to the effort of recovering broad features of the original path from its signature by way of a novel approach. We do so by explicitly relating the Hilbert-Schmidt norm of projected signatures with p -Wasserstein distances between discrete probability measures, allowing the study of the former using tools of the latter. These measures are characterised in terms of γ only through an integral equation, making the contribution of the geometrical properties of γ (such as its curvature) in the limit of the norm explicit.

¹ We refer to the right-hand side term $c(\gamma)$ as the *Hambly-Lyons limit*.

Remark 4.0.2. *To ease notation, from this point forward we consider unit speed paths parameterised on $[0, 1]$ rather than $[0, l]$. Indeed, it is clear from the left-hand-side of Eq. (4.4) that the limit is invariant under the scaling of γ . Additionally for $\sigma = \lambda\gamma$ with $\lambda \in \mathbb{R}_+$ we see that*

$$\begin{aligned} \int_0^{\lambda l} (B_{0,s})^2 \langle \sigma'_s, \sigma'''_s \rangle ds &= \int_0^{\lambda l} (B_{0,s})^2 \lambda^2 \langle \gamma'_{s/\lambda}, \gamma'''_{s/\lambda} \rangle ds \\ &= \int_0^l (B_{0,\lambda s})^2 \lambda \langle \gamma'_s, \gamma'''_s \rangle ds \\ &\stackrel{d}{=} \int_0^l (\tilde{B}_{0,s})^2 \langle \gamma'_s, \gamma'''_s \rangle ds, \end{aligned}$$

where the second equality follows from a change of variable and the third is equality in distribution where $\tilde{B}_{0,\cdot}$ is a Brownian bridge returning to 0 at time l . The claimed equality in distribution follows from the well-known scale invariance of Brownian bridges.

The insight to this connection originates from a theorem by del Barrio, Giné, and Utze (see Theorem 4.1.3 below), which can be used to recast the Hambly-Lyons limit $c(\gamma)$ as the limit of a 2-Wasserstein distance between empirical measures, see Section 4.1. Formally, for a twice continuously differentiable unit speed path $\gamma : [0, 1] \rightarrow V$ that is regular enough (see exact conditions in Proposition 4.1.11), we construct and show the existence of a measure μ on $\mathbb{R}$, prescribed in terms of the following integral equation for its cumulative distribution function F

$$F(t) = \int_0^t |\gamma''_{F(s)}|^{-1} ds, \quad \lim_{s \rightarrow 0^+} F(s) = 0, \quad \lim_{s \rightarrow 1^-} F(s) = 1. \quad (4.6)$$

When coupled with the B-G-U theorem we show that

$$c(\gamma) = \lim_{n \rightarrow \infty} \mathbb{E} [\exp(-nW_2^2(\mu_n, \mu))], \quad (4.7)$$

so that the Hambly-Lyons limit $c(\gamma)$ can be seen as the limit of 2-Wasserstein distances between μ and an empirical version $\mu_n := \sum_{i=1}^n \delta_{F^{-1}(U_i)}$. Here δ denote the Dirac delta and $\{U_i\}_{i \in \{1, \dots, n\}}$ is a sample of n independent uniform random variables on $[0, 1]$.

In Section 4.2 (motivated by the above) we derive relationships between the signature inner product $\langle \mathcal{S}(\gamma)^n, \mathcal{S}(\sigma)^n \rangle$ and a series of p -Wasserstein distances. Then as an application, we re-derive a generalised version of the Hambly-Lyons limit theorem (Section 4.2.4) through the lens of discrete optimal transport by exploiting asymptotic results of Wasserstein distances between empirical measures [BL19].

Theorem 4.0.3 (Generalised Hambly-Lyons limit theorem). *Let $\gamma : [0, 1] \rightarrow V$ be a twice continuously differentiable unit-speed path such that the map $s \mapsto |\gamma_s''|$ is non-vanishing, and differentiable with bounded derivative. Then*

$$\lim_{n \rightarrow \infty} n! \|\mathcal{S}(\gamma)^n\| = \mathbb{E} \left[\exp \left(- \int_0^1 (B_{0,s})^2 |\gamma_s''|^2 ds \right) \right]^{1/2}. \quad (4.8)$$

By combining [Remark 4.0.2](#) and the right-hand side of [Eq. \(4.8\)](#), the Hambly-Lyons limit $c(\gamma)$ can be viewed as a weighted (scale-invariant) functional of the curvature of γ . [Section 4.2.5](#) concludes this chapter by presenting a way to practically compute the limit through the solving of a second order distributional differential equation.

4.1 The Hambly-Lyons limit and Wasserstein distances

This section outlines the proof of [Theorem 4.0.3](#). We do so by first recalling a theorem by del Barrio, Giné, and Utze ([Theorem 4.1.3](#) below) and observing that under some conditions, the former allows the rewriting of the Hambly-Lyons limit $c(\gamma)$ as a limit in terms of the 2-Wasserstein distance between empirical measures. The rest of this section will investigate the assumptions needed on γ for these conditions to be fulfilled.

4.1.1 The B-G-U theorem

Before presenting the aforementioned theorem of del Barrio, Giné, and Utze, we start by recalling the definition of the p -Wasserstein distances that form a core part of our analysis.

Definition 4.1.1 (Wasserstein Distance). *Let μ and ν be probability measures supported on $\mathbb{R}$. The p^{th} Wasserstein distance $W_p(\mu, \nu)$ between μ and ν is defined by*

$$W_p^p(\mu, \nu) := \inf_{\pi \in \Gamma(\mu, \nu)} \int_{\mathbb{R}} \int_{\mathbb{R}} |x - y|^p d\pi(x, y), \quad (4.9)$$

where $\Gamma(\mu, \nu)$ denotes the set of all couplings of μ and ν . If μ and ν have bounded support then one may also define the $p = \infty$ distance by

$$W_\infty(\mu, \nu) := \lim_{p \rightarrow \infty} W_p(\mu, \nu). \quad (4.10)$$

Definition 4.1.2 (J_p functional and I -function $I(t)$). *Let μ be a probability measure on $\mathbb{R}$ which has a density f with respect to the Lebesgue measure, and let F be the associated*

distribution function. The J_p functional is defined as

$$J_p(\mu) := \int_{-\infty}^{\infty} \frac{[F(x)(1 - F(x))]^{p/2}}{f(x)^{p-1}} dx \quad (4.11)$$

for $p \in \mathbb{N}$. Moreover, if F admits an absolutely continuous inverse on $(0, 1)$ (or equivalently by virtue of [BL19, Proposition A.17], if μ is supported on an interval, finite or not, and the absolutely continuous component of μ has an almost everywhere positive density on that interval), define the I-function $I(t)$ for almost all $t \in (0, 1)$ as

$$I(t) := f(F^{-1}(t)) \stackrel{(A.18)}{=} \frac{1}{(F^{-1})'(t)} \quad (4.12)$$

where the last equality exploited [BL19, Proposition A.18].

For the rest of this chapter we denote by μ_n the empirical measure defined as

$$\mu_n := \frac{1}{n} \sum_{i=1}^n \delta_{X_i}, \quad (4.13)$$

where for $i \in \{1, 2, \dots, n\}$, $X_i \stackrel{\text{i.i.d.}}{\sim} \mu$.

Theorem 4.1.3 (B-G-U Theorem; [dBGU05]). *Let μ be a measure supported on $(a, b) \subset \mathbb{R}$ admitting a density f . Assume further that f is positive, differentiable, satisfies*

$$\sup_{a < x < b} \frac{F(x)(1 - F(x))}{f(x)^2} |f'(x)| < \infty, \quad (4.14)$$

and that $J_2(\mu) < \infty$. Denote by F its distribution function. Then,

$$nW_2^2(\mu_n, \mu) \rightarrow \int_0^1 \frac{(B_{0,t})^2}{I(t)^2} dt \quad (4.15)$$

as $n \rightarrow \infty$ weakly in $\mathbb{R}$ where $B_{0,t}$ is a Brownian bridge starting at 0 and vanishing at $t = 1$.

Remark 4.1.4 (Origin of assumption Eq. (4.14)). *The condition in Eq. (4.14) has been a recurrent assumption in several asymptotic results involving quantile processes and ultimately goes back to Csorgo and Revesz [CR78]. To better understand its connection with the B-G-U Theorem 4.1.3, recall the following identity*

$$W_2^2(\mu_n, \mu) = \int_0^1 (F_n^{-1}(s) - F^{-1}(s))^2 ds = \frac{1}{n} \int_0^1 \xi_n(s)^2 ds, \quad (4.16)$$

with $\xi_n(s) := \sqrt{n} (F_n^{-1}(s) - F^{-1}(s))$ denoting the quantile process. It is shown in [dBGU05] that ξ_n converges weakly in $L^2(0, 1)$ to $B_{0,t}/I(t)$. For a slightly more general object, the so-called normed sample quantile process ρ_n , it can be shown that requiring Eq. (4.14) allows one to asymptotically control ρ_n . More details can be found in [Csö83].

The following corollary will form a key component of our generalisation of the Hambly-Lyons result.

Corollary 4.1.5. *Let μ be a measure supported on $(a, b) \subset \mathbb{R}$ which satisfies the assumptions of Theorem 4.1.3, then*

$$nW_2^2(\mu_n^1, \mu_n^2) \rightarrow 2 \int_0^1 \frac{(B_{0,t})^2}{I(t)^2} dt \quad (4.17)$$

weakly in $\mathbb{R}$ as $n \rightarrow \infty$, where μ_n^1 and μ_n^2 are independent copies of empirical measures from μ .

Proof. Let ξ_n^1 and ξ_n^2 denote the quantile processes of μ_n^1 and μ_n^2 defined in Remark 4.1.4. Since $L^2(0, 1)$ is separable, [Bil68, Theorem 2.8] implies that $\xi_n^1 - \xi_n^2$ converges weakly in $L^2(0, 1)$ to $(B_{0,t}^1 - B_{0,t}^2)/I(t)$ for independent Brownian bridges B^1 and B^2 . Since the difference of independent Brownian bridges is itself a Brownian bridge with twice the variance, we can apply the continuous mapping theorem and Remark 4.1.4 to conclude that

$$nW_2^2(\mu_n^1, \mu_n^2) = \int_0^1 (\xi_n^1(t) - \xi_n^2(t))^2 dt \rightarrow 2 \int_0^1 \frac{(B_{0,t})^2}{I(t)^2} dt$$

weakly in $\mathbb{R}$. □

The existence of a bridge between signature asymptotics and the theory of empirical processes is hinted at when one considers a special instance of the B-G-U Theorem 4.1.3. Indeed, when applied to a regular enough class of measures μ satisfying $I(t) = |\gamma_t''|^{-1}$, the right-hand side of B-G-U exactly coincides with the Hambly-Lyons limit. The following remark formalises this observation.

Corollary 4.1.6 (Recovering the Hambly-Lyons limit). *Let $\gamma : [0, 1] \rightarrow V$ be a twice continuously differentiable unit-speed path with non-vanishing second derivative. Consider a probability measure μ on $\mathbb{R}$ supported on $(0, b)$ having density f . Assume that the following four conditions are satisfied.*

- A) *The density function f is positive and differentiable everywhere;*
- B) *Its associated I -function satisfies $I(t) = |\gamma_t''|^{-1}$ almost everywhere;*

C) The distribution function F and density function f satisfy

$$\sup_{0 < x < 1} \frac{F(x)(1 - F(x))}{f(x)^2} |f'(x)| < \infty; \quad (4.18)$$

D) $J_2(\mu) < \infty$.

Now the B-G-U [Theorem 4.1.3](#) implies that the Hambly-Lyons limit in $c(\gamma)$ can be rewritten as the limit of a 2-Wasserstein distance, that is

$$c(\gamma) = \lim_{n \rightarrow \infty} \mathbb{E} \left[\exp \left(-n W_2^2(\mu_n, \mu) \right) \right]^{1/2}. \quad (4.19)$$

Proof. The assumptions for the B-G-U [Theorem 4.1.3](#) are satisfied and the limit

$$\lim_{n \rightarrow \infty} n W_2^2(\mu_n, \mu) = \int_0^1 B_{0,t}^2 |\gamma''|^2 dt < \infty \quad (4.20)$$

exists under the stated assumptions. If one further assumes that the assumptions of the Hambly-Lyons limit [Eq. \(4.4\)](#) holds (i.e. that γ is of class C^3), then the limits coincide as γ is unit-speed and satisfies $\langle \gamma', \gamma''' \rangle = -|\gamma''|^2$. $\square$

4.1.2 Existence and characterisation of admissible measures

The rest of this section will focus on characterising the paths, γ , for which a measure μ exists satisfying the conditions A , B , C and D of [Corollary 4.1.6](#). Firstly we recast the determination of such μ into the solving of an integral equation in terms of γ only, and then formulate the assumptions on γ ensuring the existence of a solution to this equation whose associated measure μ satisfies the relevant assumptions.

We start by rewriting condition B as an explicit condition on the cumulative distribution function F associated to the measure μ .

Remark 4.1.7 (Reformulating condition B as an integral equation for F). *Let $\gamma : [0, 1] \rightarrow V$ be a twice continuously differentiable unit-speed path with non-vanishing second derivative. If F admits an absolutely continuous inverse on $(0, 1)$ (which is the case whenever condition A holds; see [\[BL19, Proposition A.17\]](#)), then condition B holds if and only if F satisfies the integral equation*

$$F(t) = \int_0^t |\gamma''_{F(s)}|^{-1} ds, \quad (4.21)$$

for all $t \in (0, b)$ with boundary conditions $\lim_{s \rightarrow 0^+} F(s) = 0$, $\lim_{s \rightarrow b^-} F(s) = 1$. To see this, first suppose that conditions A and B hold, then

$$\int_0^t I(F(s)) ds \stackrel{(4.12)}{=} \int_0^t f(s) ds = F(t) \stackrel{B}{=} \int_0^t |\gamma''_{F(s)}|^{-1} ds.$$

Assume now that condition A and Eq. (4.21) hold, then the Lebesgue differentiation theorem implies that the density function associated to F can be written as

$$f(s) = |\gamma''_{F(s)}|^{-1}, \quad \text{a.e..} \quad (4.22)$$

And so

$$I(s) = f(F^{-1}(s)) = |\gamma''_{F(F^{-1}(s))}|^{-1} = |\gamma''_s|^{-1}, \quad \text{almost everywhere.}$$

Lemma 4.1.8. Let $\gamma : [0, 1] \rightarrow V$ be a twice continuously differentiable unit-speed path with non-vanishing second-derivative. Assume the existence of a twice-differentiable function $F : (0, b) \rightarrow [0, 1]$ satisfying

$$F(t) = \int_0^t |\gamma''_{F(s)}|^{-1} ds, \quad \lim_{s \rightarrow 0^+} F(s) = 0, \quad \lim_{s \rightarrow b^-} F(s) = 1, \quad (4.23)$$

with bounded second derivative F'' . Then

- (i) F is a cumulative distribution function with support on $(0, b)$ and admits an absolutely continuous inverse F^{-1} on $(0, 1)$;
- (ii) F fulfils condition A, B, C , and D .

Proof. That F is a cumulative distribution function corresponding to some measure μ supported on $(0, b)$ follows immediately from Eq. (4.23) and differentiability of F . By virtue of proposition A.17 in [BL19], F admits an absolutely continuous inverse on $(0, 1)$ if and only if μ is supported on an interval (finite or not) and the absolutely continuous component of μ has on that interval an almost-everywhere positive density (with respect to Lebesgue measure). We show that the latter statement holds. Indeed, since γ'' is non-vanishing, the Lebesgue differentiation theorem implies that the density function of the absolutely continuous component of the probability measure μ associated to F is positive almost everywhere and satisfies

$$f(s) = |\gamma''_{F(s)}|^{-1}, \quad \text{a.e.,} \quad (4.24)$$

implying the existence of an absolutely continuous inverse F^{-1} . This concludes (i). Regarding point (ii), as F is twice-differentiable its underlying measure μ is

absolutely continuous and f is its density which is positive and differentiable, implying the fulfilment of condition A . Condition B follows from [Remark 4.1.7](#). Finally, since $|\gamma''|$ is bounded from below (since it is continuous and non-vanishing) and F'' is bounded from above, it follows that conditions C and D are both satisfied. $\square$

The above result states that a twice-differentiable solution F prescribed by the integral equation [Eq. \(4.23\)](#) with the property that F'' is bounded, is the cumulative distribution function of a measure μ required for the application of [Corollary 4.1.6](#). We now determine the conditions on γ for which the existence of such an F is guaranteed. Let γ be as in [Lemma 4.1.8](#). Then the following observations can be made.

- 1) Suppose that the map $s \mapsto |\gamma_s''|^{-1}$ is Lipschitz on the interval $[0, 1]$, then the Picard-Lindelöf theorem gives existence and uniqueness to the initial value problem

$$F(t) = \int_0^t |\gamma_{F(s)}''|^{-1} ds, \quad F(0) = 0$$

as long as $F(t) \leq 1$. Since γ has bounded curvature, i.e. there exists a constant $c > 0$ such that $|\gamma_s''| \leq \frac{1}{c} \in \mathbb{R}_+$, it holds that

$$F(t) \geq \int_0^t c \, ds = ct$$

so that F hits 1 at some finite time $b > 0$.

- 2) To ensure that F is twice differentiable, requiring Lipschitz continuity of $s \mapsto |\gamma_s''|$ is not enough. Indeed, the latter assumption only implies the differentiability of F' *almost everywhere* on any open subset of the definition domain by virtue of the Rademacher's Theorem (which is not sufficient as it can lead to the breaking of condition A). However if we assume the map $s \mapsto |\gamma_s''|$ to be differentiable, then F is guaranteed to be twice-differentiable for all $t \in (0, b)$ by the quotient rule and non-vanishing property of γ'' .
- 3) As it is now assumed that $s \mapsto |\gamma_s''|^{-1}$ is differentiable, the fundamental theorem of calculus implies that the probability density function f can be exactly written as

$$f(s) = |\gamma_{F(s)}''|^{-1} \geq c_+, \quad s \in (0, b).$$

Additionally, since γ'' is non-vanishing and continuous, the curvature $|\gamma_s''|$ is bounded from below, i.e. $|\gamma_s''| \geq \frac{1}{c_-} \in \mathbb{R}_+$ for all $s \in [0, 1]$ (or equivalently

said, if the map $s \mapsto |\gamma_s''|^{-1}$ is bounded above by c_-), then we may conclude that

$$f(s) = |\gamma_{F(s)}''|^{-1} \leq c_-, \quad s \in (0, b). \quad (4.25)$$

Assuming further that the derivative of the map $s \mapsto |\gamma_s''|$ is bounded, this implies the boundedness of F'' by the chain and quotient rules.

These observations are collected and combined with [Lemma 4.1.8](#) in the following lemma.

Lemma 4.1.9 (Condition on γ for existence of admissible F). *Let $\gamma : [0, 1] \rightarrow V$ be a twice continuously differentiable unit-speed path such that the map $s \mapsto |\gamma_s''|$ is non-vanishing, and is differentiable with bounded derivative. Then there exists a constant $b \in \mathbb{R}$ and a twice-differentiable function $F : (0, b) \rightarrow [0, 1]$ such that the integral equation and boundary conditions [Eq. \(4.23\)](#) are satisfied as well as the conditions of [Corollary 4.1.6](#).*

Remark 4.1.10 (Boundedness of the derivative of $|\gamma_s''|$ as a condition on the total curvature). *Observe that the boundedness condition on $(|\gamma_s''|)'$ (which we recall is there to ensure that F'' is bounded) can be reformulated as a condition on the total curvature $T(t) := \int_0^t |\gamma_s''| ds$ of γ as follows. For a $c_2 \in \mathbb{R}_+$,*

$$c_2 \geq (|\gamma_s''|^{-1})' = -\frac{|\gamma_s''|'}{|\gamma_s''|^2} = -\left[\log |\gamma_s''|\right]' \cdot |\gamma_s''|^{-1},$$

which happens if and only if

$$|\gamma_s''| \geq -\frac{1}{c_2} \left[\log |\gamma_s''|\right]'.$$

Taking the integral from 0 to t gives

$$\frac{1}{c_2} \log \left(\frac{|\gamma_0''|}{|\gamma_t''|} \right) \leq T(t), \quad \text{for all } t \in [0, 1].$$

Finally, the direct application of [Lemma 4.1.9](#) followed by [Theorem 4.1.3](#) yields the following result.

Proposition 4.1.11 (Hambly-Lyons limit as the limit of a Wasserstein distance). *Let γ be as in [Lemma 4.1.9](#) and F the solution to the integral equation [Eq. \(4.23\)](#). Let μ be the measure associated with the cumulative distribution function F and let $X_1, \dots, X_n$ be a sample of n i.i.d. random variables drawn from μ . Let μ_n be its associated empirical*

measure, i.e.

$$\mu_n := \frac{1}{n} \sum_{i=1}^n \delta_{X_i}. \quad (4.26)$$

Then, the Hambly-Lyons limit in $c(\gamma)$ is the limit of a 2-Wasserstein distance

$$c(\gamma) = \lim_{n \rightarrow \infty} \mathbb{E} \left[\exp \left(-n W_2^2(\mu_n, \mu) \right) \right]^{1/2}. \quad (4.27)$$

4.2 Signature projections and Wasserstein distances

Due to the known connection between the Hambly-Lyons limit and the limit of the Hilbert-Schmidt tensor norm of projected signatures [HL10], the insights developed in the previous section naturally lead one to ask whether the Hilbert-Schmidt tensor norm of projected signatures can be related to Wasserstein distances. This section answers this question positively. By using this relationship we are able to characterise a class of curves, which are not necessarily C^3 as originally considered in [HL10], that satisfy

$$\lim_{n \rightarrow \infty} n! \|\mathcal{S}(\gamma)^n\| = c(\gamma). \quad (4.28)$$

We proceed as follows.

- 1) First, in [Section 4.2.1](#), we prove a technical augmentation of a lemma in [HL10] and then derive a probabilistic representation for the inner product of two signature terms;
- 2) We then proceed in [Section 4.2.2](#) to exploit the characterisation of the Wasserstein distances between empirical measures to relate the quantities derived in the first step to these Wasserstein distances and hence derive lower and upper bounds on $\|\mathcal{S}(\gamma)\|$ in terms of the former;
- 3) By leveraging the results of the previous section, we generalise the Hambly-Lyons limit [Eq. \(4.4\)](#) in [Section 4.2.4](#) and present the proof of [Theorem 4.0.3](#);
- 4) Finally, we show a practical way to compute the limit in [Section 4.2.5](#) and illustrate it in a simple case.

4.2.1 A probabilistic expression for signature inner products

In this subsection we generalise [HL10, Lemma 3.9] to the inner product between signatures before presenting a probabilistic formula in terms of the angles between the derivatives of the two underlying curves.

Definition 4.2.1 (Uniform order statistics sample). *Let $\{U_i\}_{i \in \{1, 2, \dots, n\}}$ and $\{V_i\}_{i \in \{1, 2, \dots, n\}}$ be two independent collections of n i.i.d. uniform random variables in $[0, 1]$. Consider the relabelling $\{U_{(i)}\}_{i \in \{1, 2, \dots, n\}}$ where*

$$0 \leq U_{(1)} \leq U_{(2)} \leq \dots \leq U_{(n)} \leq 1, \quad (4.29)$$

and similarly for $\{V_{(i)}\}_{i \in \{1, 2, \dots, n\}}$. In the rest of this chapter, we will denote by $\{U_{(i)}\}_{i \in \{1, 2, \dots, n\}}$ and $\{V_{(i)}\}_{i \in \{1, 2, \dots, n\}}$ two independent collections of i.i.d. uniform order statistics on $[0, 1]$.

Lemma 4.2.2 (Generalisation of Lemma 3.9 in [HL10]). *Let $\gamma, \sigma : [0, 1] \rightarrow \mathbb{R}^n$ be two C^1 unit-speed curves, and let $\{U_{(i)}\}_{i \in \{1, 2, \dots, n\}}$ and $\{V_{(i)}\}_{i \in \{1, 2, \dots, n\}}$ be two i.i.d. uniform order statistics collections (Definition 4.2.1). Then,*

$$\langle \mathcal{S}(\gamma)^n, \mathcal{S}(\sigma)^n \rangle = \frac{1}{(n!)^2} \mathbb{E} \left[\prod_{i=1}^n \langle \gamma'_{U_{(i)}}, \sigma'_{V_{(i)}} \rangle \right]. \quad (4.30)$$

Proof. It is known [HL10] that

$$n! \mathcal{S}(\gamma)^n = \mathbb{E} \left[\bigotimes_{i=1}^n \gamma'_{U_{(i)}} \right]. \quad (4.31)$$

Hence for an orthonormal basis $\{e_\rho\}_{\rho \in \{1, \dots, n\}}$ of $\mathbb{R}^n$, we have

$$(n!)^2 \langle \mathcal{S}(\gamma)^n, \mathcal{S}(\sigma)^n \rangle \stackrel{(4.31)}{=} \mathbb{E} \left[\prod_{i=1}^n \langle \gamma'_{U_{(i)}}, \sigma'_{V_{(i)}} \rangle \right]. \quad (4.32)$$

For more details, we refer the reader to the arguments of [HL10]. $\square$

This result states that the inner product between signatures of deterministic paths can be represented statistically through the mean of the product of $\langle \gamma'_{U_{(i)}}, \sigma'_{V_{(i)}} \rangle$. Observe that for unit-speed curves the inner products $\langle \gamma'_{U_{(i)}}, \sigma'_{V_{(i)}} \rangle$ encode only the information on the angles Θ_i between the vectors $\gamma'_{U_{(i)}}$ and $\sigma'_{V_{(i)}}$, i.e.

$$\langle \gamma'_{U_{(i)}}, \sigma'_{V_{(i)}} \rangle = \cos(\Theta_i) := \cos \left(\angle(\gamma'_{U_{(i)}}, \sigma'_{V_{(i)}}) \right). \quad (4.33)$$

In this case, also observe that the angles Θ_i can be exactly recovered from the norm of the difference between the two above random variables,

$$\Theta_i = \cos^{-1} \left(\langle \gamma'_{U(i)}, \sigma'_{V(i)} \rangle \right) = \cos^{-1} \left(1 - \frac{1}{2} \left| \gamma'_{U(i)} - \sigma'_{V(i)} \right|^2 \right). \quad (4.34)$$

Proposition 4.2.3 (Inner product as a probabilistic expression). *Suppose that $\gamma, \sigma : [0, 1] \rightarrow V$ are two absolutely continuous curves, such that $|\gamma'_t| = |\sigma'_t| = 1$ for almost every $t \in [0, 1]$. Then for every $n \in \mathbb{N}$ we have*

$$\begin{aligned} \langle \mathcal{S}(\gamma)^n, \mathcal{S}(\sigma)^n \rangle &= \frac{1}{(n!)^2} \mathbb{E} \left[\exp \left(- \sum_{k=1}^{\infty} \frac{1}{k 2^k} \sum_{i=1}^n \left| \gamma'_{U(i)} - \sigma'_{V(i)} \right|^{2k} \right) \mathbb{1}_{\{\max \Theta_i < \frac{\pi}{2}\}} \right] \\ &\quad + \frac{1}{(n!)^2} \mathbb{E} \left[\mathbb{1}_{\{\max \Theta_i \geq \frac{\pi}{2}\}} \prod_{i=1}^n \langle \gamma'_{U(i)}, \sigma'_{V(i)} \rangle \right], \end{aligned} \quad (4.35)$$

where Θ_i in $[0, \pi]$ is defined by $\langle \gamma'_{U(i)}, \sigma'_{V(i)} \rangle = \cos \Theta_i$ for $i = 1, \dots, n$, and $\mathbb{1}_A$ denotes the indicator function on a set A .

Proof. On the set $\{\max_{i=1, \dots, n} \Theta_i < \frac{\pi}{2}\}$, it holds that

$$\frac{1}{2} \left| \gamma'_{U(i)} - \sigma'_{V(i)} \right|^2 < 1, \quad \text{for } i = 1, \dots, n.$$

Consequently on this set, we have the expansion

$$\begin{aligned} \log \left(\prod_{i=1}^n \langle \gamma'_{U(i)}, \sigma'_{V(i)} \rangle \right) &= \sum_{i=1}^n \log \left(1 - \frac{1}{2} \left| \gamma'_{U(i)} - \sigma'_{V(i)} \right|^2 \right) \\ &= - \sum_{i=1}^n \sum_{k=1}^{\infty} \frac{1}{k} \left(\frac{1}{2} \left| \gamma'_{U(i)} - \sigma'_{V(i)} \right|^2 \right)^k \\ &= - \sum_{k=1}^{\infty} \frac{1}{k 2^k} \sum_{i=1}^n \left| \gamma'_{U(i)} - \sigma'_{V(i)} \right|^{2k}, \end{aligned}$$

where the second equality utilises the standard Talyor expansion for $\log(1 - x)$, which is valid for $|x| < 1$. Combining this expansion with [Lemma 4.2.2](#) yields [Eq. \(4.35\)](#). $\square$

The following two subsections leverage the preceding probabilistic expression to attain upper and lower bounds for $\|\mathcal{S}(\gamma)^n\|$. We first collate the recurrent objects and assumptions that will be used in several subsequent arguments.

Assumption 4.2.4 (Standing assumptions). *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. The following standing assumptions will be used throughout*

- i) *Let $\gamma : [0, 1] \rightarrow V$ be a twice continuously differentiable unit-speed path such that the map $s \mapsto |\gamma_s''|$ is both non-vanishing and differentiable with bounded derivative. Let $F : [0, b] \rightarrow [0, 1]$ be its associated cumulative distribution as prescribed in Lemma 4.1.8.*
- ii) *Let $\{U_{(i)}\}_{i \in \{1, \dots, n\}}$ and $\{V_{(i)}\}_{i \in \{1, \dots, n\}}$ be two i.i.d. collections of uniform order statistics (Definition 4.2.1). Define the collections $X_{(i)}$ and $Y_{(i)}$ as*

$$X_{(i)} := F^{-1}(U_{(i)}), \quad Y_{(i)} := F^{-1}(V_{(i)}), \quad i = 1, \dots, n. \quad (4.36)$$

- iii) *Let μ_n^X and μ_n^Y be their respective empirical distributions defined by*

$$\mu_n^X := \frac{1}{n} \sum_{i=1}^n \delta_{X_{(i)}}, \quad \mu_n^Y := \frac{1}{n} \sum_{i=1}^n \delta_{Y_{(i)}}. \quad (4.37)$$

- iv) *Let μ and f respectively be the probability measure and density distribution function associated with the cumulative distribution function F .*

To conclude this subsection we state the explicit characterisation of the p -Wasserstein distances in the case that μ and ν are both discrete measures.

Lemma 4.2.5 (Characterisation of the p -Wasserstein distance for empirical measures. Lemma 4.2 of [BL19]). *Let $\{A_j\}_{j \in \{1, \dots, n\}}$ and $\{B_j\}_{j \in \{1, \dots, n\}}$ be two samples of i.i.d. random variables. Denote by $\{A_{(j)}\}_{j \in \{1, \dots, n\}}$ and $\{B_{(j)}\}_{j \in \{1, \dots, n\}}$ their respective order statistics and let $\mu_n^A := \frac{1}{n} \sum_{i=1}^n \delta_{A_i}$ and $\mu_n^B = \frac{1}{n} \sum_{i=1}^n \delta_{B_i}$ be their associated empirical probability measures. Then for every $p \geq 1$ we have*

$$\sum_{j=1}^n |A_{(j)} - B_{(j)}|^p = n W_p^p(\mu_n^A, \mu_n^B),$$

where W_p is the p -Wasserstein distance. Moreover, for $p = \infty$, we have

$$\max_{j=1, \dots, n} |A_{(j)} - B_{(j)}| = W_\infty(\mu_n^A, \mu_n^B).$$

4.2.2 Lower bound on signature norms in terms of Wasserstein distances

We now use the probabilistic expression of the signature inner product ([Proposition 4.2.3](#)) when $\sigma = \gamma$ and derive a lower-bound for $\|S(\gamma)\|$ in terms of Wasserstein distances.

Proposition 4.2.6 (Lower bound on $\|S(\gamma)\|$ in terms of Wasserstein distances). *Let γ, μ_n^X, μ_n^Y be as defined in [Assumption 4.2.4](#). Then*

$$(n!)^2 \|S(\gamma)^n\|^2 \geq \mathbb{E} \left[\exp \left(-n \sum_{k=1}^{\infty} \frac{1}{k 2^k} W_{2k}^{2k}(\mu_n^X, \mu_n^Y) \right) \mathbb{1}_{\{\max_{i=1, \dots, n} |X_{(i)} - Y_{(i)}| < 1\}} \right] - \mathbb{P} \left(\max_{i=1, \dots, n} |X_{(i)} - Y_{(i)}| \geq 1 \right), \quad (4.38)$$

for every $n \in \mathbb{N}$.

Proof. We use the fact that

$$|\gamma'_{U_{(i)}} - \gamma'_{V_{(i)}}| = |g(X_{(i)}) - g(Y_{(i)})| \quad (4.39)$$

where $g := \gamma' \circ F$. The assumptions on γ give that g is once continuously differentiable and so the mean value inequality may be employed to see that

$$|\gamma'_{U_{(i)}} - \gamma'_{V_{(i)}}| \leq |X_{(i)} - Y_{(i)}| \cdot |g'(\xi_i)| \text{ for some } \xi_i \in [X_{(i)}, Y_{(i)}] \text{ for } i = 1, \dots, n. \quad (4.40)$$

Furthermore as F satisfies the integral equation of [Lemma 4.1.8](#), we have that

$$|g'(s)| = |\gamma''_{F(s)}| \cdot |F'(s)| = 1, \quad \forall s \in [0, b].$$

By applying [Lemma 4.2.5](#) we see that

$$\sum_{i=1}^n \left| \gamma'_{U_{(i)}} - \gamma'_{V_{(i)}} \right|^{2k} \leq \sum_{i=1}^n |X_{(i)} - Y_{(i)}|^{2k} = n W_{2k}^{2k}(\mu_n^X, \mu_n^Y). \quad (4.41)$$

Observe that $\max_{i=1, \dots, n} |X_{(i)} - Y_{(i)}| < 1$ is a strictly stronger condition than $\max_{i=1, \dots, n} \Theta_i < \frac{\pi}{2}$. Moreover, the product in the second term on the right-hand-side of [Eq. \(4.35\)](#) may be lower bounded by -1 by the Cauchy-Schwarz inequality. The lower bound [Eq. \(4.38\)](#) then follows by combining this observation, [Eq. \(4.35\)](#), and [Eq. \(4.41\)](#). $\square$

Remark 4.2.7. We can also multiply the sum inside the exponential term in Eq. (4.38) by the indicator function on the set $\{\max_{i=1,\dots,n} |X_{(i)} - Y_{(i)}| < 1\}$, without changing the random variable inside the expectation. Doing so will prove convenient in the proof of our main result [Theorem 4.0.3](#).

4.2.3 Upper bound on signature norms in terms of Wasserstein distances

Similarly to [Section 4.2.2](#), we use [Proposition 4.2.3](#) to derive an upper bound for $\|\mathcal{S}(\gamma)^n\|$ in terms of a series of Wasserstein distances.

Proposition 4.2.8 (Upper bound on $\|\mathcal{S}(\gamma)\|$ in terms of Wasserstein distances). *Let γ , μ_n^X , μ_n^Y be as defined in [Assumption 4.2.4](#). Then there exists an $0 < \varepsilon' \leq 1$ so that for every $n \in \mathbb{N}$ and $0 < \varepsilon < \varepsilon'$*

$$\begin{aligned} (n!)^2 \|\mathcal{S}(\gamma)^n\|^2 \leq & \mathbb{E} \left[\exp \left(-n \sum_{k=1}^{\infty} \frac{(1 - \phi(\varepsilon))^k}{k! 2^k} W_{2k}^{2k}(\mu_n^X, \mu_n^Y) \right) \mathbb{1}_{\{\max_{i=1,\dots,n} |X_{(i)} - Y_{(i)}| < \varepsilon\}} \right] \\ & + \mathbb{P} \left(\max_{i=1,\dots,n} |X_{(i)} - Y_{(i)}| \geq \varepsilon \right), \end{aligned} \quad (4.42)$$

where $\phi(u) := \sup_{|s-t|<u} |g'(t) - g'(s)|$ is the modulus of continuity of the derivative of $g := \gamma' \circ F$.

Proof. An application of the fundamental theorem of calculus and [Eq. \(4.39\)](#) gives

$$\begin{aligned} \left| \gamma'_{U_{(i)}} - \gamma'_{V_{(i)}} \right|^2 &= |g(X_{(i)}) - g(Y_{(i)})|^2 \\ &= \left\langle \int_{X_{(i)}}^{Y_{(i)}} g'(r) dr, \int_{X_{(i)}}^{Y_{(i)}} g'(s) ds \right\rangle \\ &= \int_{[X_{(i)}, Y_{(i)}]^2} \langle g'(r), g'(s) \rangle - |g'(r)|^2 + 1 dr ds \\ &= |X_{(i)} - Y_{(i)}|^2 + \int_{[X_{(i)}, Y_{(i)}]^2} \langle g'(r), g'(s) - g'(r) \rangle dr ds. \end{aligned}$$

And hence, by applying the Cauchy-Schwarz inequality to the integrand in the last line

$$\left| \gamma'_{U_{(i)}} - \gamma'_{V_{(i)}} \right|^2 - |X_{(i)} - Y_{(i)}|^2 \leq |X_{(i)} - Y_{(i)}|^2 \phi(|X_{(i)} - Y_{(i)}|), \quad (4.43)$$

and so in particular

$$\left| \gamma'_{U(i)} - \gamma'_{V(i)} \right|^2 \geq \left(1 - \phi \left(|X(i) - Y(i)| \right) \right) |X(i) - Y(i)|^2. \quad (4.44)$$

To ensure that the series inside the exponential in Eq. (4.42) is finite, we note that ϕ is a modulus of continuity, and so there exists some $0 < \varepsilon' \leq 1$ for which $\phi(\varepsilon) < 1$ for any $\varepsilon < \varepsilon'$. Now, the product in the second term on the right-hand side of Eq. (4.35) may be upper bounded by 1 by the Cauchy-Schwarz inequality, so it follows from Eq. (4.44), Lemma 4.2.5 and Proposition 4.2.3 that Eq. (4.42) holds provided that $\varepsilon < \varepsilon'$. $\square$

4.2.4 Generalising the Hambly-Lyons Limit theorem

Combined, Propositions 4.2.6 and 4.2.8 provide lower and upper bounds for $\|\mathcal{S}(\gamma)^n\|$ in terms of a series of Wasserstein distances. What remains is to show that the lower bound converges to the square of the Hambly-Lyons limit as $n \rightarrow \infty$, and that the same applies to the upper bound when taking $n \rightarrow \infty$ and then $\varepsilon \rightarrow 0$. The following pair of lemmas provide the necessary results for this conclusion.

Lemma 4.2.9. *Let μ , μ_n^X , and μ_n^Y be as in the standing assumptions, then for any $\varepsilon > 0$*

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\max_{i=1, \dots, n} |X(i) - Y(i)| \geq \varepsilon \right) = 0. \quad (4.45)$$

Proof. Using mean value and inverse function theorems, we deduce that

$$\begin{aligned} X(i) - Y(i) &= F^{-1}(U(i)) - F^{-1}(V(i)) \\ &= (F^{-1})'(\eta_i) (U(i) - V(i)) \\ &= |\gamma''_{\eta_i}| (U(i) - V(i)) \end{aligned} \quad (4.46)$$

for some $\eta_i \in [U(i), V(i)]$. An application of Markov's inequality gives that

$$\begin{aligned} \mathbb{P} \left(\max_{i=1, \dots, n} |X(i) - Y(i)| \geq \varepsilon \right) &\leq \varepsilon^{-1} \|\gamma''\|_{\infty} \mathbb{E} \left(\max_{i=1, \dots, n} |U(i) - V(i)| \right) \\ &\leq 2C\varepsilon^{-1} \|\gamma''\|_{\infty} \frac{1}{\sqrt{n}}, \end{aligned}$$

for some absolute constant $C > 0$. The first inequality utilises Eq. (4.46) and the second uses Lemma 4.2.5, the triangle inequality, and Theorem 4.9 of [BL19]. Taking the limit as $n \rightarrow \infty$ in the above inequality concludes the proof. $\square$

Lemma 4.2.10. *Let μ , μ_n^X , and μ_n^Y be as in the standing assumptions and $0 < a, \varepsilon \leq 1$, then*

$$\Delta_n := n \mathbb{1}_{\left\{\max_{i=1, \dots, n} |X_{(i)} - Y_{(i)}| < \varepsilon\right\}} \sum_{k=2}^{\infty} \frac{a^k}{k 2^k} W_{2k}^{2k}(\mu_n^X, \mu_n^Y) \rightarrow 0 \quad (4.47)$$

in L_1 as $n \rightarrow \infty$.

Proof. We have that

$$\begin{aligned} \mathbb{E} [\Delta_n] &= n \mathbb{E} \left[\mathbb{1}_{\left\{\max_{i=1, \dots, n} |X_{(i)} - Y_{(i)}| < \varepsilon\right\}} \sum_{k=2}^{\infty} \frac{a^k}{k 2^k} W_{2k}^{2k}(\mu_n^X, \mu_n^Y) \right] \\ &= \mathbb{E} \left[\mathbb{1}_{\left\{\max_{i=1, \dots, n} |X_{(i)} - Y_{(i)}| < \varepsilon\right\}} \sum_{k=2}^{\infty} \frac{a^k}{k 2^k} \sum_{i=1}^n |X_{(i)} - Y_{(i)}|^{2k} \right] \\ &\leq \mathbb{E} \left[\sum_{k=2}^{\infty} \frac{a^k}{k 2^k} \sum_{i=1}^n |X_{(i)} - Y_{(i)}|^4 \right] \\ &\leq \mathbb{E} \left[n W_4^4(\mu_n^X, \mu_n^Y) \sum_{k=1}^{\infty} \frac{a^k}{k 2^k} \right] \\ &= \mathbb{E} \left[n W_4^4(\mu_n^X, \mu_n^Y) \right] \log \left(\frac{2}{2-a} \right). \end{aligned}$$

Using the following inequality (Section 4.1, [BL19]) for $p \in \mathbb{N} \setminus \{0\}$

$$\mathbb{E} \left[W_p^p(\mu_n^X, \mu_n^Y) \right] \leq 2^p \mathbb{E} \left[W_p^p(\mu_n^X, \mu) \right]$$

together with the bounds from Theorem 5.3 in [BL19] gives

$$\mathbb{E} \left[W_4^4(\mu_n^X, \mu_n^Y) \right] \leq 2^4 \left(\frac{20}{\sqrt{n+2}} \right)^4 J_4(\mu),$$

where $J_4(\mu)$ is finite since μ is compactly supported with density f bounded below by Lemma 4.1.9. It follows that

$$\lim_{n \rightarrow \infty} \mathbb{E} [\Delta_n] \leq \lim_{n \rightarrow \infty} 40^4 \log \left(\frac{2}{2-a} \right) \frac{n}{(n+2)^2} J_4(\mu) = 0.$$

□

We are now ready to prove our main result.

Theorem 4.2.11 (Generalisation of the Hambly-Lyons Limit Theorem). *Let γ, μ, μ_n^X , and μ_n^Y be as in Assumption 4.2.4. Then,*

$$\lim_{n \rightarrow \infty} n! \|\mathcal{S}(\gamma)^n\| = \mathbb{E} \left[\exp \left(- \int_0^1 (B_{0,s})^2 |\gamma_s''|^2 ds \right) \right]^{1/2}. \quad (4.48)$$

Proof. For $0 < \varepsilon \leq 1$, define A_n^ε to be the set $\{\max_{i=1,\dots,n} |X_{(i)} - Y_{(i)}| < \varepsilon\}$. Then, by Lemma 4.2.9, $\mathbb{1}_{A_n^\varepsilon}$ converges to 1 in probability, while $\mathbb{P}((A_n^\varepsilon)^c)$ converges to 0. It follows from Corollaries 4.1.5 and 4.1.6 and Slutsky's theorem that

$$\frac{n}{2} \mathbb{1}_{A_n^\varepsilon} W_2^2(\mu_n^X, \mu_n^Y) \rightarrow \int_0^1 (B_{0,s})^2 |\gamma_s''|^2 ds \quad (4.49)$$

in distribution as $n \rightarrow \infty$. As such, by Eq. (4.49), Lemma 4.2.10 and another application of Slutsky's theorem we obtain the following convergence in distribution

$$n \mathbb{1}_{A_n^\varepsilon} \sum_{k=1}^{\infty} \frac{1}{k 2^k} W_{2k}^{2k}(\mu_n^X, \mu_n^Y) \rightarrow \int_0^1 (B_{0,s})^2 |\gamma_s''|^2 ds. \quad (4.50)$$

The combining of Eq. (4.50) with Proposition 4.2.6, Remark 4.2.7, Lemma 4.2.9, yet another application of Slutsky's theorem and the continuous mapping theorem, results in the lower bound

$$\lim_{n \rightarrow \infty} (n!)^2 \|\mathcal{S}(\gamma)^n\|^2 \geq \mathbb{E} \left[\exp \left(- \int_0^1 (B_{0,s})^2 |\gamma_s''|^2 ds \right) \right]. \quad (4.51)$$

Similar analysis for the upper bound from Proposition 4.2.8 shows that

$$\lim_{n \rightarrow \infty} (n!)^2 \|\mathcal{S}(\gamma)^n\|^2 \leq \mathbb{E} \left[\exp \left(-(1 - \phi(\varepsilon))^2 \int_0^1 (B_{0,s})^2 |\gamma_s''|^2 ds \right) \right] \quad (4.52)$$

for suitably small ε . Using the fact that ϕ is continuous at zero, we may take the limit in the preceding as $\varepsilon \rightarrow 0$ and combine it with the lower bound Eq. (4.51) to conclude that Eq. (4.48) holds. $\square$

This generalisation allows us to compute the limit of the signature norm for curves that are not C^3 .

Remark 4.2.12. *Due to our restrictions on γ , it is not possible to exhibit 2-dimensional examples satisfying our assumptions that are not C^3 .*

Example 4.2.13 (3-dimensional non- C^3 curves). Let $\alpha_t : [0, 1] \rightarrow \mathbb{R}$ be a continuous function which is non-differentiable at $t = \frac{1}{2}$. Define

$$\theta_t := \frac{\pi}{2} + \int_0^t \cos(\alpha_s) ds, \quad \varphi_t := \int_0^t \frac{\sin(\alpha_s)}{\sin(\theta_s)} ds.$$

By the Fundamental Theorem of Calculus, θ and φ are continuously differentiable but not twice-differentiable at $t = \frac{1}{2}$, provided α is chosen carefully. Note that φ is well defined, since $\frac{\pi}{2} - 1 \leq \theta_t \leq \frac{\pi}{2} + 1$ so that $\sin(\theta_t)$ is bounded away from 0. Construct the curve $\gamma' : [0, 1] \rightarrow \mathbb{R}^3$ using spherical coordinates

$$\gamma'_1(t) := \sin(\theta_t) \cos(\varphi_t), \quad \gamma'_2(t) := \sin(\theta_t) \sin(\varphi_t), \quad \gamma'_3(t) = \cos(\theta_t),$$

and set $\gamma(t) := \int_0^t \gamma'(s) ds$. Then by construction:

- 1) γ is twice continuously differentiable;
- 2) γ is unit speed and so of length 1;
- 3) $\gamma''_3 = -\theta'_t \sin(\theta_t)$ which is not differentiable at $t = \frac{1}{2}$;
- 4) $|\gamma''(t)|^2 = \theta'^2_t + (\varphi'_t \sin(\theta'_t))^2 = \cos^2(\alpha_s) + \sin^2(\alpha_s) = 1$, which is differentiable with bounded derivative.

4.2.5 Computing the Hambly-Lyons limit explicitly

Finally we propose a way to practically compute the limit presented in [Theorem 4.0.3](#). Using the work of Yor and Revuz on Bessel bridges [\[RY99\]](#), we relate the computation of the expected integral $c(\gamma)$ to the solving of a second order distributional differential equation.

Lemma 4.2.14 (High order term and curvature). Let V be a finite dimensional inner product space. Suppose that $\gamma : [0, l] \rightarrow V$ is parameterised at unit speed and is three-times continuously differentiable (under this parameterisation). Let μ denote the finite Borel measure on $\mathbb{R}_+$ which is absolutely continuous with respect to the Lebesgue measure $\lambda = \lambda_{\mathbb{R}_+}$ with density given by $\frac{d\mu}{d\lambda}(t) = \frac{1}{l^2} |\gamma''_t|^2 \mathbb{1}_{[0,1]}(t)$ a.e. Then,

- i) There exists a unique continuous function $\phi : [0, 1] \rightarrow \mathbb{R}$ which solves, in the distributional sense, the second-order differential equation

$$\psi'' = 2\psi\mu, \text{ with } \psi_0 = 0, \psi'_0 = 1. \quad (4.53)$$

ii) This solution satisfies

$$\psi_1^{-1/4} = \mathbb{E} \left[\exp \left(- \int_0^1 (B_{0,t})^2 \mu(dt) \right) \right]^{1/2} = c(\gamma). \quad (4.54)$$

Proof. We recall from [Remark 4.0.2](#) that $c(\gamma)$ is invariant under rescaling γ . As such, it is enough to prove [Eq. \(4.54\)](#) for unit-speed curves of unit length. Indeed, for the path $\tilde{\gamma} := \frac{1}{l}\gamma_{lt}$ defined on $[0, 1]$, we see that the density defining μ is given by

$$|\tilde{\gamma}_t''|^2 \mathbb{1}_{[0,1]}(t) = \frac{1}{l^2} |\gamma_{lt}''|^2 \mathbb{1}_{[0,1]}(t).$$

We therefore assume that $l = 1$ and seek to prove that $c(\gamma) = \psi_1^{-1/4}$. To this end, we first observe that the unit-speed parameterisation of γ gives that $\langle \gamma_t', \gamma_t''' \rangle = -|\gamma_t''|^2$ for every t in $[0, 1]$. When used together with [Eq. \(4.4\)](#) this gives that

$$c(\gamma) = \mathbb{E} \left[\exp \left(- \int_0^1 (B_{0,t})^2 \mu(dt) \right) \right]^{1/2}.$$

By noticing that $(B_{0,\cdot})^2$ is a squared Bessel bridge of dimension 1 which returns to zero at time 1, we can then use Theorem 3.2 of [\[RY99, Chapter 11\]](#) to see that

$$c(\gamma) = \left(\phi_1 \int_0^1 \frac{1}{\phi_t^2} dt \right)^{-1/4} \quad (4.55)$$

where ϕ is the unique continuous positive function solving the distributional equation $\phi'' = 2\phi\mu$, with $\phi_0 = 1$, $\phi'_0 = 0$. Using Exercise 1.34 in [\[RY99, Chapter 11\]](#), we know that the function ψ in the statement and ϕ are related by

$$\psi_t = \phi_t \int_0^t \frac{1}{\phi_s^2} ds$$

and therefore [Eq. \(4.55\)](#) becomes $c(\gamma) = \psi_1^{-1/4}$. □

This result reduces the computation of $c(\gamma)$ for a planar circle to the solving of a simple differential equation.

Example 4.2.15 (The planar circle). Let $\gamma_t = \frac{1}{2\pi} (\sin(2\pi t), \cos(2\pi t))$ for t in $[0, 1]$, then γ is a smooth unit-speed curve of length $l = 1$ with $|\gamma_t''| = 2\pi$ so that $\mu(dt) = 2\pi \mathbb{1}_{[0,1]}(t) dt$.

By solving the differential equation [Eq. \(4.53\)](#) we find that

$$c(\gamma) = \left(\frac{2\sqrt{\pi}}{\sinh 2\sqrt{\pi}} \right)^{1/4}. \quad (4.56)$$

Remark 4.2.16 (Role of $c(\gamma)$ in the asymptotics of the norm). By starting from [Eq. \(4.3\)](#) and [Eq. \(4.4\)](#) we notice the latter statement can be rewritten as a statement of asymptotic equivalence, namely

$$\left(n! \|\mathcal{S}(\gamma)^n\|_{HS} \right)^{1/n} \sim l + \frac{l}{n} \log c(\gamma) + o\left(\frac{1}{n}\right) \text{ as } n \rightarrow \infty, \quad (4.57)$$

where we write $\|\cdot\|_{HS}$ for the moment to emphasise the dependence of this expansion on the choice of tensor norm. By contrast for the projective tensor norm, it follows from [Eq. \(4.2\)](#) that we have

$$\left(n! \|\mathcal{S}(\gamma)^n\|_{proj} \right)^{1/n} \sim l + o\left(\frac{1}{n}\right) \text{ as } n \rightarrow \infty.$$

When written in this way, the $c(\gamma)$ limit has the interpretation of being the second term in the asymptotic expansion of $(n! \|\mathcal{S}(\gamma)^n\|_{HS})^{1/n}$ as $n \rightarrow \infty$. Natural questions would be to explore the higher order terms in these asymptotic expansions, and relate them to geometric features of the underlying curve.

CHAPTER 5

GENERATIVE MODELLING OF LÉVY AREA

The final chapter of this thesis is concerned with constructing weak estimators for the Lévy area of a multi-dimensional Brownian motion. Lévy area – used in high order numerical schemes for SDEs – is the anti-symmetric component of the second level of the signature of Brownian motion. As a random variable, its distribution is non-Gaussian and difficult to approximate. Traditional approaches involve either truncated Fourier expansions of Brownian motion or designing custom random variables whose low order moments match those of Lévy area. We adopt an alternative approach: generative modelling, which uses a custom discriminator that leverages symmetries of the distribution of Lévy area. The results presented this chapter are found in [Jel+23].

The numerical simulation of stochastic differential equations (SDEs) is a ubiquitous task encountered in a wide variety of fields, ranging from mathematical finance [Shr04] and systems biology [Bro+20] to molecular dynamics [LM15] and data science [Li+19]. Real-world phenomena arising in these areas are often described well by SDEs formulated through *Itô calculus* and of the general form:

$$dX_t = f(X_t)dt + \sum_{i=1}^d g_i(X_t)dW_t^{(i)}, \quad X_0 = x_0, \quad (5.1)$$

where the solution $X = \{X_t\}_{t \in [0, T]}$ takes values in $\mathbb{R}^e$, $W = (W^{(1)}, \dots, W^{(d)})$ denotes a standard d -dimensional Brownian motion and $f, g_i : \mathbb{R}^e \rightarrow \mathbb{R}^e$ are suitably regular vector fields on $\mathbb{R}^e$. In practice, one is often concerned with estimating quantities of the form $\bar{\varphi} := \mathbb{E} [\varphi(X) \mid X_0 = x_0]$, where the function φ may depend on the whole trajectory of X or simply on the terminal value X_T ; such as the payoff of a European

call-option. On occasion, it may be possible to obtain $\bar{\varphi}$ by solving certain PDEs (e.g. through the backward Kolmogorov equation or the Feynman-Kac formula).

However, the standard approach is to use Monte Carlo simulation where one uses a discretisation scheme to generate approximate sample paths $\{\hat{X}_i\}_{i=1}^N$ of X , which can be used to approximate $\bar{\varphi}$ by taking the average of $\{\varphi(\hat{X}_i)\}_{i=1}^N$. Given the importance of Monte Carlo simulation in applications, there is a rich literature concerning numerical methods for SDEs and their theoretical properties. A broad range of discretisation schemes are available, such as the classical Euler-Maruyama and Milstein schemes as well as the higher order Talay [Tal90] and Ninomiya-Victoir [NV08] schemes. For more details on the numerical simulation of SDEs, we refer the reader to [KP92] and [PT85].

There are two standard measures for the effectiveness of numerical schemes: strong error (or MSE) and weak error. It is a well-known result of Clark and Cameron [CC80] that numerical schemes using only increments of Brownian motion are limited in general to a strong convergence rate of at most $O(\sqrt{h})$, where h denotes the step size. Furthermore, to the best of our knowledge, all numerical schemes achieving second order weak convergence require the generation of random variables in addition to the Brownian increments. In particular, the Talay scheme [Tal90] and the Ninomiya-Victoir scheme [NV08] require the generation of Rademacher random variables. Further examples include the Ninomiya-Ninomiya scheme [NN09] and a stochastic Runge-Kutta method due to Rößler [Röß09]. In all of these cases, the additional random variable(s) is (are) generated to replace certain second order iterated integrals of Brownian motion which are commonly referred to as its *Lévy area*.

Definition 5.0.1. *The Lévy area of a d -dimensional Brownian motion over $[s, t]$ is a $d \times d$ antisymmetric matrix whose (i, j) -th entry is given by*

$$A_{s,t}^{(i,j)} := \frac{1}{2} \left(\int_s^t (W_r^{(i)} - W_s^{(i)}) dW_r^{(j)} - \int_s^t (W_r^{(j)} - W_s^{(j)}) dW_r^{(i)} \right).$$

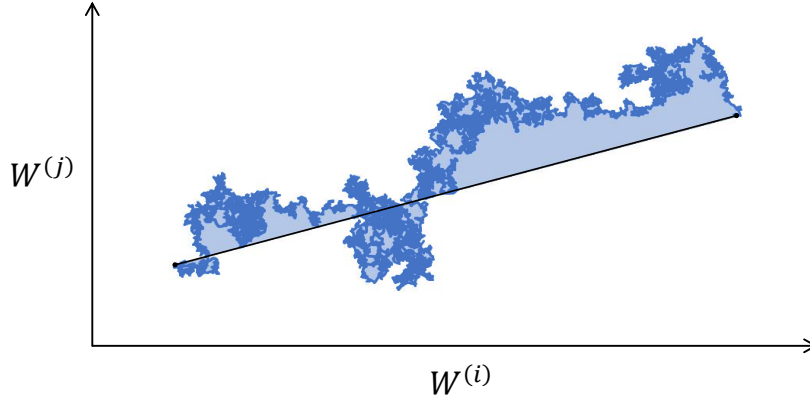


Figure 5.1: Each entry $A^{(i,j)}$ is the signed area between the independent Brownian motions $W^{(i)}$ and $W^{(j)}$ (diagram adapted with permission from [Fos20]).

When the vector fields of the SDE Eq. (5.1) do not satisfy the commutativity condition $[g_i, g_j] = 0$ (where $[g_i, g_j](x) := g'_j(x)g_i(x) - g'_i(x)g_j(x)$ denotes the standard Lie bracket of vector fields), then schemes that achieve high order Strong convergence (such as the Milstein and log-ODE methods) require the simulation of Lévy area. Consequently, the approximation of Lévy area has received much interest in recent decades, with the view towards both high order weak and strong convergence.

Approximations to Lévy area have been well studied [Dav14; Dic07; GL94; KP92; KPW92; Fos20; FH23; MR22; Wik01], with the majority of approximations concerning strong estimation. Strong estimators aim to approximate Lévy area by minimising the mean-squared error so that the resulting estimator may be incorporated into the strong analysis of the discretisation scheme. Typically such estimators rely on truncated expansions of Brownian motion, such as the Fourier series expansion [KP92; KPW92; KR23], the Karhunen-Loève expansion [Loè78] and more recently the polynomial expansion [FH23; FLO20]. These estimators are often improved by estimating the tail sum of the expansion in an appropriate manner, see for example [MR22; Wik01].

To the best of our knowledge there is no known scheme which simulates Lévy area exactly, with even the “rectangle-wedge-tail” algorithm of Gaines and Lyons [GL94] requiring numerical integration. Moreover, this approach is only applicable in $d = 2$. On the other hand, the main drawback of the truncated expansion methods is the cost of simulation. In practice one is often required to generate millions of Lévy area samples, and the aforementioned methods typically require a high truncation level to achieve good performance. Consequently there has been a renewed focus in recent years on approximations of Lévy area that are suitable for weak discretisation schemes, where the estimator is less costly to generate. The aim of a weak estimator is to match some moments of the Lévy area given a

Brownian increment. These estimators differ in complexity depending on their intended usage. Basic estimators include the Rademacher random variables that appear in the Talay scheme [Tal90] and Davie’s approximation [Dav14; FL15] which uses a Gaussian random variable with the correct variance (this may be improved to have the correct conditional variance given the Brownian increment [Fos20]). Perhaps the most sophisticated weak approximation is due to Foster [Fos20], which matches the first five conditional moments of Lévy area given the Brownian increment when $d \leq 3$.

In this chapter we provide a new approach to the construction of weak estimators of Lévy area given a Brownian increment, through the use of generative modelling techniques. To the best of our knowledge, this is the first time that the powerful toolkits provided by modern machine learning have been applied to the problem of Lévy area simulation. Arguably the main obstacle to this approach is the computation effort required to generate considerable amounts of precise samples of Lévy area. We note this is possible in principle, through the use of a truncated Fourier series method [KR23], with other options also possible. However in the context of Lévy area generation, we present a novel training algorithm based on Chen’s relation [Che57] that allows for the training of a generative model without access to a dataset of Lévy area samples.

5.0.1 Contributions

In this chapter we present LévyGAN, a deep-learning based generative model that simulates the Lévy area of arbitrary dimensional Brownian motions. Deep-learning based generative models have been widely used for data synthesis where a parametric model is trained to learn the target distribution. Among the variety of generative models, GAN-type models [MO14; Goo+14] have been particularly noteworthy for their performances. GAN, short for Generative Adversarial Network, operates on the compelling principle of adversarial training. This approach consists of two neural networks - a generator and a discriminator - that are trained simultaneously. The generator’s task is to create synthetic data, while the discriminator’s role is to distinguish between real and generated data.

With no exception, like other generative models, GAN-type models also possess drawbacks. The necessity of real data as reference sets for training is one. Machine learning models are often data-driven and sometimes data-greedy; normally, practitioners collect real-world data and approximate its distribution using an empirical distribution output by the generator. This is especially pertinent in the context of Lévy area generation, where a competitive method needs to achieve

very high accuracy, incurring a high statistical complexity, thusly requiring large amounts of data. In contrast to standard GANs, score-based diffusion models and variational autoencoders, LévyGAN is designed to learn the target distribution without requiring any samples from it at all. We term this approach “Chen-training”, which is theoretically underpinned by the unique invariance of the joint law of Brownian motion and Lévy area under Chen’s relation, see [Theorem 5.3.5](#).

We have designed both the generator and discriminator by exploring the features of the joint law of Brownian motion and its Lévy area [[FH23](#); [FLO20](#); [Fos20](#)]. In particular, for the generator we ensure that the joint distribution of our generator is permutation invariant and that each component of the generated area depends only on the relevant components of the Brownian increment. We also ensure that all odd cross moments of our generator are exact through the multiplication by certain Rademacher random variables. For the discriminator, we have chosen a characteristic function based discriminator initially proposed in [[FSS20](#); [Li+20a](#)]. Inspired by this approach and the more general method of [[LLN23](#)], we define the unitary characteristic function of a random variable as a generalization of the characteristic function into higher degree Lie algebras.

For our numerical results, we train the LévyGAN in $d = 4$. It is noteworthy that the model is able to generate the Lévy area for arbitrary Brownian dimensions, with no loss in performance guaranteed in $d' < d$ and empirically strong performance for $d' > d$. We show empirically that LévyGAN attained the best performance among other weak estimators (such as those found in [[Dav14](#)] and [[Fos20](#)]) in terms of distributional metrics and is comparable in generation speed to the method in [[Fos20](#)]. Finally, we provide an application of weak approximations of Lévy area to high order multi-level Monte Carlo numerical schemes. In this example, we demonstrate that the inclusion of an approximate Lévy area term in the Strang splitting method achieves higher order variance reduction and weak convergence. Moreover, we provide evidence that an estimator that only matches the variance of Lévy area (such as the Rademacher random variables found in the Talay scheme) is not appropriate for this application. The LévyGAN implementation, together with a trained model for $d = 4$ can be found at <https://github.com/andyElking/LevyGAN>.

5.0.2 Outline

This chapter is divided into five main sections. In [Section 5.1](#) we recall the standard setup of generative adversarial networks and discuss why the traditional approach to generative modelling is not easily applicable to our setting. In [Section 5.2](#) we outline the structure of our generator. This section focuses on the symmetries

of the joint law of Brownian motion and Lévy area that we hard code into our generator. This includes so-called “bridge-flipping”, a precise multiplication by certain Rademacher random variables to ensure all joint odd moments are correctly estimated and to help the generator train evenly across all quadrants in space. In [Section 5.2.2](#) we introduce a network architecture dubbed “pair-net”, inspired by graph neural networks that ensures the correct dependence structure between the coordinates of Brownian motion and the coordinates of Lévy area. The structure of our discriminator is outlined in [Section 5.3](#). Here we discuss two alternatives for a loss function based on the analytical form of the joint characteristic function of Lévy area and Brownian motion, and a generalisation termed the unitary characteristic function proposed in [\[LLN23\]](#). Our novel training approach “Chen-training” is covered in [Section 5.3.2](#), before the whole training procedure is summarised in [Section 5.4](#). Finally, in [Section 5.5](#), we compare the distributional performance of our generator in comparison to the state-of-the-art Foster method [\[Fos20\]](#) and demonstrate the applicability of weak estimators for Lévy area to multilevel Monte-Carlo.

To conclude the introduction, we outline some common notation to be used throughout the chapter.

- 1) $\mathbb{P}_{X|Y=y}$ for the conditional distribution of X given $Y = y$.
- 2) With abuse of notation $(X | Y = y) \stackrel{d}{=} (Z | Y = y)$ if $\mathbb{P}_{X|Y=y} = \mathbb{P}_{Z|Y=y}$.
- 3) $(W_t)_{t \in [0,1]}$ for a d -dimensional Brownian motion with coordinates $\{W_t^{(i)}\}_{1 \leq i \leq d}$ and the process $(A_t)_{t \in [0,1]}$ with $A_t = \{A_t^{(i,j)}\}_{1 \leq i < j \leq d}$ a $\frac{d(d-1)}{2}$ -dimensional vector representing the flattened upper triangle of the Lévy area matrix associated with the Brownian motion. Unless stated otherwise, we denote by a the dimension of Lévy area, i.e. $a = \frac{d(d-1)}{2}$.
- 4) For any process $(X_t)_{t \in [0,1]}$, X_t denotes the process evaluated at time t .
- 5) $\mathbb{P}_{(W_t, A_t)}$ for the joint law of Brownian motion and Lévy area at time t .
- 6) $\mathcal{N}^d(\mu, \sigma^2)$ for a d -dimensional Gaussian distribution with independent coordinates, each with mean μ and variance σ^2 .
- 7) $\text{Rad}^d(p)$ for a d -dimensional Rademacher random variable with independent entries each of which takes 1 and -1 with probability p and $1 - p$ respectively.
- 8) Φ_X for the characteristic function of a random variable X .
- 9) $W_{s,t} := W_t - W_s$ for the increments of a Brownian motion $(W_t)_{t \in [0,1]}$, we will use $W_t = W_{0,t}$ interchangeably.

- 10) For any *tensor*, i.e. an element $\mathbf{x} \in \mathbb{R}^{\cdot \times d}$, we always denote by $x^{(i)} \in \mathbb{R}$ the i^{th} coordinate of the second dimension.

5.1 The GAN architecture

The goal of this chapter is to build an efficient and accurate estimator that approximates the conditional law $\mathbb{P}_{A_t|W_t}$, and hence the joint law, $\mathbb{P}_{(W_t, A_t)}$, of a Brownian increment and its Lévy area. Thanks to the scaling property of Brownian motion, it is enough to consider the problem when $t = 1$.

We adopt a conditional GAN (Generative Adversarial Network) approach as proposed by [MO14]. GAN-typed models, initially proposed by [Goo+14] consist of a pair of competing neural networks - the generator and the discriminator. The aim of the generator is to create “fake” data $\tilde{\mathbf{x}}$ from some noise distribution that tries to mimic a target distribution. On the other hand, the discriminator will be given both $\tilde{\mathbf{x}}$ and data $\mathbf{x}$ from the true distribution and will try to distinguish the ground truth from the fake data. The dynamics between the generator and discriminator are controlled by a min-max game acting on a loss function, which usually represents the distance between two distributions. In a conditional GAN, the generator is given not only samples from the noise distribution but also the conditioning variable – in our case that will be the Brownian increment W – or later, the space-time Lévy area H (see Section 5.2.1). Here we provide the description of a classical conditional GAN adapted to our interest.

Definition 5.1.1 (Classical conditional GAN for Lévy area generation). *Let $d \geq 2$ be the Brownian dimension, and $a := \frac{d(d-1)}{2}$ be the dimension of the associated Lévy area vector. Assume z is an n -dimensional noise vector distributed according to $\mathbb{P}_z$. The conditional generator G_θ and the discriminator D_η are maps*

$$G_\theta : \mathbb{R}^d \times \mathbb{R}^n \rightarrow \mathbb{R}^d \times \mathbb{R}^a \quad D_\eta : \mathbb{R}^d \times \mathbb{R}^a \rightarrow \mathbb{R}$$

that are parametrized by θ and η respectively. When restricted to the first d -coordinates of the domain and co-domain, G_θ is enforced to be the identity, i.e. $G_\theta(w, z) = (w, \tilde{A})$. Let $\mathbb{P}_{(W_1, A_1)}$ be the ground truth distribution of the coupled process. An example loss function might be given by

$$L(\theta, \eta) := \mathbb{E}_{(w, A) \sim \mathbb{P}_{(W_1, A_1)}} [D_\eta(w, A)] - \mathbb{E}_{w \sim \mathbb{P}_{W_1}, z \sim \mathbb{P}_z} [D_\eta(G_\theta(w, z))],$$

where one restricts D_η to be at most 1-Lipschitz. The models are trained using the min-max game $\min_\theta \max_\eta L(\theta, \eta)$ until convergence. The generator obtained is then used to simulate Lévy areas.

Since there are no known methods for exact simulation of Lévy areas in $d > 2$, the “true” Lévy area samples $\mathbf{A} \sim \mathbb{P}_{A_1|W_1}$ must themselves be obtained through approximate simulation. In particular we might generate the “true” Lévy area samples using a Julia-language package created by [KR23], which complements their paper about approximate strong simulation of Lévy area through truncated Fourier series methods. We also present the flowchart of this methodology in Fig. 5.2.

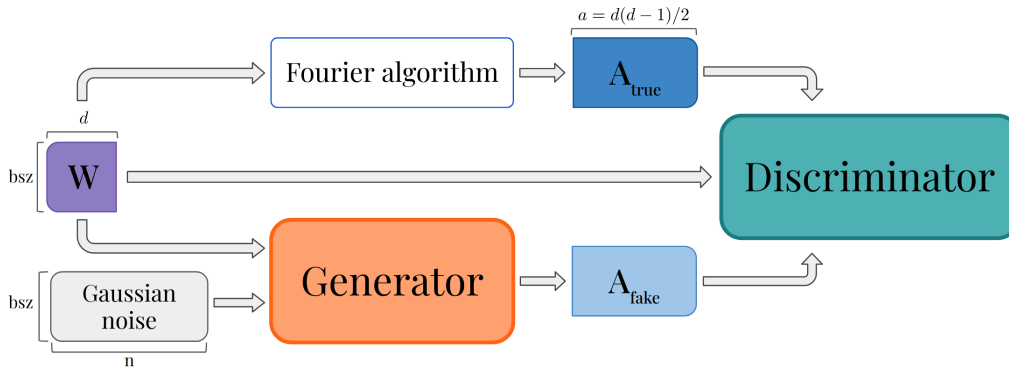


Figure 5.2: A schematic of Lévy generation for classical conditional GAN. Throughout this article, bsz represents the training batch-size.

The main drawback of this methodology is the need for simulating approximate real samples of Lévy area. Using real data not only slows the training procedure, but also introduces simulation error (due to the truncation in the Fourier series) and finite sample error (leading to over-fitting). In order to address these problems, we propose a novel approach, LévyGAN, which completely excludes real samples from the training process and is well justified by our two main theoretical contributions [Theorem 5.3.5](#) and [Theorem 5.3.8](#).

5.2 Generator

In order to improve the accuracy of the generated distribution, we can consider the symmetries of the Lévy area distribution and hard-code them into the generator itself. This way, the generator will consist of both a neural net and additional operations applied to the network’s output. One symmetry of Lévy area that is desirable to reflect is the fact that its distribution is mean zero when conditioned

on any increment $W_{0,1} = w$. We do this through two operations, which combined we term “bridge-flipping”.

5.2.1 Bridge-Flipping

As far as possible, we would like to enforce any symmetries of the joint distribution of Brownian motion and Lévy area. For example, it would be beneficial to hard-code the symmetry of Lévy area about zero into the generator’s architecture. Even though each dimension of Lévy area $A^{(i,j)}$ is symmetric about zero, their joint distribution is not invariant to multiplying any individual dimension by -1 . However, the dimensions of the underlying Brownian motion are independent, and hence each can be mirrored separately without violating their joint law.

The goal is then to find a symmetry of Lévy area corresponding to independently “flipping” individual dimensions of the Brownian motion. Notice, however, that we are trying to generate Lévy area conditional on a fixed input $W_{0,1} = w$, and so we do not wish to flip the increment of Brownian motion itself. We can circumvent this issue by considering the polynomial expansion of Lévy area [FH23]. This decomposes the Brownian motion into the components dependent on w and components independent of w , the latter of which can then be flipped independently. To this end, we first define the Brownian bridge and its accompanying “space-time” and “space-space” Lévy areas.

Definition 5.2.1 (Brownian bridge). *Let $0 \leq s \leq t < \infty$. Then the Brownian bridge of W on $[s, t]$ returning to zero at time t is defined as*

$$B_{s,u} := W_{s,u} - \frac{u-s}{t-s} W_{s,t} \quad \text{for } u \in [s, t].$$

The “space-time” Lévy area $H_{s,t} \in \mathbb{R}^d$ and “space-space” Lévy area $b_{s,t} \in \mathbb{R}^{d \times d}$ of B over $[s, t]$ are

$$H_{s,t}^{(i)} := \frac{1}{t-s} \int_s^t B_{s,u}^{(i)} du = \frac{1}{t-s} \int_s^t W_{s,u}^{(i)} - \frac{u-s}{t-s} W_{s,t}^{(i)} du \quad \text{for } 1 \leq i \leq d, \quad (5.2)$$

$$b_{s,t}^{(i,j)} := \int_s^t B_{s,u}^{(i)} \circ dB_u^{(j)} \quad \text{for } 1 \leq i, j \leq d. \quad (5.3)$$

Whenever $s = t$, we define $b_{s,t} = H_{s,t} = 0$. We write H_t and b_t for $H_{0,t}$ and $b_{0,t}$ respectively.

It turns out that (H, b) and $W_{s,t}$ are independent, the marginal distribution of H is Gaussian, and the marginal of b is logistic.

Proposition 5.2.2 (Distribution of Brownian bridge Lévy area [FLO20; Fos20]). *For fixed $0 \leq s < t < \infty$, the process $\{(H_{s,u}, b_{s,u})\}_{u \in [s,t]}$ and the increment $W_{s,t}$ are independent. Furthermore, H is distributed as a d -dimensional Gaussian with independent coordinates and the marginal distribution of each Brownian bridge Lévy area is logistic:*

$$H_{s,t} \sim \mathcal{N}^d \left(0, \frac{1}{12}(t-s) \right) \quad \text{and} \quad b_{s,t}^{(i,j)} \sim \text{Logistic} \left(0, \frac{1}{2\pi}(t-s) \right).$$

This yields the first two terms of the polynomial expansion of Lévy area.

Proposition 5.2.3 (Polynomial expansion of Lévy area [FH23]). *The Lévy area of a d -dimensional Brownian motion W has the following decomposition:*

$$A_{s,t} = H_{s,t} \otimes W_{s,t} - W_{s,t} \otimes H_{s,t} + b_{s,t},$$

where $\otimes$ denotes the outer-product of vectors.

This decomposition reduces our conditional generative task to the estimation of the Brownian bridge Lévy area conditional on a Brownian increment. In this case, the target distribution and conditioning variable are independent. Indeed since B and the increment $W_{0,1} = w$ are independent, the conditional distribution $(B_t \mid W_{0,1} = w)$ is symmetric around 0. Furthermore, each dimension of B is its own independent process, so we can flip each individually without affecting the distribution (as established in the following lemma). This approach may be generalised to the estimation of the tail sum of the polynomial expansion of Lévy area truncated at a higher level.

Lemma 5.2.4. *Let W be a d -dimensional Brownian motion on $[0, 1]$ and let B, H, b be the corresponding derived processes from Definition 5.2.1. Fix some $\xi \in \{-1, 1\}^d$, and let H' , and b' be the space-time and space-space Lévy area processes associated with the process $\xi \odot B = \{\xi \odot B_t\}_{t \in [0,t]}$ where $\odot$ denotes the Hadamard (coordinate-wise) product. Then*

$$\{B_t\}_{t \in [0,1]} \stackrel{d}{=} \{\xi \odot B_t\}_{t \in [0,1]}, \quad H' = \xi \odot H, \quad b' = (\xi \otimes \xi) \odot b, \quad \text{and} \quad (H', b') \stackrel{d}{=} (H, b).$$

Proof. See Lemma B.2.3. □

We also include a multiplication final value $\xi_0 \in \{-1, 1\}$, whose role is explained in Proposition 5.2.7. Combining this with Lemma 5.2.4 we obtain the “bridge-flipping” function.

Definition 5.2.5. Let $w, h, \xi \in \mathbb{R}^d$, $b \in \mathbb{R}^{d \times d}$ and $\xi_0 \in \mathbb{R}$. Then the Bridge-flipping function is defined as

$$\text{BF}(w, h, b, \xi_0, \xi) := \xi_0 ((\xi \odot h) \otimes w - w \otimes (\xi \odot h) + (\xi \otimes \xi) \odot b). \quad (5.4)$$

For this function we have the following as a consequence of [Proposition 5.2.3](#) and [Lemma 5.2.4](#).

Theorem 5.2.6 (Bridge-flipping). Let $\xi_0, \dots, \xi_d \stackrel{i.i.d.}{\sim} \text{Rad}(\frac{1}{2})$ be random variables so that $W, (H, b)$, and $(\xi_0, \dots, \xi_d)$ are independent. Write $\xi = (\xi_1, \dots, \xi_d)$ and fix some $w \in \mathbb{R}^d$. Let H, b be as in [Definition 5.2.1](#). Then for every $t \in [0, 1]$

$$(A_{0,t} \mid W_{0,1} = w) \stackrel{d}{=} (\text{BF}(W_{0,t}, H_{0,t}, b_{0,t}, \xi_0, \xi) \mid W_{0,1} = w).$$

For our purpose, we will utilise the result for $t = 1$:

$$(A_{0,1} \mid W_{0,1} = w) \stackrel{d}{=} \text{BF}(w, H_{0,1}, b_{0,1}, \xi_0, \xi).$$

Recall that $H_{s,t} \sim \mathcal{N}^d(0, \frac{1}{12}(t-s))$ and that b and H are correlated, but that (H, b) and $W_{0,1}$ are independent. Hence, given a neural net $\text{NN}_\theta : \mathbb{R}^{d+n} \rightarrow \mathbb{R}^d$, we define the “Bridge-flipping generation”.

Algorithm 1 Area generation using Bridge-flipping

Input: θ - neural net parameters, d - Brownian dimension, w - Brownian increment, n - dimension of noise vector.

- 1: $\xi_0 \leftarrow \text{Rad}(1/2)$; $\xi \leftarrow \text{Rad}^d(1/2)$
 - 2: $h \leftarrow \mathcal{N}^d(0, \frac{1}{12}(t-s))$; $z \leftarrow \mathcal{N}^n(0, 1)$
 - 3: $\tilde{b} \leftarrow \text{NN}_\theta(h, z)$ $\triangleright$ we want $\text{NN}_\theta(h, z) \stackrel{d}{\approx} (b_{0,1} \mid H_1 = h)$
 - 4: **return** $\text{BF}(w, h, \tilde{b}, \xi_0, \xi)$
-

This construction has some desirable properties.

- 1) Informally speaking, the use of ξ effectively makes the generator behave identically on all orthants of $\mathbb{R}^d$ and hence any learning done in one orthant transfers equally to the other orthants. This speeds up training and significantly improves the generator’s accuracy, as it now perfectly mimics the symmetric structure of Lévy area.
- 2) The neural net can be trained directly on the distribution of $(b_{0,1} \mid H_{0,1} = h)$, and is then used for generation of $(A_{0,1} \mid W_{0,1} = w)$ using the BF algorithm.

- 3) The structure of the BF algorithm allows for efficient implementation of back-propagation.

The use of the extra Rademacher random variable ξ_0 is to guarantee that all odd joint and conditional moments are correctly matched. This is summarised in the following proposition. We denote by $\mathbb{P}_{\text{BF}}^{\theta,w}$ the output distribution of the BF generator given inputs w and θ , and the joint distribution of Brownian increments and “fake” area defined by the generator is denoted by $\mathbb{P}_{\text{BF}}^\theta$.

Proposition 5.2.7. *The following facts about the joint and conditional distributions of Brownian motion, Lévy area, and the BF generator hold.*

- 1) If $r \sim \text{Rad}(\frac{1}{2})$ then $(rW_{0,t}, rA_{0,t}) \stackrel{d}{=} (W_{0,t}, rA_{0,t}) \stackrel{d}{=} (W_{0,t}, A_{0,t})$.
- 2) For any $n_i, n_{ij}, m_{ij} \in \mathbb{N}$, $1 \leq i < j \leq d$ it holds that

$$\mathbb{E}_{(W_{0,1}, A_{0,1}) \sim \mathbb{P}_{(W_{0,1}, A_{0,1})}} \left[\prod_{i=1}^d (W_{0,1}^{(i)})^{n_i} \left(\prod_{1 \leq i < j \leq d} (A_{0,1}^{(i,j)})^{n_{ij}} \right) \right] = 0, \quad (5.5)$$

$$\mathbb{E}_{(W_{0,1}, \tilde{A}) \sim \mathbb{P}_{\text{BF}}^\theta} \left[\prod_{i=1}^d (W_{0,1}^{(i)})^{n_i} \left(\prod_{1 \leq i < j \leq d} (\tilde{A}^{(i,j)})^{n_{ij}} \right) \right] = 0, \quad (5.6)$$

$$\mathbb{E}_{A_{0,1} \sim \mathbb{P}_{A_{0,1}|W_{0,1}=w}} \left[\prod_{1 \leq i < j \leq d} (A_{0,1}^{(i,j)})^{m_{ij}} \right] = 0, \quad (5.7)$$

$$\mathbb{E}_{\tilde{A} \sim \mathbb{P}_{\text{BF}}^{\theta,w}} \left[\prod_{1 \leq i < j \leq d} (\tilde{A}^{(i,j)})^{m_{ij}} \right] = 0, \quad (5.8)$$

provided that $\sum_{i=1}^d n_i + \left(\sum_{1 \leq i < j \leq d} n_{ij} \right)$ and $\sum_{1 \leq i < j \leq d} m_{ij}$ are odd.

Proof. See [Appendix B.2](#). □

Having $\tilde{A}$ unbiased means that some of the usual error analysis from stochastic numerics can be applied, such as in the proof of [Theorem 5.3.5](#) where one of the requirements is that $\tilde{A}$ be unbiased. Recall that the Milstein scheme requires a subroutine which generates samples of Lévy area given a Brownian increment. Since $\tilde{A} \sim \mathbb{P}_{\text{BF}}^{\theta,w}$ is unbiased, one can establish theoretical guarantees on the convergence of Milstein’s method with the BF generator as this subroutine. Note however, the order of convergence: with exact Lévy area samples one would obtain a rate of $O(h)$ not $O(\sqrt{h})$. The following result can be proven by applying [[MT21](#), Theorem 1.1] with $p_1 = \frac{3}{2}$, $p_2 = 1$.

Proposition 5.2.8. *Given a time horizon $T > 0$ and a step size $h = \frac{T}{N}$ where $N \geq 1$ denotes the number of steps, let $\{\tilde{X}_n\}_{n \in \{0, \dots, N\}}$ be the output of Milstein's scheme (see Eq. (B.4) in Appendix B.6.2 for the definition) applied to the SDE*

$$dX_t = \mu(X_t, t)dt + \sigma(X_t, t)dW_t,$$

where the Lévy areas provided as input were generated by the BF generator. Then if μ, σ are continuous and globally Lipschitz, there exists a constant $C > 0$ such that for sufficiently small h

$$\sup_{0 \leq n \leq N} \left[\mathbb{E} |\tilde{X}_n - X_{nh}|^2 \right]^{\frac{1}{2}} \leq Ch^{\frac{1}{2}}.$$

Remark 5.2.9. *Although Theorem 5.2.6 and Algorithm 1 leads to the generation of $(A_{0,1} | W_{0,1} = w)$, we emphasise that by the scale invariance of Brownian motion, it can be generalized to the generation of $(A_{0,t} | W_{0,1} = w)$ for any $t \in [0, 1]$. In particular, given w we can perform the following.*

- 1) *Sample w' from the distribution of $(W_t | W_1 = w)$, i.e. $w' \sim \mathcal{N}(tw, t(1-t))$;*
- 2) *Rescale $w'' = \frac{w'}{\sqrt{t}}$ and sample $a'' \sim (A_{0,1} | W_1 = w'')$;*
- 3) *Finally rescale again $a' = ta''$, which will have the desired distribution $a' \sim (A_{0,t} | W_1 = w)$.*

5.2.2 The pair-net generator

It is clear that for any $1 \leq i, j \leq d$, both $b_1^{(i,j)}$ and $A_1^{(i,j)}$ depend only on the paths of $\{W_t^{(i)}\}_{t \in [0,1]}$ and $\{W_t^{(j)}\}_{t \in [0,1]}$ but not on $\{W_t^{(k)}\}_{t \in [0,1]}$ for $k \notin \{i, j\}$. This dependency structure can be well described using graphs, encouraging us to employ model architectures reminiscent of Graph Neural Networks (GNN) [Sca+09].

Consider a clique on d nodes, where each node corresponds to one dimension of the Brownian motion and each edge (i, j) is associated to $A^{(i,j)}$ (or $b^{(i,j)}$). Unlike GNNs, where it might be desirable for information to propagate throughout the entire graph, we want edge (i, j) to never see information at node $k \notin \{i, j\}$. Hence, our architecture should function like a 1-step GNN with edge-wise outputs.

So that $A^{(i,j)}$ only depends on $W^{(i)}$ and $W^{(j)}$, we generate a separate noise vector for each dimension of Brownian motion. We can interpret this as some embedding of the entire path $\{W_t^{(i)}\}_{t \in [0,1]} \mapsto \text{noise}^{(i)}$, but in practice we use Gaussian noise.

Definition 5.2.10 (Pair-net). *The Pair-net is defined as the mapping:*

$$\begin{aligned} \text{PairNN}_\theta : (\mathbb{R}, \mathcal{Z}) \times (\mathbb{R}, \mathcal{Z}) &\rightarrow \mathbb{R}, \\ (H, Z) \times (H', Z') &\mapsto \tilde{b}, \end{aligned}$$

where $\mathcal{Z}$ is the space of latent noise. Consider $H_{0,1}$ and $b_{0,1}$ from [Definition 5.2.1](#). We approximate $\mathbb{P}_{b_{0,1}|H_{0,1}=h}$ via

$$\tilde{\mathbf{b}}^{(i,j)} = \text{PairNN}_\theta((h^{(i)}, \mathbf{z}^{(i)}), (h^{(j)}, \mathbf{z}^{(j)})), \quad 1 \leq i < j \leq d$$

for any $\mathbf{z} \in \mathcal{Z}^d$. In practice, we choose $\mathcal{Z}$ to be $\mathbb{R}^n$ and we let $\mathbf{z}^{(i)}$ be an n -dimensional Gaussian noise for $1 \leq i \leq d$. Combining this with bridge-flipping, we describe the generation algorithm and flowchart in [Algorithm 2](#) and [Fig. 5.3](#).

Algorithm 2 Area generation using Pair-net generator and Bridge-flipping

Input: θ - neural net parameters, d - dimension of Brownian motion, w - Brownian increment, n - dimension of noise vector associated with each coordinate of Brownian motion.

- 1: $\xi_0 \leftarrow \text{Rad}(1/2)$; $\xi \leftarrow \text{Rad}^d(1/2)$
 - 2: $\mathbf{z}^{(i)} \leftarrow \mathcal{N}^n(0, 1)$ for $1 \leq i \leq d$; $h \leftarrow \mathcal{N}^d(0, \frac{1}{12}(t-s))$
 - 3: $\tilde{\mathbf{b}}^{(i,j)} \leftarrow \text{PairNN}_\theta((h^{(i)}, \mathbf{z}^{(i)}), (h^{(j)}, \mathbf{z}^{(j)}))$ for $1 \leq i < j \leq d$
 - 4: **return** $\text{BF}(w, h, \tilde{\mathbf{b}}, \xi_0, \xi)$
-

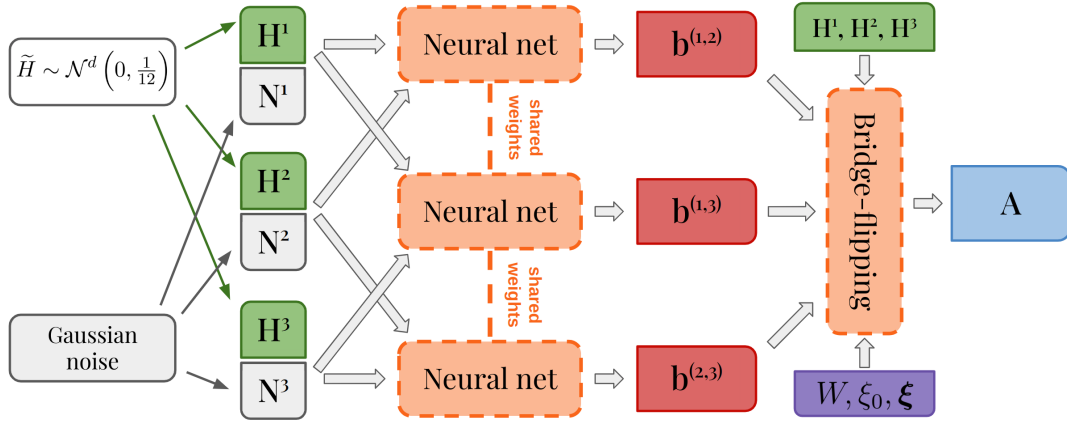


Figure 5.3: A schematic of the Pair-net architecture when $d = 3$.

Another benefit of the Pair-net architecture is that the neural net PairNN_θ can now be significantly smaller, since it does not need to capture the relationship between all the dimensions. This comes with the downside of requiring more passes (as shown in [Fig. 5.3](#); we use a single net to generate all $\mathbf{b}$ s, but each dimension of $\mathbf{b}$

requires its own forward pass), but suitable indexing and batching can make this efficient. Indeed, using PairNet results in both a speed-up (compared to truncated Fourier series methods) and greater accuracy (see [Section 5.5](#)). Furthermore, we emphasise that the PairNet structure allows us to generate Lévy area for any dimension d' , even though we train the model in a lower dimension $d < d'$. From now on, we will take nsz to be the total noise dimension for the complete Lévy area generation, namely $\text{nsz} = n \times d$ where n is the latent noise dimension and d denotes the Brownian dimension. In the framework of [Definition 5.1.1](#), our generator is now $G_\theta : \mathbb{R}^d \times \mathbb{R}^{\text{nsz}} \rightarrow \mathbb{R}^d \times \mathbb{R}^a$ and the evaluation follows [Algorithm 2](#).

5.3 Discriminator

One potential discriminator is the characteristic function GAN (CFGAN) approach introduced in [\[FSS20; Li+20a\]](#), which aims to learn the law of an underlying process by approximating its characteristic function. Compared to traditional GANs, we list some advantages of this approach.

- 1) The characteristic function always exists and is uniformly bounded;
- 2) The characteristic function fully describes the law of the random variable, hence offering good theoretical support.

In this section we introduce two different choices for the characteristic function. Firstly, we assess the distance between two distributions using the characteristic function distance.

Definition 5.3.1 (Characteristic function distance [\[FSS20\]](#)). *Let $\mathbb{P}_X$ and $\mathbb{P}_Y$ be the distributions of two $\mathbb{R}^d$ -valued random variable X and Y respectively. The characteristic function distance (CFD) between X and Y associated with an $\mathbb{R}^d$ -valued random variable $\Lambda \sim \nu$ is given by*

$$\text{CFD}_\nu(X, Y) = \mathbb{E}_{\Lambda \sim \nu} \left[\left\| \Phi_X(\Lambda) - \Phi_Y(\Lambda) \right\| \right],$$

where $\Phi_X(\Lambda) := \mathbb{E}_{X \sim \mathbb{P}_X} [\exp i \langle \Lambda, X \rangle]$.

The characteristic function distance has the following properties:

- 1) Definiteness: if the support of ν is $\mathbb{R}^d$, then $\text{CFD}_\nu(X, Y) = 0$ iff $\mathbb{P}_X = \mathbb{P}_Y$;
- 2) The distance is bounded and differentiable almost everywhere;
- 3) For certain ν , $\text{CFD}_\nu(X, Y_n) \rightarrow 0 \implies Y_n \xrightarrow{d} X$ ([\[Sri+10\]](#), [Proposition B.1.2](#)).

Supposing we restrict ν to be a member some parametric family of distributions $\{f_\theta : \theta \in \Theta\}$, e.g. Gaussian with $\theta = (\mu, \sigma^2)$, we can parametrise the distribution ν by learnable coefficients which will then play the role of the discriminator in the GAN setting. Backpropagation on these coefficients is well-understood by utilising techniques such as the reparameterisation trick. In practice, the characteristic function of both the generator and true data is approximated by empirical measures: if $\Lambda \sim \nu$, and $x_1, \dots, x_N$ are samples from $\mathbb{P}_X$, then we estimate $\Phi_X(\Lambda)$ by

$$\hat{\Phi}_X(\Lambda) = \frac{1}{N} \sum_{i=1}^N \exp(i\langle \Lambda, x_i \rangle).$$

The use of the empirical characteristic function in place of the analytical characteristic function is justified in [Proposition B.1.1](#). In the case that $X = (W_t, A_t)$, i.e. the joint process of d -dimensional Brownian motion and the corresponding Lévy area at any time t , we may obtain $\Phi_{(W_t, A_t)}$ analytically, see [Theorem B.2.1](#). With access to the analytical form of the target characteristic function, it would be possible to train our generator without access to any samples from Lévy area. While this proves a successful approach, especially in low dimensions, there is a better option: Chen training, which will be discussed in [Section 5.3.2](#).

5.3.1 Unitary characteristic function

In this subsection we introduce an extension of the classical characteristic function of a random variable, which was originally proposed in [\[CL16\]](#) and [\[LLN23\]](#). Denote by U_N the set of unitary matrices of dimension $N \times N$. Then U_N is a matrix Lie group with the group operation given by matrix multiplication. The Lie algebra corresponding to U_N , denoted by $\mathfrak{u}_N$ is the set of anti-Hermitian matrices, i.e. $\mathfrak{u}_N := \{A \in \mathbb{C}^{N \times N} : A^* + A = 0\}$. We now give the definition of the unitary characteristic function.

Definition 5.3.2 (Unitary characteristic function). *Let $X \sim \mathbb{P}_X$ be an $\mathbb{R}^d$ -valued random variable. Let $B \geq 1$ be an integer and let $M \in \mathcal{L}(\mathbb{R}^d, \mathfrak{u}_n)$, then the unitary representation function of X at M is given by*

$$\mathcal{U}_M(X) := \exp(M(X)), \tag{5.9}$$

where $\exp$ denotes the matrix exponential. The unitary characteristic function (UCF) of X is a mapping from $\mathcal{L}(\mathbb{R}^d, \mathfrak{u}_N)$ to $GL(N)$ such that

$$UCF_N(X)(M) := \mathbb{E}_{X \sim \mathbb{P}_X} [\mathcal{U}_M(X)]. \tag{5.10}$$

In practice, $UCF_N(X)(M)$ is approximated by Monte-Carlo sampling. Let $\mathbf{X} = \{X_i\}_{i=1}^m$ be m samples from $\mathbb{P}_X$. The empirical unitary characteristic function is given by

$$EUCF_m(\mathbf{X})(M) := \frac{1}{m} \sum_{i=1}^m \mathcal{U}_M(X_i).$$

Remark 5.3.3. When $N = 1$, $\mathfrak{u}_1$ is simply the set of pure imaginary numbers and M can be identified with an element $\lambda \in \mathbb{R}^d$ such that $M(x) = i\langle \lambda, x \rangle$ for all $x \in \mathbb{R}^d$ where $\langle \cdot, \cdot \rangle$ denotes the standard inner product. Then, $UCF_1(X)$ recovers the ordinary characteristic function of a d -dimensional random variable.

Similarly to [Definition 5.3.1](#), for any $N \geq 1$ we can define the *unitary characteristic function distance* for two random variables X and Y as

$$\text{UCFD}_N^2(X, Y) = \mathbb{E}_{M \sim \mathbb{P}_M} \left[\left\| UCF_N(X)(M) - UCF_N(Y)(M) \right\|_{HS}^2 \right],$$

where $\mathbb{P}_M$ denotes the distribution of the linear mapping M and $\|\cdot\|_{HS}$ denotes the Hilbert–Schmidt norm. Since Monte-Carlo over any fully supported measure on $\mathcal{L}(\mathbb{R}^d, \mathfrak{u}_N)$ quickly becomes computationally infeasible even for moderate N , we will instead parameterise the law $\mathbb{P}_M$ to be a discrete measure. That is, we take

$$\mathbb{P}_M = \frac{1}{m} \sum_{i=1}^m \delta_{M_i},$$

where δ denotes the Dirac measure and each M_i can be parametrized as learnable coefficients and optimized using gradient-based methods. For computational and optimisation details of the UCFD, we refer the reader to [\[LLN23\]](#). Let X be any d -dimensional random variable, then benefits of UCFD compared to the standard characteristic function distance include the following.

- 1) Standard uniqueness results hold as $UCF_1(X)$ recovers the standard characteristic function;
- 2) Since $\mathfrak{u}_1$ is a subspace of $\mathfrak{u}_N$ for any $N \geq 2$, $\mathcal{U}_M$ should possess a richer structure than $\mathbb{C}$ for any linear mapping M into $\mathfrak{u}_n$. Although using $\mathfrak{u}_1$ already encodes enough information to determine the random variable, embedding the random variable into a Lie algebra of a higher degree appears to provide a more efficient way of representing this information. Moreover empirically, using the unitary characteristic function led to a more stable training procedure compared to using the ordinary characteristic function.

5.3.2 Chen training

Whilst Brownian motion over different intervals can be concatenated simply via addition, i.e. $W_{0,t} = W_{0,s} + W_{s,t}$, concatenation of Lévy areas requires an additional term, specified by Chen's relation. In its general version within rough path theory, Chen's relation establishes the homomorphism property of path signatures under concatenation as we saw earlier in this thesis. However, we recall here only the special case relating to Lévy area.

Proposition 5.3.4 (Chen's relation [Che57]). *For times $0 < s < t$*

$$A_{0,t}^{(i,j)} = A_{0,s}^{(i,j)} + A_{s,t}^{(i,j)} + \frac{1}{2} \left(W_{0,s}^{(i)} W_{s,t}^{(j)} - W_{0,s}^{(j)} W_{s,t}^{(i)} \right).$$

We now present the main theoretical contribution of this section, which can be viewed as a partial converse of the above proposition.

Theorem 5.3.5. *[Distributional uniqueness of Lévy area under Chen's relation]*

Suppose μ is a mean zero probability distribution on $\mathbb{R}^d \times \mathbb{R}^{d \times d}$, where the first marginal has finite second moment. Let $(V_i, Z_i) \stackrel{i.i.d.}{\sim} \mu$ for $i = 1, 2$; if it holds that the pair

$$\begin{aligned} V_3 &:= \frac{1}{\sqrt{2}} (V_1 + V_2), \\ Z_3 &:= \frac{1}{2} Z_1 + \frac{1}{2} Z_2 + \frac{1}{4} (V_1 \otimes V_2 - V_2 \otimes V_1) \end{aligned} \tag{5.11}$$

is also distributed according to μ , then μ is the distribution of $(W_{0,1}, A_{0,1})$ where A is the Lévy area process associated with a d -dimensional Brownian motion W .

Proof. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space carrying for each $N \in \mathbb{N}$ a sequence of random variables $(V_i^{N,N}, Z_i^{N,N})_{i=1}^{2^N} \stackrel{i.i.d.}{\sim} \mu$. For each N define via backward recursion the sequences $(V_i^{k,N}, Z_i^{k,N})_{i=1}^{2^k}$ for $k = 0, 1, \dots, N-1$ by

$$V_i^{k,N} = \frac{1}{\sqrt{2}} (V_{2i-1}^{k+1,N} + V_{2i}^{k+1,N}) \quad \text{and} \tag{5.12}$$

$$Z_i^{k,N} = \frac{1}{2} (V_{2i-1}^{k+1,N} + V_{2i}^{k+1,N}) + \frac{1}{4} (V_{2i-1}^{k+1,N} \otimes V_{2i}^{k+1,N} - V_{2i}^{k+1,N} \otimes V_{2i-1}^{k+1,N}). \tag{5.13}$$

Using the assumptions on μ , it follows that $(V_i^{k,N}, Z_i^{k,N})_{i=1}^{2^k} \stackrel{i.i.d.}{\sim} \mu$ for every k . We now define

$$X_k^N = \left(\frac{1}{2^{k/2}} \sum_{i=1}^{2^k} V_i^{k,N}, \frac{1}{2^n} \sum_{i=1}^{2^k} Z_i^{k,N} \right) \quad \text{for } k = 0, \dots, N, \tag{5.14}$$

noting that $X_0^N = (V_1^{0,N}, Z_1^{0,N}) \sim \mu$. For every N we have the telescoping sum

$$X_0^N = \sum_{k=1}^{N-1} (X_{k-1}^N - X_k^N) + X_N^N,$$

and since by [Eqs. \(5.12\)](#) and [\(5.13\)](#)

$$X_{k-1}^N - X_k^N = \frac{1}{2^{k+1}} \left(0, \sum_{i=1}^{2^{k-1}} V_{2i-1}^{k,N} \otimes V_{2i}^{k,N} - V_{2i}^{k,N} \otimes V_{2i-1}^{k,N} \right),$$

this relation may be rewritten as

$$\begin{aligned} X_0^N = & \left(\frac{1}{2^{N/2}} \sum_{i=1}^{2^N} V_i^{N,N}, \sum_{k=1}^N \frac{1}{2^{k+1}} \sum_{i=1}^{2^{k-1}} V_{2i-1}^{k,N} \otimes V_{2i}^{k,N} - V_{2i}^{k,N} \otimes V_{2i-1}^{k,N} \right) \\ & + \left(0, \frac{1}{2^N} \sum_{i=1}^{2^N} Z_i^{N,N} \right). \end{aligned}$$

The left-hand side has distribution μ independent of N . On the right-hand side the second term tends to zero in probability, while the first term can be recognised as having the distribution

$$(W_{0,1}^{D_N}, \text{Area}(W_{0,1}^{D_N})) \xrightarrow{d} (W_{0,1}, A_{0,1}) \text{ as } N \rightarrow \infty,$$

where W^{D_N} is the piecewise linear interpolation of a rescaled random walk along the N^{th} diadic partition D_N of $[0, 1]$. Here, $\text{Area}(W_{0,1}^{D_N})$ denotes the Lévy area of the piecewise linear interpolation. The convergence in distribution follows from Donsker's theorem for enhanced Brownian motion [\[BFH09, Theorems 2 and 3\]](#). The proof is concluded by an application of Slutsky's theorem. $\square$

Remark 5.3.6. *To see that*

$$\sum_{k=1}^N \frac{1}{2^{k+1}} \sum_{i=1}^{2^{k-1}} V_{2i-1}^{k,N} \otimes V_{2i}^{k,N} - V_{2i}^{k,N} \otimes V_{2i-1}^{k,N}$$

has the distribution of the Lévy area of the piecewise linear approximation of Brownian motion on the diadic partition D^N , one can note that for each $k = 1, \dots, N$, the summand is the signed area between the piecewise linear approximation on D^k and the piecewise linear approximation on the coarser partition D^{k-1} . For a piecewise linear path, the Lévy area is given by the sum of these enclosed areas. Alternatively, one can show via induction that the sum may be rearranged to give the trapezium rule applied on D^N to the Lévy area

of the piecewise linear approximation on the same partition. The trapezium rule in this case will be exact.

Including a finite-variance assumption on the second coordinate measure μ together with a Gaussian assumption on the first marginal yields an alternative proof using Wasserstein distances, see [Jel+23] for details. The preceding motivates the following procedure, which takes samples Z_1, Z_2 from some distribution and concatenates them using Chen's relation to produce samples that are closer in distribution to $A_{0,1}$. With abuse of notation, if $\mathbf{X} \in \mathbb{R}^{m \times d}$ consists of m samples of a d -dimensional random variable, we denote by $\mathbf{X}^{(i)} \in \mathbb{R}^d$ the i^{th} coordinate of each sample. Adopting this notation, we describe the Chen-combine operation in [Algorithm 3](#).

We also note that the proof of [Jel+23, Theorem B.2] shows that the repeated application of Chen-combine gives convergence of order at least $\frac{1}{2}$ in the 2-Wasserstein metric. Faster rates can be proven assuming a suitable starting distribution, such as Davie's approximation [Dav14; FL15]. To fully justify the Chen training procedure, we prove that the conditional Wasserstein distance [Che+24a] between

- 1) Lévy area conditioned on a Brownian increment; and
- 2) An estimator conditioned on the same Brownian increment

is bounded by an explicit constant times the conditional Wasserstein distance between

- 1) The estimator conditioned on the Brownian increment; and
- 2) One iteration of Chen's relation applied to the estimator, still conditioned on the Brownian increment.

Definition 5.3.7 (Conditional Wasserstein distance). *Let μ and ν be probability measures on the product of two metric spaces $X_1 \times X_2$ (equipped with the L_2 product metric ρ) for which the restrictions to the first coordinate are equal, that is $\lambda := \pi_{\#}^{(1)} \mu = \pi_{\#}^{(1)} \nu$. Define the set of couplings*

$$\Gamma^*(\mu, \nu) = \left\{ \gamma \in \Gamma(\mu, \nu) : \pi_{\#}^{(1,3)} \gamma = \Delta_{\#} \lambda \right\},$$

where $\Gamma(\mu, \nu)$ is the set of all couplings of μ and ν and $\Delta : X_1 \rightarrow X_1 \times X_1$ is given by $x \mapsto (x, x)$. The conditional Wasserstein distance between μ and ν is then defined to be

$$\mathcal{W}_{2, \pi^{(1)}}^2(\mu, \nu) := \inf_{\gamma \in \Gamma^*(\mu, \nu)} \mathbb{E}_{x_1, x_2 \sim \gamma} [\rho(x_1, x_2)^2].$$

Theorem 5.3.8. Let μ be a mean-zero square integrable measure on $X_1 \times X_2 := \mathbb{R}^d \times \mathbb{R}^{d \times d}$, where $\pi_{\#}^{(1)} \mu = \mathcal{N}^d(0, 1)$, and let ν be the joint distribution of a Brownian increment and its Lévy area over the time interval $[0, 1]$. Define the “Lévy-error” of μ as

$$\varepsilon_{\text{Lévy}} := \mathcal{W}_{2, \pi^{(1)}}(\mu, \nu). \quad (5.15)$$

Let $(V_i, Z_i) \stackrel{\text{i.i.d.}}{\sim} \mu$, and let $\tilde{\mu}$ be the law of the random variable

$$\tilde{V} = \frac{1}{\sqrt{2}}(V_1 + V_2) \quad \text{and} \quad \tilde{Z} = \frac{1}{2}(Z_1 + Z_2) + \frac{1}{4}(V_1 \otimes V_2 - V_2 \otimes V_1).$$

The “Chen-error” of μ is then given by $\varepsilon_{\text{Chen}} := \mathcal{W}_{2, \pi^{(1)}}(\mu, \tilde{\mu})$. Then

$$\varepsilon_{\text{Lévy}} \leq (2 + \sqrt{2})\varepsilon_{\text{Chen}}. \quad (5.16)$$

Remark 5.3.9. Note that $\varepsilon_{\text{Chen}}$ is well-defined, since the $\pi_{\#}^{(1)} \tilde{\mu}$ is also $\mathcal{N}^d(0, 1)$.

Proof. Fix $\varepsilon > 0$ and let $\gamma^1 \in \Gamma^*(\mu, \nu)$ be a coupling for which

$$\mathbb{E}_{(V, Z), (W, A) \sim \gamma^1} \left[\|Z - A\|_2^2 \right] \leq \varepsilon_{\text{Lévy}}^2 + \varepsilon.$$

Note that we are able to omit V and W inside the expectation due to the conditional nature of the coupling. Additionally, let $\gamma^2 \in \Gamma^*(\mu, \tilde{\mu})$ be a coupling for which

$$\mathbb{E}_{(V, Z), (\tilde{V}, \tilde{Z}) \sim \gamma^2} \left[\|Z - \tilde{Z}\|_2^2 \right] \leq \varepsilon_{\text{Chen}}^2 + \varepsilon.$$

Suppose there exists a probability measure γ^* on $(X_1 \times X_2)^5$, such that any

$$((V_1, Z_1), (V_2, Z_2), (V_3, Z_3), (W_1, A_1), (W_2, A_2)) \sim \gamma^*$$

has the following properties.

- (i) $((V_i, Z_i), (W_i, A_i)) \stackrel{\text{i.i.d.}}{\sim} \gamma^1$ for $i = 1, 2$;
- (ii) If $\tilde{V} = \frac{1}{\sqrt{2}}(V_1 + V_2)$ and $\tilde{Z} = \frac{1}{2}(Z_1 + Z_2) + \frac{1}{4}(V_1 \otimes V_2 - V_2 \otimes V_1)$, then $((V_3, Z_3), (\tilde{V}, \tilde{Z})) \sim \gamma^2$.

We will return to the construction of such a measure later. For now, define

$$W_3 := \frac{1}{\sqrt{2}}(W_1 + W_2) \quad \text{and} \quad A_3 := \frac{1}{2}(A_1 + A_2) + \frac{1}{4}(W_1 \otimes W_2 - W_2 \otimes W_1).$$

By construction of γ^* and invariance of ν under the rescaled Chen's relation, it holds that the law of $((V_3, Z_3), (W_3, A_3))$ is in $\Gamma^*(\mu, \nu)$. Hence

$$\begin{aligned}\epsilon_{\text{Lévy}}^2 &\leq \mathbb{E}_{\gamma^*} \left[\|A_3 - Z_3\|_2^2 \right] \\ &= \mathbb{E}_{\gamma^*} \left[\|A_3 - \tilde{Z} - Z_3 + \tilde{Z}\|_2^2 \right] \\ &= \mathbb{E}_{\gamma^*} \left[\|A_3 - \tilde{Z}\|_2^2 \right] + \mathbb{E}_{\gamma^*} \left[\|Z_3 - \tilde{Z}\|_2^2 \right] - 2\mathbb{E}_{\gamma^*} \left[\langle A_3 - \tilde{Z}, Z_3 - \tilde{Z} \rangle \right].\end{aligned}$$

The second term in the final line is bounded by $\mathbb{E}_{\gamma^*} \left[\|A_3 - \hat{Z}\|_2^2 \right] \leq \epsilon_{\text{Chen}}^2 + \varepsilon$ by the construction of γ^* . Now

$$A_3 - \tilde{Z} = \frac{1}{2}(A_1 + A_2) - \frac{1}{2}(Z_1 + Z_2) = \frac{1}{2}(A_1 - Z_1) + \frac{1}{2}(A_2 - Z_2),$$

where the equalities hold almost surely by the nature of the couplings, so that

$$\begin{aligned}\mathbb{E}_{\gamma^*} \left[\|A_3 - \tilde{Z}\|_2^2 \right] &= \mathbb{E}_{\gamma^*} \left[\left\| \frac{1}{2}(A_1 - Z_1) + \frac{1}{2}(A_2 - Z_2) \right\|_2^2 \right] \\ &= \frac{1}{4} \left(\mathbb{E}_{\gamma^*} \left[\|A_1 - Z_1\|_2^2 \right] + \mathbb{E}_{\gamma^*} \left[\|A_2 - Z_2\|_2^2 \right] \right. \\ &\quad \left. + 2\mathbb{E}_{\gamma^*} \left[\langle A_1 - Z_1, A_2 - Z_2 \rangle \right] \right) \\ &\leq \frac{1}{2} (\epsilon_{\text{Lévy}}^2 + \varepsilon).\end{aligned}$$

Note that the cross terms vanish by independence. Finally, we bound the third term using the Cauchy-Schwarz inequality

$$\begin{aligned}2 \left| \mathbb{E} \left[\langle A_3 - \tilde{Z}, Z_3 - \hat{Z} \rangle \right] \right| &\leq 2\mathbb{E} \left[\|A_3 - \tilde{Z}\|_2^2 \right]^{\frac{1}{2}} \mathbb{E} \left[\|Z_3 - \hat{Z}\|_2^2 \right]^{\frac{1}{2}} \\ &\leq \sqrt{2} (\epsilon_{\text{Lévy}}^2 + \varepsilon)^{\frac{1}{2}} (\epsilon_{\text{Chen}}^2 + \varepsilon)^{\frac{1}{2}}.\end{aligned}$$

Substituting all bounds into the original inequality and taking $\varepsilon \rightarrow 0$, we obtain

$$\epsilon_{\text{Lévy}}^2 \leq \frac{1}{2} \epsilon_{\text{Lévy}}^2 + \sqrt{2} \epsilon_{\text{Lévy}} \epsilon_{\text{Chen}} + \epsilon_{\text{Chen}}^2 = \left(\frac{1}{\sqrt{2}} \epsilon_{\text{Lévy}} + \epsilon_{\text{Chen}} \right)^2.$$

Since all quantities are non-negative, we see that $\left(1 - \frac{1}{\sqrt{2}}\right) \epsilon_{\text{Lévy}} \leq \epsilon_{\text{Chen}}$ as required.

We now return to the problem of constructing the measure γ^* . To this end, disintegrate γ^1 as $d\gamma^1((v, z), (w, a)) = d\gamma_{(v, z)}^1(a) d\mu(v, z)$ (which is possible since $w = v$, γ^1 -a.s.), i.e.

$$\int_{(X_1 \times X_2)^2} f(v, z, w, a) d\gamma^1(v, z, w, a) = \int_{X_1 \times X_2} \left(\int_{X_2} f(v, z, v, a) d\gamma_{(v, z)}^1(a) \right) d\mu(v, z)$$

for any Borel measurable f . Also disintegrate γ^2 onto its second coordinate as $d\gamma^2((v, z), (\tilde{v}, \tilde{z})) = d\gamma_{(\tilde{v}, \tilde{z})}^2(z) d\tilde{\mu}(\tilde{v}, \tilde{z})$. Now define the product kernel ρ by

$$d\rho_{(v_1, z_1, v_2, z_2)}(a_1, a_2, z_3) = d\gamma_{(v_1, z_1)}^1(a_1) d\gamma_{(v_2, z_2)}^1(a_2) d\gamma_{(\tilde{v}, \tilde{z})}^2(z_3)$$

where $\tilde{v} = \frac{1}{\sqrt{2}}(v_1 + v_2)$ and $\tilde{z} = \frac{1}{2}(v_1 + v_2) + \frac{1}{4}(v_1 \otimes v_2 - v_2 \otimes v_1)$. The required measure γ^* is then defined by the integration of bounded Borel measurable f as

$$\begin{aligned} \int f d\gamma^* &= \int_{(X_1 \times X_2)^2} \left[\int_{X_2^3} f((v_1, z_1), (v_1, a_1), (v_2, z_2), (v_2, a_2), (v_3, a_3)) \right. \\ &\quad \left. \times d\rho_{(v_1, z_1, v_2, z_2)}(a_1, a_2, z_3) \right] d\mu(v_1, z_1) d\mu(v_2, z_2) \end{aligned}$$

□

Since $\mathcal{W}_2(\mu, \nu)$ is bounded by $\mathcal{W}_{2, \pi_{\#}^{(1)}}(\mu, \nu)$ we immediately obtain the following.

Corollary 5.3.10. *The error $\mathcal{W}_2(\mu, \nu)$ is bounded by $(\sqrt{2} + 2)\epsilon_{\text{Chen}}$.*

Motivated by the above results, we now present our training procedure. Assume that $(\mathbf{W}, \tilde{\mathbf{A}}) \in \mathbb{R}^{2m \times (d+a)}$ is $2m$ samples of $(d+a)$ -dimensional random variables. By Chen-combine $(\mathbf{W}, \tilde{\mathbf{A}})$ we mean: evenly split $(\mathbf{W}, \tilde{\mathbf{A}})$ into two blocks with equal size, denoted by $(\mathbf{W}, \tilde{\mathbf{A}})_{\text{first}}, (\mathbf{W}, \tilde{\mathbf{A}})_{\text{second}} \in \mathbb{R}^{m \times (d+a)}$, and apply the operation described in Algorithm 3. The output of Chen-combine will be in $\mathbb{R}^{m \times (d+a)}$.

Algorithm 3 Chen-combine

Input: m - batch size, d - Brownian dimension, a - Lévy dimension,

$(\mathbf{W}, \tilde{\mathbf{A}})_{\text{first}} \in \mathbb{R}^{m \times (d+a)}, (\mathbf{W}, \tilde{\mathbf{A}})_{\text{second}} \in \mathbb{R}^{m \times (d+a)}$ - Brownian increments and Lévy area samples.

- 1: $\hat{\mathbf{W}} \leftarrow \mathbf{0} \in \mathbb{R}^{m \times d}, \hat{\mathbf{A}} \leftarrow \mathbf{0} \in \mathbb{R}^{m \times a}$
 - 2: $\mathbf{W}_{\text{first}} \leftarrow \frac{1}{\sqrt{2}} \mathbf{W}_{\text{first}}, \mathbf{W}_{\text{second}} \leftarrow \frac{1}{\sqrt{2}} \mathbf{W}_{\text{second}}, \quad \triangleright \text{Brownian scaling } W_{0, \frac{1}{2}} \stackrel{d}{=} \frac{1}{\sqrt{2}} W_{0, 1}$
 - 3: $\tilde{\mathbf{A}}_{\text{first}} \leftarrow \frac{1}{2} \tilde{\mathbf{A}}_{\text{first}}, \tilde{\mathbf{A}}_{\text{second}} \leftarrow \frac{1}{2} \tilde{\mathbf{A}}_{\text{second}}, \quad \triangleright \text{Brownian scaling } A_{0, \frac{1}{2}} \stackrel{d}{=} \frac{1}{2} A_{0, 1}$
 - 4: **for** $i \in \{1, \dots, d\}$ **do**
 - 5: $\hat{\mathbf{W}}^{(i)} \leftarrow \mathbf{W}_{\text{first}}^{(i)} + \mathbf{W}_{\text{second}}^{(i)}$
 - 6: **for** $j \in \{i+1, \dots, d\}$ **do**
 - 7: $\mathbf{D} \leftarrow \frac{1}{2} \left(\mathbf{W}_{\text{first}}^{(i)} \odot \mathbf{W}_{\text{second}}^{(j)} - \mathbf{W}_{\text{first}}^{(j)} \odot \mathbf{W}_{\text{second}}^{(i)} \right)$
 - 8: $\hat{\mathbf{A}}^{(i, j)} \leftarrow \tilde{\mathbf{A}}_{\text{first}}^{(i, j)} + \tilde{\mathbf{A}}_{\text{second}}^{(i, j)} + \mathbf{D}, \quad \triangleright \text{Chen's relation}$
 - 9: **return** $(\hat{\mathbf{W}}, \hat{\mathbf{A}})$
-

5.4 LévyGAN

In this section we incorporate the ideas presented in [Sections 5.2](#) and [5.3](#) into our tailored model “LévyGAN”, used to generate the associated Lévy area conditioned on the Brownian increments. Similar to [Definition 5.1.1](#), we provide the definition of the proposed model as follows.

Definition 5.4.1 (LévyGAN). *Let PairNN_θ denote a PairNet generator defined in [Definition 5.2.10](#). Given a d -dimensional Brownian increment at $t = 1$, $W \sim \mathcal{N}^d(0, 1)$, for all $1 \leq i < j \leq d$ we generate the estimated Lévy area associated to W as follows:*

$$\tilde{b}^{(i,j)} = \text{PairNN}_\theta((H^{(i)}, Z^{(i)}), (H^{(j)}, Z^{(j)})) \quad (5.17)$$

$$\tilde{A}^{(i,j)} = H^{(i)}W^{(j)} - H^{(j)}W^{(i)} + \tilde{b}^{(i,j)} \quad (5.18)$$

Let $\mathbf{w} \in \mathbb{R}^{N \times d}$ be N samples of Brownian increment and the associated $\tilde{\mathbf{A}} \in \mathbb{R}^{N \times a}$ generated according to [Eqs. \(5.17\)](#) and [\(5.18\)](#). Then, we construct new samples using Chen-combine defined in [Algorithm 3](#):

$$(\mathbf{w}_{\text{Chen}}, \tilde{\mathbf{A}}_{\text{Chen}}) := \text{Chen-combine}(\mathbf{w}, \tilde{\mathbf{A}}).$$

For $N \geq 1$, let $\mathfrak{u}_N$ be the Lie algebra of the unitary matrix group U_N . Recall EUCF_N from [Definition 5.3.2](#), and let $\mathcal{M} = \{M_i\}_{i=1}^M$, $M_i \in L(\mathbb{R}^{d+a}, \mathfrak{u}_N)$ be a collection of linear mappings in $\mathfrak{u}_n$, each of them parametrised by an element in $\mathbb{R}^{(d+a) \times \dim(\mathfrak{u}_N)}$. $\mathcal{M}$ will play the role of the discriminator. Finally, training is performed with respect to the following min-max game

$$\min_{\theta} \max_{\mathcal{M}} \text{Loss}(\theta, \mathcal{M}; \mathbf{w}),$$

where $\text{Loss}(\theta, \mathcal{M}; \mathbf{w}) := \text{EUCFD}_m((\mathbf{w}, \tilde{\mathbf{A}}), (\mathbf{w}_{\text{Chen}}, \tilde{\mathbf{A}}_{\text{Chen}}))$. The training algorithm and flowchart are described in [Algorithm 4](#) and [Fig. 5.4](#).

Remark 5.4.2. One can interpret Chen training as a type of adaptive training, where the “almost true” target data $\tilde{\mathbf{A}}_{\text{Chen}}$ is always just better than the generator’s output, so that training can progress effectively. Thus, instead of requiring large datasets of “true” samples, which are costly to generate and difficult to handle, we can now efficiently generate new “true” data on the fly. This new true data can be of any desired quantity and of just the right precision required for training. One could choose to iteratively apply Chen-combine several times, but we have observed that training is slightly faster and more efficient when only a single application of Chen combine is used in [Algorithm 4](#). This is partly because using two Chen-combines produces twice fewer “true” samples and reducing the sample size leads to a poorer estimate of EUCFD.

Algorithm 4 Training algorithm for LévyGAN

Input: d - Brownian dimension, a - Lévy area dimension, n - noise dimension, m - Lie algebra degree, M - number of linear mappings onto u_N , PairNN_θ - generator, $\mathcal{M} \in \mathbb{R}^{M \times (d+a) \times \dim(u_N)}$ - discriminator, iter_d - number of discriminator updates per generator update, bsz - batch size, η_g, η_d - generator and discriminator learning rates.

```

1: while  $\theta, \mathcal{M}$  not converge do
2:   for  $i \in (1, \dots, \text{iter}_d)$  do
3:     Sample  $\mathbf{w} \sim \mathcal{N}^d(0, 1)$ ,  $(\mathbf{h}, \mathbf{z}) \sim \mathcal{N}^d(0, \frac{1}{12}) \times \mathcal{N}^{d \times n}(0, 1)$  of size  $2 \times \text{bsz}$ .
4:      $\tilde{\mathbf{b}}^{(i,j)} \leftarrow \text{PairNN}_\theta((\mathbf{h}^{(i)}, \mathbf{z}^{(i)}), (\mathbf{h}^{(j)}, \mathbf{z}^{(j)}))$  for  $1 \leq i < j \leq d$ 
5:      $\xi_0 \leftarrow \text{Rad}(1/2)$ ,  $\xi \leftarrow \text{Rad}^d(1/2)$ 
6:      $\tilde{\mathbf{A}} \leftarrow \text{BF}(\mathbf{w}, \mathbf{h}, \tilde{\mathbf{b}}, \xi_0, \xi)$ 
7:      $\mathbf{w}_{\text{Chen}}, \tilde{\mathbf{A}}_{\text{Chen}} \leftarrow \text{Chen-combine}(\mathbf{w}, \tilde{\mathbf{A}})$ 
8:      $\text{Loss}(\theta, \mathcal{M}; \mathbf{w}) \leftarrow \text{EUCFD}_m((\mathbf{w}, \tilde{\mathbf{A}}), (\mathbf{w}_{\text{Chen}}, \tilde{\mathbf{A}}_{\text{Chen}}))$ 
9:      $\mathcal{M} \leftarrow \mathcal{M} - \eta_d \cdot \nabla_{\mathcal{M}}(-\text{Loss}(\theta, \mathcal{M}; \mathbf{w}))$  ▷ Maximize the loss
10:  Sample  $\mathbf{w} \sim \mathcal{N}^d(0, 1)$ ,  $(\mathbf{h}, \mathbf{z}) \sim \mathcal{N}^d(0, \frac{1}{12}) \times \mathcal{N}^{d \times n}(0, 1)$  of size  $2 \times \text{bsz}$ .
11:   $\tilde{\mathbf{b}}^{(i,j)} \leftarrow \text{PairNN}_\theta((\mathbf{h}^{(i)}, \mathbf{z}^{(i)}), (\mathbf{h}^{(j)}, \mathbf{z}^{(j)}))$  for  $1 \leq i < j \leq d$ 
12:   $\xi_0 \leftarrow \text{Rad}(1/2)$ ,  $\xi \leftarrow \text{Rad}^d(1/2)$ 
13:   $\tilde{\mathbf{A}} \leftarrow \text{BF}(\mathbf{w}, \mathbf{h}, \tilde{\mathbf{b}}, \xi_0, \xi)$ 
14:   $\mathbf{w}_{\text{Chen}}, \tilde{\mathbf{A}}_{\text{Chen}} \leftarrow \text{Chen-combine}(\mathbf{w}, \tilde{\mathbf{A}})$ 
15:   $\text{Loss}(\theta, \mathcal{M}; \mathbf{w}) \leftarrow \text{EUCFD}_m((\mathbf{w}, \tilde{\mathbf{A}}), (\mathbf{w}_{\text{Chen}}, \tilde{\mathbf{A}}_{\text{Chen}}))$ 
16:   $\theta \leftarrow \theta - \eta_d \cdot \nabla_{\theta} \text{Loss}(\theta, \mathcal{M}; \mathbf{w})$  ▷ Minimize the loss
17: return  $\text{PairNN}_\theta, \mathcal{M}$ 
    
```

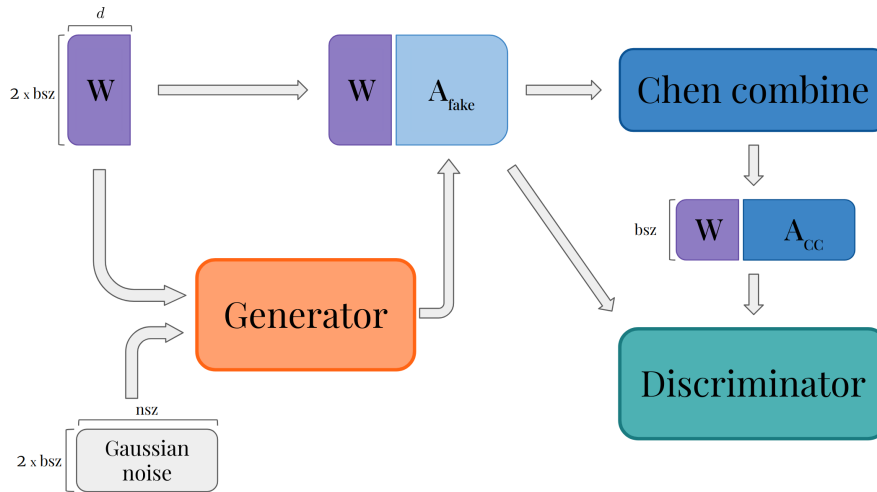


Figure 5.4: A schematic of LévyGAN. Here bsz denotes the batch dimension and we recall that nsz denotes the total noise dimension, namely $\text{nsz} = n \times d$.

5.5 Numerical experiments

We trained the model in $d = 4$. Note that by the architecture of the generator, the model can be used to generate Brownian Lévy area for any $d' \leq d$. The model can also be used to generate Lévy area for any $d' > d$, however the performance might deteriorate as the training was not performed in higher dimensions.

We performed the training procedure as illustrated in [Algorithm 4](#). On the generator side we used a feed-forward neural network, with the activation function chosen to be LeakyRelu. On the discriminator side, we parameterise 128 linear maps onto the Lie algebra of degree 3 to mimic the empirical distribution used to compute UCFD mentioned in [Section 5.3.1](#). The total number of training iterations is set to be 2500, which was approximately the point that we observed convergence in the marginal 2-Wasserstein metric to reference samples. We optimised both the generator and discriminator using stochastic gradient descent with the Adam optimiser. We set the batch size to 2^{13} and the learning rate for generator/discriminator was chosen to be 0.001/0.01 respectively. We also employed a learning rate scheduler, with both learning rates decaying every 500 iterations. Finally, we set iter_d to be 3.

We conducted a hyperparameter grid-search (see [Appendix B.4](#) evaluating the model performance according to the marginal 2-Wasserstein metric, with our optimal architecture as follows.

- (i) Feed-forward neural network with 3 hidden layers each of dimension 16;
- (ii) LeakyRelu activation function with slope = 0.01;
- (iii) Gaussian noise with $n = 3$.

Finally, we assessed the performance of our model on the generation of the coupled process for $d = 2, 3, 4$, and 8. We consider the following test metrics:

- 1) Marginal 2-Wasserstein metric;
- 2) Cross moment metric;
- 3) Characteristic function distance, computed using maximum mean discrepancy with different kernels;
- 4) Empirical Unitary Characteristic Function.

A detailed explanation of each test metric can be found in [Appendix B.3](#). We compare with two baselines: Foster's and Davie's moment matching generators [[Fos20](#); [Dav14](#)]. We regard the truncated Fourier series [[KR23](#)] of Lévy area up to

an L^2 precision of 10^{-4} as “true” samples. Finally, we provide a numerical SDE example for the log-Heston model using different estimators for fake Lévy area.

Table 5.1: Marginal distribution fitting and computational efficiency for the different generative models. The generation is done using NVIDIA Quadro RTX 8000. The marginal W_2 error is calculated with respect to the joint process generated by the Fourier series. Tests are performed with 2^{20} samples. The final column contains the results of the Fourier algorithm with 19 terms in the expansion (far more terms were used to generate “true” samples). This truncation has been chosen so that the performance is comparable to LévyGAN and Foster’s method.

Test Metric	LévyGAN	Foster	Davie	Fourier series
Computational time (s)	0.019	0.0071	0.002	3.1
Marginal W_2 (10^{-2})	.246 ± .013	.254 ± .010	2.03 ± .013	.27 ± 0.008

Table 5.2: Fourth moment and MMD-based metrics across different models and Brownian dimensions.

Dim	Test Metrics	LévyGAN	Foster	Davie
2	Fourth moment	.004 ± .002	.002 ± .002	.042 ± .001
	Polynomial MMD (10^{-5})	.341 ± .070	.654 ± .131	.646 ± .188
	Gaussian MMD (10^{-6})	1.47 ± .125	1.44 ± .128	34.6 ± .683
	EUCFD (10^{-2})	1.52 ± .213	1.92 ± .113	10.1 ± .851
3	Fourth moment	.004 ± .002	.004 ± .002	.043 ± .001
	Polynomial MMD (10^{-5})	2.18 ± .568	2.30 ± .732	2.26 ± .773
	Gaussian MMD (10^{-6})	1.87 ± .002	1.84 ± .001	16.3 ± .001
	EUCFD (10^{-2})	1.88 ± .063	2.03 ± .034	18.5 ± 1.11
4	Fourth moment	.004 ± .000	.006 ± .002	.043 ± .002
	Polynomial MMD (10^{-5})	4.04 ± .436	4.65 ± 1.31	5.62 ± .808
	Gaussian MMD (10^{-6})	1.90 ± .001	1.90 ± .001	263 ± .003
	EUCFD (10^{-2})	1.92 ± .026	2.03 ± .036	17.5 ± .483
8	Fourth moment	.006 ± .001	.006 ± .002	.044 ± .000
	Polynomial MMD (10^{-2})	1.13 ± .019	1.15 ± .030	1.31 ± .066
	Gaussian MMD (10^{-6})	1.91 ± .001	1.91 ± .000	1.92 ± .003
	EUCFD (10^{-2})	1.99 ± .002	1.99 ± .001	2.05 ± .003

5.5.1 SDE example

In this section, we will demonstrate how “fake” Lévy area can be used within SDE numerics to achieve both high order weak convergence as well as Multilevel Monte Carlo (MLMC) variance reduction. Although the synthetic Lévy area only needs to exhibit the correct mean and covariance to give high order weak convergence, we show that the bias introduced by the MLMC estimator is negligible in practice due to the small Chen error inherent in our generative model. A secondary motivation is to compare the various Lévy area generators and show that our GAN-based approach performs indistinguishably from previous state-of-the-art methods, whilst taking less time to generate samples.

Consider the Itô SDE from [Eq. \(5.1\)](#)

$$dX_t = f(X_t)dt + \sum_{i=1}^d g_i(X_t)dW_t^{(i)}, \quad X_0 = x_0, \quad (5.1 \text{ revisited})$$

where the solution X takes values in $\mathbb{R}^e$. To estimate the solution to [Eq. \(5.1\)](#), one typically uses a discretisation scheme that generates approximate sample paths of the solution X . Often the objective is to approximate quantities of the form

$$\mathbb{E} [\varphi(X) \mid X_0 = x_0], \quad (5.19)$$

where φ may depend on the whole sample path $(X_t)_{t \in [0, T]}$ though commonly is only a function of the solution X_T at the terminal time T . To measure the error of a particular discretisation scheme, there are two standard metrics: weak and strong error. We will only evaluate the weak error for reasons discussed in [Appendix B.6.1](#). To accurately determine the error of various numerical schemes, we seek a multidimensional SDE and a quantity of the form (5.19) which is known semi-analytically. Thankfully, such an example exists: the price of a European call option under the log-Heston model. The stochastic volatility model is defined by the following two-dimensional SDE:

$$\begin{aligned} dU_t &= \left(r - \frac{1}{2}V_t \right) dt + \sqrt{V_t}dW_t^{(1)}, \quad U_0 \in \mathbb{R} \\ dV_t &= \kappa(\theta - V_t)dt + \sigma\sqrt{V_t}dW_t^{(2)}, \quad V_0 > 0, \end{aligned} \quad (5.20)$$

for a pair of independent Brownian motions $W^{(1)}$ and $W^{(2)}$. To ensure the volatility term V remains positive, we must enforce the Feller condition $2\kappa\theta - \sigma^2 > 0$. The

payoff of a European call option for a price process S with $S := \exp(U)$ is given by

$$\varphi(S) := e^{-rT} \left(e^{U_T} - K \right)^+,$$

where r is the discount rate, K the strike price, and T the maturity. For the derivation and form of the semi-analytic formula for the expected value of the above, we refer the reader to [Ter22] and [Cri15].

5.5.2 Numerical results

We compare four discretisation schemes combined with the MLMC procedure [Gil08]. We briefly recall that MLMC is based on the idea of a telescoping sum of expectations. Indeed, assume we have L levels, and that Y_l is an estimator for X , based on a discretisation scheme with step-size h_l . Then we may write

$$\mathbb{E} \left[\varphi(Y_L) \right] = \sum_{l=1}^L \mathbb{E} \left[\varphi(Y_l) - \varphi(Y_{l-1}) \right],$$

with $Y_0 \equiv 0$. The MLMC estimator is then defined by

$$\bar{\varphi}_{n_1, \dots, n_L} = \sum_{l=1}^L \hat{\varphi}_{n_l}, \quad \text{where} \quad \hat{\varphi}_{n_l} = \frac{1}{n_l} \sum_{i=1}^{n_l} \left(\varphi(Y_l^{i,l}) - \varphi(Y_{l-1}^{i,l}) \right), \quad (5.21)$$

where $Y_l^{i,l}$ is the i^{th} sample of the estimator Y_l used on level l and $Y_{l-1}^{i,l}$ is the i^{th} sample of the estimator Y_{l-1} that is used on level l . It is important to note that the pairs $(Y_l^{i,l}, Y_{l-1}^{i,l})$ are coupled: the underlying Brownian path for each member of the pair is the same. In our case, the path on the lower level will be coarse (i.e. a large step size) and the higher level will be fine (i.e. a small step size). The standard condition used to ensure convergence of the telescoping sum of expectations is given by

$$\mathbb{E} \left[\varphi(Y_l^{i,l}) \right] = \mathbb{E} \left[\varphi(Y_l^{i,l+1}) \right].$$

However when incorporating a fake Lévy area term, our coupling at each level is defined as follows:

- 1) The Brownian increments for the fine path $Y_l^{i,l}$ are generated with step size h_l , with the increments of fake Lévy area generated using an estimator $\tilde{A}^{h_l}$;
- 2) The Brownian increments on the coarse path $Y_{l-1}^{i,l}$ are computed by pairwise summing the increments of the fine path. The fake Lévy area used on the coarse path is computed using one iteration of Chen's identity applied to the increments and areas of the fine path.

However, this scheme introduces a bias; namely, the distribution of the fake Lévy area used for the fine path at level l will not be the same as the distribution of Lévy area used for the coarse path at level $l + 1$. We may write the effect of this by amending the telescoping expectation to be

$$\begin{aligned} \sum_{l=1}^L \mathbb{E} \left[\varphi(Y_l(\tilde{A}^{h_l})) - \varphi(Y_{l-1}(\tilde{A}_{CC}^{h_l})) \right] &= \underbrace{\sum_{l=1}^L \mathbb{E} \left[\varphi(Y_l(\tilde{A}^{h_l})) - \varphi(Y_{l-1}(\tilde{A}^{h_{l-1}})) \right]}_{\text{desired telescoping expectation}} \\ &+ \underbrace{\sum_{l=1}^L \mathbb{E} \left[\varphi(Y_{l-1}(\tilde{A}^{h_{l-1}})) - \varphi(Y_{l-1}(\tilde{A}_{CC}^{h_l})) \right]}_{\text{bias term}}. \end{aligned} \quad (5.22)$$

We have emphasised the dependence of the fine level on the fake area $\tilde{A}^{h_l}$ and dependence of the coarse level of one Chen iteration of this fake area, denoted by $\tilde{A}_{CC}^{h_l}$. The bias introduced is exactly the second sum. We aim to show empirically that this sum is small in comparison to the size of the weak error due to the SDE discretisation. In order to minimise this sum, the distribution of an estimator $\tilde{A}$ must be as close as possible to the distribution $\text{Chen-combine}(\tilde{A})$ – exactly the criterion used to train our generator.

The four numerical schemes in the comparison are: no-area Milstein, antithetic Milstein [GS14a], the Strang splitting method, and a “Strang” log-ODE method. Only the final method incorporates the fake Lévy area. For details on the schemes see [Appendix B.6.2](#). The first two schemes were included to demonstrate that the rate of variance reduction is comparable to two popular schemes, while the weak error rate of the Strang log-ODE method is (conjectured to be) $O(h^2)$ where the other methods achieve a weak error rate of $O(h)$. The numerical simulations were performed with a constant time-step h_l on each level satisfying $h_l = \frac{1}{2}h_{l-1}$. On the coarsest level we use the timestep $h_0 = \frac{1}{2}$ for the Milstein methods and $h_0 = 1$ for the Strang methods; this results in the variance on this level being approximately equal across all three methods. The number of sample paths on each level satisfies $n_l = \frac{1}{2}n_{l-1}$ with $n_0 = 2^{27}$, so the computational effort on each level is approximately constant. We repeat the experiment forty times and report the average result. We fix the log-Heston model parameters to be $T = 1$, $r = 0.1$, $K = 20$, $\kappa = 2$, $\theta = 0.1$, $\sigma = 0.5$, $U_0 = \log(20)$, and $V_0 = 0.4$. The following plots report the multilevel variance defined by $\text{Var} \left[\varphi(Y_l^{:,l}) - \varphi(Y_{l-1}^{:,l}) \right]$ and the empirical error given by $|\bar{\varphi}_{n_1, \dots, n_l} - P_{\text{true}}|$, where P_{true} is the true price of the call option under the log-Heston model.

In Fig. 5.5a we see that the multilevel level variance of the Strang log-ODE method decreases at an approximate rate of $O(h^2)$. The rate for the Milstein antithetic scheme appears slightly higher, while the variance reduction rate for the standard Milstein method is clearly lower. As expected, the weak convergence rate of both the Milstein and Milstein antithetic schemes is of order $O(h)$ whilst the weak rate for the Strang log-ODE with fake Lévy area is approximately $O(h^2)$. It was conjectured in [FdRS24] that the Strang log-ODE method should attain this weak convergence rate, and the experiments corroborate this hypothesis.

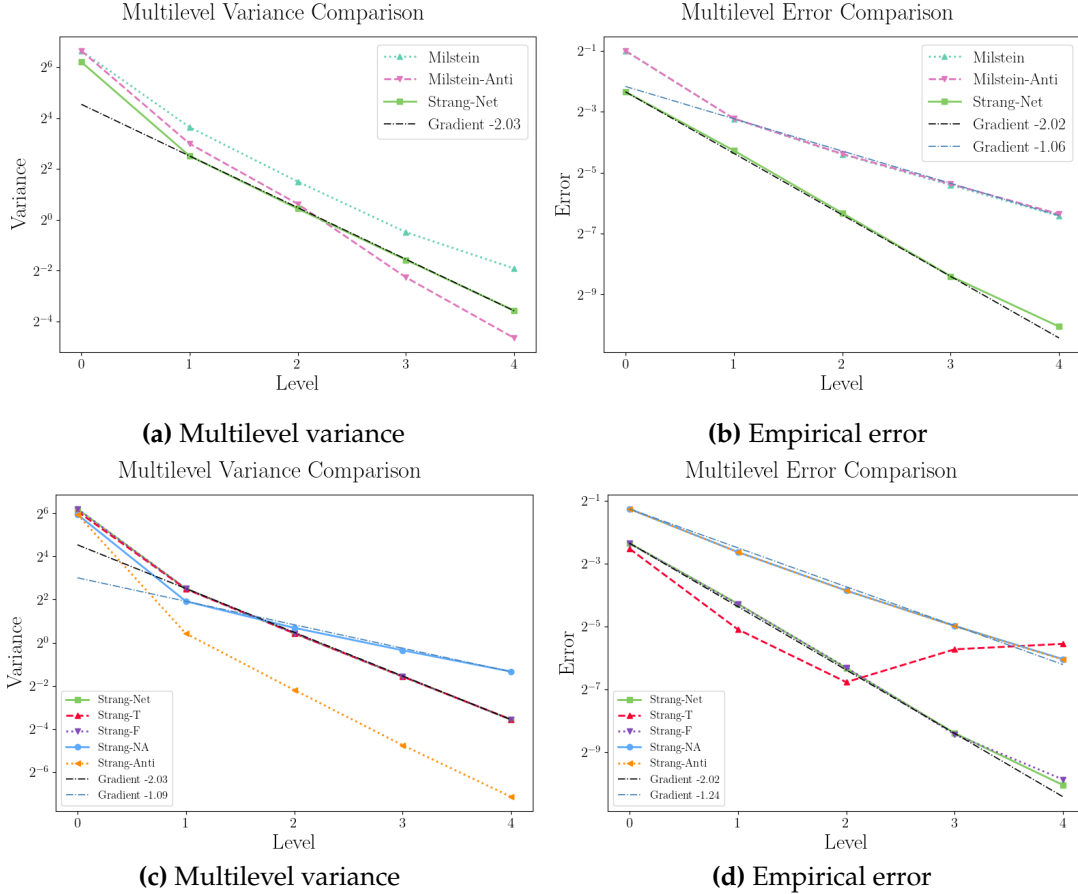


Figure 5.5: Plots of multilevel variance and empirical error. The top pair of plots compare the Milstein scheme without area, Milstein antithetic and Strang log-ODE scheme with fake Lévy area from our generator (labelled “Strang-Net”). The bottom pair compare the Strang log-ODE scheme using three different fake Lévy areas. “Strang-T” indicates that the fake Lévy area is an independent Rademacher random variable with the correct variance (the same random variable appearing in the Talay scheme), and “Strang-F” denotes Foster’s approximation. The “Strang-NA” line is the usual Strang splitting method with “Strang-Anti” being the antithetic version of this scheme.

When using a fake Lévy area in MLMC, a key factor for the performance of the scheme is how close the distribution of a Chen combined sample of Brownian motion and fake Lévy area is to the distribution before performing the combine operation. Since we employed Chen training, our model succeeds in matching the distributions well enough to match and even outperform Foster’s method. In

Table 5.3 we record the bias introduced at each level, as in [Eq. \(5.22\)](#), for each of the fake Lévy areas.

Table 5.3: Approximate bias introduced per level by the use of different fake Lévy areas. We use 2^{29} samples paths on each level.

Level	0	1	2	3
Strang-Net (10^{-3})	-0.402	0.313	-0.133	-0.652
Strang-F (10^{-3})	0.361	0.848	0.0714	1.31
Strang-T (10^{-3})	-17.0	4.33	-1.79	-1.10

For LévyGAN and Foster’s method, the bias introduced on each level of the telescoping sum is of order 2^{-11} ; this is far smaller than the weak error seen in [Fig. 5.5d](#). However, it is possible that the accumulated bias may then be of order 2^{-9} on the finest level, which may account in part for the slight deviation from the line at level 4. It is also clear that it is not enough for the fake Lévy area simply to match the mean and variance of true Lévy area, as demonstrated by the poor performance of the “Strang-T” method. However, we do note here that a scheme matching the conditional variance of Lévy area given a Brownian increment performed similarly to Foster’s method in previous experiments. We may also see from the performance of the Strang splitting method that without the fake Lévy area terms, one achieves only a weak order convergence rate of $O(h)$. It is interesting to note however, that [Fig. 5.5c](#) indicates that the fake Lévy area needs only to match the mean and variance of true Lévy area to obtain improved variance reduction at each level. Even the “Strang-T” method has the same variance reduction rate as the more sophisticated techniques despite having poor empirical error.

In practice, one usually wishes to obtain some target root mean-squared error (RMSE) with minimal computational cost. In this setting, for two numerical schemes with variance reduction $O(h^\beta)$ with $\beta > 1$, the scheme with higher order weak convergence is not necessarily the preferred one. By the complexity theorem of Giles [[Gil08](#), Theorem 3.1], the optimal number of sample paths on each level should be asymptotically proportional to $O(h_l^{(\beta+1)/2})$. As such, the computational effort should be expended mostly on the coarse levels in the regime $\beta > 1$, driving one towards discretisations that are computationally cheap on the lower levels. Since it is difficult to measure the computational complexity of the Strang log-ODE scheme with fake Lévy area produced by a generative model, we use the following approach. We implement the algorithm of Giles [[Gil08](#), Section 5] for the Milstein

scheme and Strang log-ODE scheme and compare the average time taken to achieve a selection of target RMSEs between 0.1 and 0.0025.

Table 5.4: The average time taken across 25 runs for the Milstein and Strang log-ODE methods to attain a target RMSE. We use the algorithm of [Gil08, Section 5] to determine the number of samples on each level and the stopping condition on the number of levels used on each run. All random variables are generated in torch on GPU, with the numerical schemes implemented in numpy on CPU.

RMSE	0.1	0.0441	0.0129	0.0086	0.0057	0.0038	0.0025
Milstein (s)	0.0097	0.0256	0.376	1.03	2.86	8.63	23.6
LévyGAN (s)	0.0102	0.0128	0.142	0.311	0.806	2.25	5.83

Remark 5.5.1. We see from Fig. 5.5c that the the Strang-antithetic scheme achieves the highest order variance reduction with a first order weak error rate. As noted above, both factors play a role in the overall time required to achieve a desired RMSE and it is possible that the higher order variance reduction of the Strang-antithetic scheme may outperform the high order weak error of the LévyGAN based approach. However, it was recently shown in [Igu+24] that the antithetic method can be combined with weak estimators of Lévy area to achieve high order variance reduction. We believe that combining their approach with ours would result in a scheme with both high order variance reduction and weak error rate.

We conclude this section by reiterating that the use of extra random variables to attain higher order weak convergence has become a popular technique, see for example [Tal90; NV08; NN09]. However, to the best of our knowledge it has not yet been observed that the use of fake Lévy area combined with standard multilevel Monte-Carlo can also achieve high order weak convergence and variance reduction.

5.6 Conclusion and open directions

While stochastic analysis techniques are often used in generative deep learning, this chapter appears to be one of the first examples where deep learning methodology has provided meaningful results in an application to stochastic analysis. Indeed, we have demonstrated a proof of concept that the techniques used in LévyGAN have a place in the field of numerical solutions to SDEs. However, we remark here that careful consideration of the domain-specific analytical properties was required. In particular, regardless of the network size or architecture, LévyGAN in its initial form was an order of magnitude less accurate without the inclusion of both bridge-flipping and Pair-net.

One open direction for future research is a careful analysis of the conditions required on the fake Lévy area in order to achieve optimal convergence rates for the MLMC scheme discussed in [Section 5.5.2](#). More applications of fake Lévy area may be found in the field of stochastic flows and accurate approximations to the level two rough path of Brownian motion.

An application of particular interest would be a GAN-based adaptive SDE solver. That is, a method that first generates a coarsely discretised path before checking whether the step-size of the solver should be reduced. Such functionality is desirable for use in Neural SDEs [[Li+20b](#); [Kid+21a](#); [Kid+21b](#)] and Logsig-RNN generators [[Ni+21](#)], which are both powerful methods for modelling noisy time series data. In the context of Lévy area generation, this would require the ability to generate Lévy area and Brownian increments over two half intervals given the Lévy area and Brownian increment over the larger interval. One approach would be to use the analytical characteristic function given in [[GQ12](#)] which provides the joint characteristic function evaluated at multiple time points. However, we expect the training time to be rather long, since the evaluation of the characteristic function involves solving a recursive system of matrix Riccati equations in addition to a system of independent linear matrix ODEs of order one.

Finally, it is possible to extend the Chen-training approach. This might take several forms: one might derive a Chen type relation for higher order terms in the polynomial expansion of Brownian motion (e.g. for H and b); use the ordinary Chen's relation for the generation of higher order terms in the log-signature of Brownian motion; or generating Lévy areas for certain Lévy processes.

For the third application, our approach may be applicable to the generation of Lévy areas of α -stable Lévy processes, where moment matching approaches are not possible, since these processes have unbounded variance for $\alpha < 2$. Indeed, the α -stable Lévy process X_t^α satisfies the scaling property $X_t^\alpha = t^{1/\alpha} X_1^\alpha$ in addition to independent and stationary increments. Combining these properties together, the distribution of its Lévy area (defined using Itô integration with jumps) should also be invariant under a suitably re-scaled version of Chen's relation. One could then attempt to train a network under this Chen relation, analogous to our approach for Brownian motion.

– BIBLIOGRAPHY –

- [AP20] B. Adlam & J. Pennington The neural tangent kernel in high dimensions: triple descent and a multi-scale theory of generalization. In: *International Conference on Machine Learning*. PMLR. 2020, pp. 74–84.
- [AGZ10] G. W. Anderson, A. Guionnet & O. Zeitouni *An introduction to random matrices*. Vol. 118. Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge, 2010, pp. xiv+492. ISBN: 978-0-521-19452-5.
- [Bay+23] C. Bayer, P. P. Hager, S. Riedel & J. Schoenmakers Optimal stopping with signatures. *Ann. Appl. Probab.* 33 (1) (2023), 238–273. ISSN: 1050-5164,2168-8737. DOI: [10.1214/22-aap1814](https://doi.org/10.1214/22-aap1814).
- [BDG01] G. Ben Arous, A. Dembo & A. Guionnet Aging of spherical spin glasses. *Probab. Theory Related Fields*. 120 (1) (2001), 1–67. ISSN: 0178-8051,1432-2064. DOI: [10.1007/PL00008774](https://doi.org/10.1007/PL00008774).
- [BG97] G. Ben Arous & A. Guionnet Large deviations for Wigner’s law and Voiculescu’s non-commutative entropy. *Probab. Theory Related Fields*. 108 (4) (1997), 517–542. ISSN: 0178-8051,1432-2064. DOI: [10.1007/s004400050119](https://doi.org/10.1007/s004400050119).
- [BP21] L. Benigni & S. Péché Eigenvalue distribution of some nonlinear models of random matrices. *Electron. J. Probab.* 26 (2021), Paper No. 150, 37. ISSN: 1083-6489. DOI: [10.1214/21-ejp699](https://doi.org/10.1214/21-ejp699).
- [Bil68] P. Billingsley *Convergence of probability measures*. John Wiley & Sons, Inc., New York-London-Sydney, 1968, pp. xii+253.
- [BL19] S. Bobkov & M. Ledoux One-dimensional empirical measures, order statistics, and Kantorovich transport distances. *Mem. Amer. Math. Soc.* 261 (1259) (2019), v+126. ISSN: 0065-9266,1947-6221. DOI: [10.1090/memo/1259](https://doi.org/10.1090/memo/1259).

- [BG15] H. Boedihardjo & X. Geng The uniqueness of signature problem in the non-Markov setting. *Stochastic Process. Appl.* 125 (12) (2015), 4674–4701. ISSN: 0304-4149,1879-209X. DOI: [10.1016/j.spa.2015.07.012](https://doi.org/10.1016/j.spa.2015.07.012).
- [BG19a] H. Boedihardjo & X. Geng Tail asymptotics of the Brownian signature. *Trans. Amer. Math. Soc.* 372 (1) (2019), 585–614. ISSN: 0002-9947,1088-6850. DOI: [10.1090/tran/7683](https://doi.org/10.1090/tran/7683).
- [Boe+21] H. Boedihardjo, J. Diehl, M. Mezzarobba & H. Ni The expected signature of Brownian motion stopped on the boundary of a circle has finite radius of convergence. *Bull. Lond. Math. Soc.* 53 (1) (2021), 285–299. ISSN: 0024-6093,1469-2120. DOI: [10.1112/blms.12420](https://doi.org/10.1112/blms.12420).
- [BG19b] H. Boedihardjo & X. Geng A non-vanishing property for the signature of a path. *C. R. Math. Acad. Sci. Paris.* 357 (2) (2019), 120–129. ISSN: 1631-073X,1778-3569. DOI: [10.1016/j.crma.2018.12.006](https://doi.org/10.1016/j.crma.2018.12.006).
- [BG23] H. Boedihardjo & X. Geng $SL_2(\mathbb{R})$ -developments and signature asymptotics for planar paths with bounded variation. *Rev. Mat. Iberoam.* 39 (5) (2023), 1973–2006. ISSN: 0213-2230,2235-0616. DOI: [10.4171/rmi/1397](https://doi.org/10.4171/rmi/1397).
- [Boe+16] H. Boedihardjo, X. Geng, T. Lyons & D. Yang The signature of a rough path: uniqueness. *Adv. Math.* 293 (2016), 720–737. ISSN: 0001-8708,1090-2082. DOI: [10.1016/j.aim.2016.02.011](https://doi.org/10.1016/j.aim.2016.02.011).
- [BGS20] H. Boedihardjo, X. Geng & N. P. Souris Path developments and tail asymptotics of signature for pure rough paths. *Adv. Math.* 364 (2020), 107043, 48. ISSN: 0001-8708,1090-2082. DOI: [10.1016/j.aim.2020.107043](https://doi.org/10.1016/j.aim.2020.107043).
- [Bog07] V. I. Bogachev *Measure theory. Vol. I, II.* Springer-Verlag, Berlin, 2007, Vol. I: xviii+500 pp., Vol. II: xiv+575. DOI: [10.1007/978-3-540-34514-5](https://doi.org/10.1007/978-3-540-34514-5).
- [BFH09] E. Breuillard, P. Friz & M. Huesmann From random walks to rough paths. *Proc. Amer. Math. Soc.* 137 (10) (2009), 3487–3496. ISSN: 0002-9939,1088-6826. DOI: [10.1090/S0002-9939-09-09930-4](https://doi.org/10.1090/S0002-9939-09-09930-4).
- [Bro+20] A. Browning, D. Warne, K. Burrage, R. Baker & M. Simpson Identifiability analysis for stochastic differential equation models in systems biology. *Journal of the Royal Society Interface.* 17 (173) (2020).
- [Cas+16] T. Cass, B. K. Driver, N. Lim & C. Litterer On the integration of weakly geometric rough paths. *J. Math. Soc. Japan.* 68 (4) (2016), 1505–1524. ISSN: 0025-5645,1881-1167. DOI: [10.2969/jmsj/06841505](https://doi.org/10.2969/jmsj/06841505).

- [CLX24] T. Cass, T. Lyons & X. Xu Weighted signature kernels. *Ann. Appl. Probab.* 34 (1A) (2024), 585–626. ISSN: 1050-5164,2168-8737. DOI: [10.1214/23-aap1973](https://doi.org/10.1214/23-aap1973).
- [CMT23] T. Cass, R. Messadene & W. F. Turner Signature asymptotics, empirical processes, and optimal transport. *Electron. J. Probab.* 28 (2023), Paper No. 150, 19. ISSN: 1083-6489. DOI: [10.1214/23-ejp1048](https://doi.org/10.1214/23-ejp1048).
- [CS24] T. Cass & C. Salvi *Lecture notes on rough paths and applications to machine learning*. 2024. ArXiv. [Preprint] Available from: [2404.06583](https://arxiv.org/abs/2404.06583).
- [CT24a] T. Cass & W. F. Turner *Free probability, path developments and signature kernels as universal scaling limits*. 2024. ArXiv. [Preprint] Available from: [2402.12311](https://arxiv.org/abs/2402.12311) [math.PR]. Available from: URL: <https://arxiv.org/abs/2402.12311>.
- [CT24b] T. Cass & W. F. Turner Topologies on unparameterised path space. *J. Funct. Anal.* 286 (4) (2024), Paper No. 110261, 28. ISSN: 0022-1236,1096-0783. DOI: [10.1016/j.jfa.2023.110261](https://doi.org/10.1016/j.jfa.2023.110261).
- [CT24c] T. Cass & W. F. Turner *Topologies on unparameterised rough path space*. 2024. ArXiv. [Preprint] Available from: [2407.17828](https://arxiv.org/abs/2407.17828).
- [CLN18a] J. Chang, T. Lyons & H. Ni Corrigendum to: super-multiplicativity and a lower bound for the decay of the signature of a path of finite length. *C. R. Math. Acad. Sci. Paris.* 356 (7) (2018), 720–724. ISSN: 1631-073X,1778-3569. DOI: [10.1016/j.crma.2018.05.010](https://doi.org/10.1016/j.crma.2018.05.010).
- [CLN18b] J. Chang, T. Lyons & H. Ni Super-multiplicativity and a lower bound for the decay of the signature of a path of finite length. *C. R. Math. Acad. Sci. Paris.* 356 (7) (2018), 720–724. ISSN: 1631-073X,1778-3569. DOI: [10.1016/j.crma.2018.05.010](https://doi.org/10.1016/j.crma.2018.05.010).
- [Che+24a] J. Chemseddine, P. Hagemann, G. Steidl & C. Wald *Conditional wasserstein distances with applications in bayesian ot flow matching*. 2024. ArXiv. [Preprint] Available from: [2403.18705](https://arxiv.org/abs/2403.18705).
- [Che+24b] C.-F. Chen, A. M. Dalzell, M. Berta, F. G. Brandão & J. A. Tropp Sparse random hamiltonians are quantumly easy. *Physical Review X.* 14 (1) (2024), 011014.
- [Che54] K.-T. Chen Iterated integrals and exponential homomorphisms. *Proc. London Math. Soc.* (3). 4 (1954), 502–512. ISSN: 0024-6115,1460-244X. DOI: [10.1112/plms/s3-4.1.502](https://doi.org/10.1112/plms/s3-4.1.502).

- [Che57] K.-T. Chen Integration of paths, geometric invariants and a generalized Baker-Hausdorff formula. *Ann. of Math. (2)*. 65 (1957), 163–178. ISSN: 0003-486X. DOI: [10.2307/1969671](https://doi.org/10.2307/1969671).
- [Che16] I. Chevyrev *Characteristic functions of path signatures and applications*. PhD thesis. University of Oxford, 2016.
- [CF19] I. Chevyrev & P. K. Friz Canonical RDEs and general semimartingales as rough paths. *Ann. Probab.* 47 (1) (2019), 420–463. ISSN: 0091-1798,2168-894X. DOI: [10.1214/18-AOP1264](https://doi.org/10.1214/18-AOP1264).
- [CL16] I. Chevyrev & T. Lyons Characteristic functions of measures on geometric rough paths. *Ann. Probab.* 44 (6) (2016), 4049–4082. ISSN: 0091-1798,2168-894X. DOI: [10.1214/15-AOP1068](https://doi.org/10.1214/15-AOP1068).
- [CO22] I. Chevyrev & H. Oberhauser Signature moments to characterize laws of stochastic processes. *J. Mach. Learn. Res.* 23 (2022), Paper No. [176], 42. ISSN: 1532-4435,1533-7928.
- [CC80] J. M. C. Clark & R. J. Cameron The maximum rate of convergence of discrete approximations for stochastic differential equations. In: *Stochastic differential systems (Proc. IFIP-WG 7/1 Working Conf., Vilnius, 1978)*. Vol. 25. Lect. Notes Control Inf. Sci. Springer, Berlin-New York, 1980, pp. 162–171. ISBN: 3-540-10498-4.
- [Cre+25] S. Crew, C. Salvi, W. F. Turner, T. Cass & A. Jacquier *Quantum path signatures*. 2025. ArXiv. [Preprint] Available from: [2508.05103](https://arxiv.org/abs/2508.05103).
- [Cri15] R. Crisostomo *An analysis of the heston stochastic volatility model: implementation and calibration using matlab*. 2015. ArXiv. [Preprint] Available from: [1502.02963](https://arxiv.org/abs/1502.02963).
- [Csö83] M. Csörgő *Quantile processes with statistical applications*. Vol. 42. CBMS-NSF Regional Conference Series in Applied Mathematics. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1983, pp. xiii+156. ISBN: 0-89871-185-1. DOI: [10.1137/1.9781611970289](https://doi.org/10.1137/1.9781611970289).
- [CR78] M. Csörgő & P. Révész Strong approximations of the quantile process. *Ann. Statist.* 6 (4) (1978), 882–894. ISSN: 0090-5364,2168-8966.
- [Csö81a] S. Csörgő Limit behaviour of the empirical characteristic function. *Ann. Probab.* 9 (1) (1981), 130–144. ISSN: 0091-1798,2168-894X.

- [Csö81b] S. Csörgő Multivariate empirical characteristic functions. *Z. Wahrsch. Verw. Gebiete*. 55 (2) (1981), 203–229. ISSN: 0044-3719. DOI: [10.1007/BF00535160](https://doi.org/10.1007/BF00535160).
- [CGP25] C. Cuchiero, X. Guo & F. Primavera *Functional itô-formula and taylor expansions for non-anticipative maps of càdlàg rough paths*. 2025. ArXiv. [Preprint] Available from: [2504.06164](https://arxiv.org/abs/2504.06164).
- [Dav14] A. Davie KMT theory applied to approximations of SDE. In: *Stochastic analysis and applications 2014*. Vol. 100. Springer Proc. Math. Stat. Springer, Cham, 2014, pp. 185–201. DOI: [10.1007/978-3-319-11292-3_7](https://doi.org/10.1007/978-3-319-11292-3_7).
- [dBGU05] E. del Barrio, E. Giné & F. Utzet Asymptotics for L_2 functionals of the empirical quantile process, with applications to tests of fit based on weighted Wasserstein distances. *Bernoulli*. 11 (1) (2005), 131–189. ISSN: 1350-7265,1573-9759. DOI: [10.3150/bj/1110228245](https://doi.org/10.3150/bj/1110228245).
- [Dic07] A. S. Dickinson Optimal approximation of the second iterated integral of Brownian motion. *Stoch. Anal. Appl.* 25 (5) (2007), 1109–1128. ISSN: 0736-2994,1532-9356. DOI: [10.1080/07362990701540592](https://doi.org/10.1080/07362990701540592).
- [DET20a] J. Diehl, K. Ebrahimi-Fard & N. Tapia Iterated-sums signature, quasisymmetric functions and time series analysis. *Sém. Lothar. Combin.* 84B (2020), Art. 86, 12. ISSN: 1286-4889.
- [DET20b] J. Diehl, K. Ebrahimi-Fard & N. Tapia Time-warping invariants of multidimensional time series. *Acta Appl. Math.* 170 (2020), 265–290.
- [Dri13] B. K. Driver *Rough path analysis*. Lecture notes available at <http://www.math.ucsd.edu/bdriver>. 2013.
- [DP21] G. Dubach & Y. Peled On words of non-Hermitian random matrices. *Ann. Probab.* 49 (4) (2021), 1886–1916. ISSN: 0091-1798,2168-894X. DOI: [10.1214/20-aop1496](https://doi.org/10.1214/20-aop1496).
- [DT23] B. Dupire & V. Tisserot-Daguet *Functional expansions*. 2023. ArXiv. [Preprint] Available from: [2212.13628](https://arxiv.org/abs/2212.13628).
- [EKR15] B. Eynard, T. Kimura & S. Ribault *Random matrices*. 2015. ArXiv. [Preprint] Available from: [1510.04430](https://arxiv.org/abs/1510.04430).
- [FSS20] A. Fatir Ansari, J. Scarlett & H. Soh A characteristic function approach to deep implicit generative modeling. In: *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2020.

- [Faw03] T. Fawcett *Problems in stochastic analysis: connections between rough paths and non-commutative harmonic analysis*. PhD thesis. University of Oxford, 2003.
- [FM77] A. Feuerverger & R. A. Mureika The empirical characteristic function and its applications. *Ann. Statist.* 5 (1) (1977), 88–97. ISSN: 0090-5364,2168-8966.
- [FL15] G. Flint & T. Lyons *Pathwise approximation of sdes by coupling piecewise abelian rough paths*. 2015. ArXiv. [Preprint] Available from: [1505.01298](https://arxiv.org/abs/1505.01298).
- [FH23] J. Foster & K. Habermann Brownian bridge expansions for Lévy area approximations and particular values of the Riemann zeta function. *Combin. Probab. Comput.* 32 (3) (2023), 370–397. ISSN: 0963-5483,1469-2163. DOI: [10.1017/s096354832200030x](https://doi.org/10.1017/s096354832200030x).
- [FLO20] J. Foster, T. Lyons & H. Oberhauser An optimal polynomial approximation of Brownian motion. *SIAM J. Numer. Anal.* 58 (3) (2020), 1393–1421. ISSN: 0036-1429,1095-7170. DOI: [10.1137/19M1261912](https://doi.org/10.1137/19M1261912).
- [FdRS24] J. M. Foster, G. dos Reis & C. Strange High order splitting methods for SDEs satisfying a commutativity condition. *SIAM J. Numer. Anal.* 62 (1) (2024), 500–532. ISSN: 0036-1429,1095-7170. DOI: [10.1137 / 23M155147X](https://doi.org/10.1137/23M155147X).
- [Fos20] j. Foster *Numerical approximations for stochastic differential equations*. PhD thesis. University of Oxford, 2020.
- [FH14] P. K. Friz & M. Hairer *A course on rough paths*. Universitext. With an introduction to regularity structures. Springer, Cham, 2014, pp. xiv+251. DOI: [10.1007/978-3-319-08332-2](https://doi.org/10.1007/978-3-319-08332-2).
- [FV10] P. K. Friz & N. B. Victoir *Multidimensional stochastic processes as rough paths*. Vol. 120. Cambridge Studies in Advanced Mathematics. Theory and applications. Cambridge University Press, Cambridge, 2010, pp. xiv+656. DOI: [10.1017/CBO9780511845079](https://doi.org/10.1017/CBO9780511845079).
- [GL94] J. G. Gaines & T. J. Lyons Random generation of stochastic area integrals. *SIAM J. Appl. Math.* 54 (4) (1994), 1132–1146. ISSN: 0036-1399. DOI: [10.1137/S0036139992235706](https://doi.org/10.1137/S0036139992235706).
- [Gar18] D. J. H. Garling *Analysis on Polish spaces and an introduction to optimal transportation*. Vol. 89. London Mathematical Society Student Texts. Cambridge University Press, Cambridge, 2018, pp. ix+348.

- [Ge+21] J. Ge, Y.-C. Liang, Z. Bai & G. Pan Large-dimensional random matrix theory and its applications in deep learning and wireless communications. *Random Matrices Theory Appl.* 10 (4) (2021), Paper No. 2230001, 72. ISSN: 2010-3263,2010-3271. DOI: [10.1142/S2010326322300017](https://doi.org/10.1142/S2010326322300017).
- [GQ16] X. Geng & Z. Qian On an inversion theorem for Stratonovich's signatures of multidimensional diffusion paths. *Ann. Inst. Henri Poincaré Probab. Stat.* 52 (1) (2016), 429–447. ISSN: 0246-0203,1778-7017. DOI: [10.1214/14-AIHP644](https://doi.org/10.1214/14-AIHP644).
- [Gen15] X. Geng *The signature of a rough path: uniqueness*. PhD thesis. University of Oxford, 2015.
- [Gen17] X. Geng Reconstruction for the signature of a rough path. *Proc. Lond. Math. Soc.* (3). 114 (3) (2017), 495–526. ISSN: 0024-6115,1460-244X. DOI: [10.1112/plms.12013](https://doi.org/10.1112/plms.12013).
- [Gen] X. Geng *An introduction to the theory of rough paths*. Available from: URL: <https://researchers.ms.unimelb.edu.au/~xgge@unimelb/Files/Notes/Rough%20Path%20Theory.pdf>.
- [GQ12] X. Geng & Z. Qian *On the finite dimensional joint characteristic function of lévy's stochastic area processes*. 2012. ArXiv. [Preprint] Available from: [1206.1241](https://arxiv.org/abs/1206.1241).
- [Gil08] M. B. Giles Multilevel Monte Carlo path simulation. *Oper. Res.* 56 (3) (2008), 607–617. ISSN: 0030-364X,1526-5463. DOI: [10.1287/opre.1070.0496](https://doi.org/10.1287/opre.1070.0496).
- [GS14a] M. B. Giles & L. Szpruch Antithetic multilevel Monte Carlo estimation for multi-dimensional SDEs without Lévy area simulation. *Ann. Appl. Probab.* 24 (4) (2014), 1585–1620. ISSN: 1050-5164,2168-8737. DOI: [10.1214/13-AAP957](https://doi.org/10.1214/13-AAP957).
- [Gil71] R. Giles A generalization of the strict topology. *Trans. Amer. Math. Soc.* 161 (1971), 467–474. ISSN: 0002-9947,1088-6850. DOI: [10.2307/1995954](https://doi.org/10.2307/1995954).
- [Goo+14] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville & Y. Bengio Generative adversarial nets. In: *Advances in Neural Information Processing Systems*. Vol. 27. 2014.
- [GS14b] A. Guionnet & D. Shlyakhtenko Free monotone transport. *Invent. Math.* 197 (3) (2014), 613–661.

- [Gui19] A. Guionnet *Asymptotics of random matrices and related models*. Vol. 130. CBMS Regional Conference Series in Mathematics. The uses of Dyson-Schwinger equations, Published for the Conference Board of the Mathematical Sciences. American Mathematical Society, Providence, RI, 2019, pp. vii+143. ISBN: 978-1-4704-5027-4. DOI: [10.1090/cbms/130](https://doi.org/10.1090/cbms/130).
- [GM06] A. Guionnet & E. Maurel-Segala Combinatorial aspects of matrix models. *ALEA Lat. Am. J. Probab. Math. Stat.* 1 (2006), 241–279. ISSN: 1980-0436.
- [GM07] A. Guionnet & E. Maurel-Segala Second order asymptotics for matrix models. *Ann. Probab.* 35 (6) (2007), 2160–2212. ISSN: 0091-1798,2168-894X. DOI: [10.1214/009117907000000141](https://doi.org/10.1214/009117907000000141).
- [GS09] A. Guionnet & D. Shlyakhtenko Free diffusions and matrix models with strictly convex interaction. *Geom. Funct. Anal.* 18 (6) (2009), 1875–1916. ISSN: 1016-443X,1420-8970.
- [HL10] B. Hambly & T. Lyons Uniqueness for the signature of a path of bounded variation and the reduced path group. *Ann. of Math. (2)*. 171 (1) (2010), 109–167. ISSN: 0003-486X,1939-8980. DOI: [10.4007/annals.2010.171.109](https://doi.org/10.4007/annals.2010.171.109).
- [HS83] K. Helmes & A. Schwane Lévy’s stochastic area formula in higher dimensions. *J. Funct. Anal.* 54 (2) (1983), 177–192. ISSN: 0022-1236. DOI: [10.1016/0022-1236\(83\)90053-8](https://doi.org/10.1016/0022-1236(83)90053-8).
- [Hoc78] K.J. Hochberg A signed measure on path space related to Wiener measure. *Ann. Probab.* 6 (3) (1978), 433–458. ISSN: 0091-1798,2168-894X.
- [Igu+24] Y. Iguchi, A. Jasra, M. Maama & A. Beskos *Antithetic multilevel methods for elliptic and hypo-elliptic diffusions with applications*. 2024. ArXiv. [Preprint] Available from: [2403.13489](https://arxiv.org/abs/2403.13489).
- [Iss+23] Z. Issa, B. Horvath, M. Lemercier & C. Salvi *Non-adversarial training of neural sdes with signature kernel scores*. 2023. ArXiv. [Preprint] Available from: [2305.16274](https://arxiv.org/abs/2305.16274).
- [Jel+23] A. Jelinčič, J. Tao, W. F. Turner, T. Cass, J. Foster & H. Ni *Generative modelling of lévy area for high order sde simulation*. 2023. ArXiv. [Preprint] Available from: [2308.02452](https://arxiv.org/abs/2308.02452).

- [KR23] F. Kastner & A. Rößler An analysis of approximation algorithms for iterated stochastic integrals and a Julia and MATLAB simulation toolbox. *Numer. Algorithms*. 93 (1) (2023), 27–66. ISSN: 1017-1398,1572-9265. DOI: [10.1007/s11075-022-01401-z](https://doi.org/10.1007/s11075-022-01401-z).
- [Kec95] A. S. Kechris *Classical descriptive set theory*. Vol. 156. Graduate Texts in Mathematics. Springer-Verlag, New York, 1995, pp. xviii+402. DOI: [10.1007/978-1-4612-4190-4](https://doi.org/10.1007/978-1-4612-4190-4).
- [Kid+21a] P. Kidger, J. Foster, X. Li & T. Lyons Efficient and Accurate Gradients for Neural SDEs. In: *Advances in Neural Information Processing Systems*. Vol. 34. 2021, pp. 18747–18761.
- [Kid+21b] P. Kidger, J. Foster, X. Li & T.J. Lyons Neural sdes as infinite-dimensional gans. In: *Proceedings of the 38th International Conference on Machine Learning*. Vol. 139. Proceedings of Machine Learning Research. PMLR, 2021, pp. 5453–5463.
- [KO19] F. J. Király & H. Oberhauser Kernels for sequentially ordered data. *J. Mach. Learn. Res.* 20 (2019), Paper No. 31, 45. ISSN: 1532-4435,1533-7928.
- [KPW92] P. E. Kloeden, E. Platen & I. W. Wright The approximation of multiple stochastic integrals. *Stochastic Anal. Appl.* 10 (4) (1992), 431–441. ISSN: 0736-2994,1532-9356. DOI: [10.1080/07362999208809281](https://doi.org/10.1080/07362999208809281).
- [KP92] P. E. Kloeden & E. Platen *Numerical solution of stochastic differential equations*. Vol. 23. Applications of Mathematics (New York). Springer-Verlag, Berlin, 1992, pp. xxxvi+632. ISBN: 3-540-54062-8. DOI: [10.1007/978-3-662-12616-5](https://doi.org/10.1007/978-3-662-12616-5).
- [Kry60] V. J. Krylov Some properties of the distribution corresponding to the equation $\partial u / \partial t = (-1)^{q+1} \partial^{2q} u / \partial x^{2q}$. *Dokl. Akad. Nauk SSSR*. 132 (1960), 1254–1257.
- [LQ13] Y. Le Jan & Z. Qian Stratonovich’s signatures of Brownian motion determine Brownian sample paths. *Probab. Theory Related Fields*. 157 (1-2) (2013), 209–223. ISSN: 0178-8051,1432-2064. DOI: [10.1007/s00440-012-0454-z](https://doi.org/10.1007/s00440-012-0454-z).
- [LM15] B. Leimkuhler & C. Matthews *Molecular Dynamics: With Deterministic and Stochastic Numerical Methods*. Interdisciplinary Applied Mathematics, Springer, 2015.

- [LL09] D. Levin & T. Lyons A signed measure on rough paths associated to a PDE of high order: results and conjectures. *Rev. Mat. Iberoam.* 25 (3) (2009), 971–994. ISSN: 0213-2230,2235-0616. DOI: [10.4171/rmi/587](https://doi.org/10.4171/rmi/587).
- [LLN16] D. Levin, T. Lyons & H. Ni *Learning from the past, predicting the statistics for the future, learning an evolving system*. 2016. ArXiv. [Preprint] Available from: [1309.0260](https://arxiv.org/abs/1309.0260).
- [Li+20a] S. Li, Z. Yu, M. Xiang & D. Mandic Reciprocal adversarial learning via characteristic functions. In: *Advances in Neural Information Processing Systems*. Vol. 33. 2020, pp. 217–228.
- [Li+25] S. Li, Z. Lyu, H. Ni & J. Tao *Restricted path characteristic function determines the law of stochastic processes*. 2025. ArXiv. [Preprint] Available from: [2404.18661](https://arxiv.org/abs/2404.18661).
- [Li+20b] X. Li, T.-K. L. Wong, R. T. Q. Chen & D. Duvenaud Scalable gradients and variational inference for stochastic differential equations. In: *Symposium on Advances in Approximate Bayesian Inference*. Vol. 118. Proc. Mach. Learn. Res. (PMLR). Proceedings of Machine Learning Research PMLR, [place of publication not identified], 2020, p. 28.
- [Li+19] X. Li, Y. Wu, L. Mackey & M. A. Erdogdu Stochastic runge-kutta accelerates langevin monte carlo and beyond. In: *Advances in Neural Information Processing Systems*. Vol. 32. Curran Associates, Inc., 2019.
- [Loè78] M. Loève *Probability theory. II. Fourth. Vol.* Vol. 46. Graduate Texts in Mathematics. Springer-Verlag, New York-Heidelberg, 1978, pp. xvi+413. ISBN: 0-387-90262-7.
- [LLN23] H. Lou, S. Li & H. Ni PCF-GAN: generating sequential data via the characteristic function of measures on the path space. In: *Thirty-seventh Conference on Neural Information Processing Systems*. 2023.
- [LLN24] H. Lou, S. Li & H. Ni Path development network with finite-dimensional lie group. *Transactions on Machine Learning Research*. (2024). ISSN: 2835-8856.
- [LNT24] T. Lyons, H. Ni & J. Tao *A pde approach for solving the characteristic function of the generalised signature process*. 2024. ArXiv. [Preprint] Available from: [2401.02393](https://arxiv.org/abs/2401.02393) [[math.PR](https://arxiv.org/abs/2401.02393)].

- [Lyo98] T. J. Lyons Differential equations driven by rough signals. *Rev. Mat. Iberoamericana*. 14 (2) (1998), 215–310. ISSN: 0213-2230. DOI: [10.4171/RMI/240](https://doi.org/10.4171/RMI/240).
- [Lyo+07] T. J. Lyons, F. École d’été de probabilités de Saint-Flour (34th : 2004 : Saint-Flour, M. J. Caruana & T. Lévy *Differential equations driven by rough paths : ecole d’été de probabilités de saint-flour xxxiv-2004*. eng. 1st ed. 2007. École d’Été de Probabilités de Saint-Flour, 1908. Berlin, Heidelberg, Springer Berlin Heidelberg, 2007.
- [LX17] T. J. Lyons & W. Xu Hyperbolic development and inversion of signature. *J. Funct. Anal.* 272 (7) (2017), 2933–2955. ISSN: 0022-1236,1096-0783. DOI: [10.1016/j.jfa.2016.12.024](https://doi.org/10.1016/j.jfa.2016.12.024).
- [LX18] T. J. Lyons & W. Xu Inverting the signature of a path. *J. Eur. Math. Soc. (JEMS)*. 20 (7) (2018), 1655–1687. ISSN: 1435-9855,1435-9863. DOI: [10.4171/JEMS/796](https://doi.org/10.4171/JEMS/796).
- [MT21] G. N. Milstein & M. V. Tretyakov *Stochastic numerics for mathematical physics*. Second. Scientific Computation. Springer, Cham, 2021, pp. xxv+736. DOI: [10.1007/978-3-030-82040-4](https://doi.org/10.1007/978-3-030-82040-4).
- [MO14] M. Mirza & S. Osindero *Conditional generative adversarial nets*. 2014. ArXiv. [Preprint] Available from: [1411.1784](https://arxiv.org/abs/1411.1784).
- [MR22] J. Mrongowius & A. Rößler On the approximation and simulation of iterated stochastic integrals and the corresponding Lévy areas in terms of a multidimensional Brownian motion. *Stoch. Anal. Appl.* 40 (3) (2022), 397–425. ISSN: 0736-2994,1532-9356. DOI: [10.1080/07362994.2021.1922291](https://doi.org/10.1080/07362994.2021.1922291).
- [MLS23] N. Muca Cirone, M. Lemerrier & C. Salvi Neural signature kernels as infinite-width-depth-limits of controlled ResNets. In: *Proceedings of the 40th International Conference on Machine Learning*. Vol. 202. Proceedings of Machine Learning Research. PMLR, 2023, pp. 25358–25425.
- [Nag50] J.-i. Nagata On a necessary and sufficient condition of metrizable. *J. Inst. Polytech. Osaka City Univ. Ser. A*. 1 (1950), 93–100. ISSN: 0388-0516.
- [Nej18] S. Nejad *Lipschitz functions on unparameterised rough paths and the brownian motion associated to the bilaplacian*. PhD thesis. University of Oxford, 2018.

- [Ni+21] H. Ni, L. Szpruch, M. Sabate-Vidales, B. Xiao, M. Wiese & S. Liao Sig-wasserstein gans for time series generation. In: *Proceedings of the Second ACM International Conference on AI in Finance*. 2021, pp. 1–8.
- [NN09] M. Ninomiya & S. Ninomiya A new higher-order weak approximation scheme for stochastic differential equations and the Runge-Kutta method. *Finance Stoch.* 13 (3) (2009), 415–443. ISSN: 0949-2984,1432-1122. DOI: [10.1007/s00780-009-0101-4](https://doi.org/10.1007/s00780-009-0101-4).
- [NV08] S. Ninomiya & N. Victoir Weak approximation of stochastic differential equations and application to derivative pricing. *Appl. Math. Finance.* 15 (1-2) (2008), 107–121. ISSN: 1350-486X,1466-4313. DOI: [10.1080/13504860701413958](https://doi.org/10.1080/13504860701413958).
- [PT85] É. Pardoux & D. Talay Discretization and simulation of stochastic differential equations. *Acta Appl. Math.* 3 (1) (1985), 23–47. ISSN: 0167-8019. DOI: [10.1007/BF01438265](https://doi.org/10.1007/BF01438265).
- [PSG17] J. Pennington, S. Schoenholz & S. Ganguli Resurrecting the sigmoid in deep learning through dynamical isometry: theory and practice. *Advances in neural information processing systems*. 30 (2017).
- [RY99] D. Revuz & M. Yor *Continuous martingales and Brownian motion*. Third. Vol. 293. Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Springer-Verlag, Berlin, 1999, pp. xiv+602. ISBN: 3-540-64325-7. DOI: [10.1007/978-3-662-06400-9](https://doi.org/10.1007/978-3-662-06400-9).
- [Röß09] A. Rößler Second order Runge-Kutta methods for Itô stochastic differential equations. *SIAM J. Numer. Anal.* 47 (3) (2009), 1713–1738. ISSN: 0036-1429,1095-7170. DOI: [10.1137/060673308](https://doi.org/10.1137/060673308).
- [Sal+21] C. Salvi, T. Cass, J. Foster, T. Lyons & W. Yang The signature kernel is the solution of a Goursat PDE. *SIAM J. Math. Data Sci.* 3 (3) (2021), 873–899. ISSN: 2577-0187. DOI: [10.1137/20M1366794](https://doi.org/10.1137/20M1366794).
- [Sca+09] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner & G. Monfardini The graph neural network model. *IEEE Transactions on Neural Networks.* 20 (1) (2009), 61–80.
- [Sho83] R. M. Shortt The extension of measurable functions. *Proc. Amer. Math. Soc.* 87 (3) (1983), 444–446. ISSN: 0002-9939,1088-6826. DOI: [10.2307/2043628](https://doi.org/10.2307/2043628).

- [Shr04] S. E. Shreve *Stochastic calculus for finance. II*. Springer Finance. Continuous-time models. Springer-Verlag, New York, 2004, pp. xx+550. ISBN: 0-387-40101-6.
- [Sri+10] B. K. Sriperumbudur, A. Gretton, K. Fukumizu, B. Schölkopf & G. R. G. Lanckriet Hilbert space embeddings and metrics on probability measures. *J. Mach. Learn. Res.* 11 (2010), 1517–1561. ISSN: 1532-4435, 1533-7928.
- [Tal90] D. Talay Second-order discretization schemes of stochastic differential systems for the computation of the invariant law. *Stochastics and Stochastic Reports.* 29 (1) (1990), 13–36.
- [TNL24] J. Tao, H. Ni & C. Liu *High rank path development: an approach of learning the filtration of stochastic processes*. 2024. ArXiv. [Preprint] Available from: [2405.14913](https://arxiv.org/abs/2405.14913).
- [Tao12] T. Tao *Topics in random matrix theory*. Vol. 132. Graduate Studies in Mathematics. American Mathematical Society, Providence, RI, 2012, pp. x+282. ISBN: 978-0-8218-7430-1. DOI: [10.1090/gsm/132](https://doi.org/10.1090/gsm/132).
- [Ter22] K. Terada *Higher order methods and multilevel monte carlo for sdes in finance*. Masters thesis, University of Edinburgh. 2022.
- [Ver18] R. Vershynin *High-dimensional probability*. Vol. 47. Cambridge Series in Statistical and Probabilistic Mathematics. An introduction with applications in data science, With a foreword by Sara van de Geer. Cambridge University Press, Cambridge, 2018, pp. xiv+284. ISBN: 978-1-108-41519-4. DOI: [10.1017/9781108231596](https://doi.org/10.1017/9781108231596).
- [Voi85] D. Voiculescu Symmetries of some reduced free product C^* -algebras. In: *Operator algebras and their connections with topology and ergodic theory (Buşteni, 1983)*. Vol. 1132. Lecture Notes in Math. Springer, Berlin, 1985, pp. 556–588. ISBN: 3-540-15643-7. DOI: [10.1007/BFb0074909](https://doi.org/10.1007/BFb0074909).
- [Voi86] D. Voiculescu Addition of certain noncommuting random variables. *J. Funct. Anal.* 66 (3) (1986), 323–346. ISSN: 0022-1236. DOI: [10.1016/0022-1236\(86\)90062-5](https://doi.org/10.1016/0022-1236(86)90062-5).
- [Voi87] D. Voiculescu Multiplication of certain noncommuting random variables. *J. Operator Theory.* 18 (2) (1987), 223–235. ISSN: 0379-4024.
- [Voi91] D. Voiculescu Limit laws for random matrices and free products. *Invent. Math.* 104 (1) (1991), 201–220. ISSN: 0020-9910, 1432-1297.

- [WZ23] Z. Wang & Y. Zhu *Deformed semicircle law and concentration of nonlinear random matrices for ultra-wide neural networks*. 2023. ArXiv. [Preprint] Available from: [2109.09304 \[math.ST\]](#).
- [Wik01] M. Wiktorsson Joint characteristic function and simultaneous simulation of iterated Itô integrals for multiple independent Brownian motions. *Ann. Appl. Probab.* 11 (2) (2001), 470–487. ISSN: 1050-5164,2168-8737. DOI: [10.1214/aoap/1015345301](#).
- [Yan+16] W. Yang, L. Jin, H. Ni & T. Lyons Rotation-free online handwritten character recognition using dyadic path signature features, hanging normalization, and deep neural network. In: *2016 23rd International Conference on Pattern Recognition (ICPR)*. 2016, pp. 4083–4088.

APPENDIX A

TECHNICAL RESULTS IN RANDOM MATRIX THEORY

This short appendix is dedicated to proving that the largest eigenvalue of the Gaussian Unitary Ensemble (GUE) has a sub-Gaussian tail, as required in [Chapter 3](#).

A.1 Sub-Gaussian tail of the GUE

The following lemma may be derived by keeping track of the constants in the various equivalent characterisations of sub-Gaussian random variables, as in [\[Ver18, Proposition 2.5.2\]](#) for example. We include a self-contained proof for completeness.

Lemma A.1.1. *Let $(X_N)_{N \geq 1}$ be a sequence of real-valued random variables. Then the following are equivalent*

- (1) *There exist $\alpha > 0$ and $t_0 < \infty$ such that for all $t \geq t_0$ and $N \geq 1$, $\mathbb{P}(X_N \geq t) \leq e^{-\alpha N t^2}$;*
- (2) *There exists $C > 0$ such that for all s sufficiently small and all $N \geq 1$, then $\mathbb{E}[\exp(s N X_N^2)] \leq C^N$.*

Proof. Assume (1), then

$$\begin{aligned}
\mathbb{E}[\exp(sNX_N^2)] &= \int_0^\infty \mathbb{P}(\exp(sNX_N^2) \geq t) dt \\
&\leq 1 + 2 \int_0^\infty \mathbb{P}(X_N \geq u) sNu e^{sNu^2} du \\
&= 1 + \int_0^{t_0} 2sNu e^{sNu^2} du + \int_{t_0}^\infty 2sNu e^{-(\alpha-s)Nu^2} du \\
&= 1 + e^{sNt_0^2} + \frac{s}{\alpha-s} \leq C^N,
\end{aligned}$$

for $C \geq 1 + e^{st_0^2} + \frac{s}{\alpha-s}$ and $s < \alpha$. Conversely, assume (2), then Chebyshev's inequality yields

$$\mathbb{P}(X_N \geq t) \leq e^{-sNt^2} \mathbb{E}[\exp(sNX_N^2)] \leq e^{-sNt^2} C^N \leq e^{-\alpha Nt^2},$$

whenever $\alpha < s$ and $t \geq t_0 \sqrt{C/(s-\alpha)}$. □

Recall that the spectrum of an $N \times N$ GUE matrix has the distribution

$$\sigma^N(d\lambda) = \frac{1}{Z^N} \prod_{1 \leq i < j \leq N} |\lambda_i - \lambda_j|^2 \exp \left\{ -\frac{N}{2} \sum_{i=1}^N \lambda_i^2 \right\} d\lambda, \quad (\text{A.1})$$

where Z^N is the normalisation constant given by

$$Z^N = (2\pi)^{\frac{N}{2}} N^{-\frac{N^2}{2}} \prod_{i=1}^N i!, \quad (\text{A.2})$$

see [BG97, Property 3.1] with $\beta = 2$. The following pair of lemmas are key.

Lemma A.1.2. *There exists $C > 0$ such that, for all $N \geq 2$, $\frac{Z^{N-1}}{Z^N} \leq e^{CN}$.*

Proof. From (A.2), we have

$$\frac{Z^{N-1}}{Z^N} = \frac{(N-1)^{-\frac{(N-1)^2}{2}} N^{\frac{N^2}{2}}}{\sqrt{2\pi} \Gamma(N+1)} = \frac{\exp \frac{1}{2} (N^2 \log(N) - (N-1)^2 \log(N-1))}{\sqrt{2\pi} N!}.$$

It can be checked that for every $N \geq 2$, $N^2 \log(N) - (N-1)^2 \log(N-1) \leq 2N \log(N) + 2N$, and hence

$$\frac{Z^{N-1}}{Z^N} \leq \frac{\exp N + N \log N}{\sqrt{2\pi} \Gamma(N+1)}.$$

Using the bound from Stirling's formula $N! \geq \sqrt{2\pi N} e^{N \log N - N}$, we obtain, for suitably large $C > 0$,

$$\frac{Z^{N-1}}{Z^N} \leq \frac{e^{2N}}{2\pi\sqrt{N}} \leq e^{CN}.$$

□

The proof of the next lemma follows the strategy of [BDG01, Lemma 6.2]. In the following we write $\underline{\lambda}^N \sim \sigma^N$ to mean that the vector $\underline{\lambda} := (\lambda_1, \dots, \lambda_N)$ is distributed according to σ^N , as defined in (A.1).

Lemma A.1.3 (Tails of the GUE). *Let $\underline{\lambda}^N \sim \sigma^N$, then there exists $\alpha > 0$ and $t_0 > 0$ such that for all $t \geq t_0 > 0$ and $N \in \mathbb{N}$, it holds that*

$$\mathbb{P}_{\sigma_N} \left(\max_{i=1, \dots, N} |\lambda_i| \geq t \right) \leq e^{-\alpha N t^2}. \quad (\text{A.3})$$

Proof. First note that for all $|x| \geq 8$ and $\lambda \in \mathbb{R}$, then

$$|x - \lambda|^2 e^{-\frac{\lambda^2}{2}} \leq 2(|x|^2 + |\lambda|^2) e^{-\frac{\lambda^2}{2}} \leq 4|x^2| \leq e^{\frac{x^2}{4}},$$

so that for all $|x| \geq t \geq 8$,

$$\begin{aligned} \mathbb{P}_{\sigma_N}(|\lambda_1| \geq t) &\leq \frac{Z^{N-1}}{Z^N} e^{-\frac{1}{4} N t^2} \int_{|x| \geq t} e^{-\frac{x^2}{4}} \int \prod_{i=2}^N |x - \lambda_i|^2 e^{-\frac{\lambda_i^2}{2} - \frac{x^2}{4}} \sigma^{N-1}(\mathrm{d}\underline{\lambda}^{N-1}) \mathrm{d}x \\ &\leq \frac{Z^{N-1}}{Z^N} e^{-\frac{1}{4} N t^2} \int e^{-\frac{x^2}{4}} \int e^{\frac{x^2}{4} - \frac{x^2}{4}} \sigma^{N-1}(\mathrm{d}\underline{\lambda}^{N-1}) \mathrm{d}x \\ &\leq \frac{Z^{N-1}}{Z^N} e^{-\frac{1}{4} N t^2} \int e^{-\frac{x^2}{4}} \mathrm{d}x \\ &\leq e^{CN - \frac{1}{4} N t^2}, \end{aligned}$$

for some $C > 0$, where the last inequality follows from Lemma A.1.2 and where we have used

$$e^{-N \frac{x^2}{2}} = e^{-N \frac{x^2}{4}} e^{-N \frac{x^2}{4}} \leq e^{-N \frac{t^2}{4}} e^{-N \frac{x^2}{4}},$$

whenever $|x| \geq t$ to justify the first inequality. Now

$$\mathbb{P}_{\sigma_N} \left(\max_{i=1, \dots, N} |\lambda_i| \geq t \right) \leq N \mathbb{P}_{\sigma_N}(|\lambda_1| \geq t) \leq e^{CN - \frac{1}{4} N t^2},$$

for some new constant $C > 0$. Then for any $\alpha < \frac{1}{4}$ and $t \geq t_0(\alpha)$, (A.3) holds for all N . □

The proof of Lemma 3.0.13 now follows that of [GM07, Lemma 2.2].

Proof of Lemma 3.0.13. Since

$$\lambda_{\max}^N(\mathbf{A}) = \max_{1 \leq i \leq d} \sup_{\|v\|=1} \sqrt{\langle v, A_i A_i^\dagger v \rangle}$$

is a convex function of the entries of the A_i 's, then $e^{sN\lambda_{\max}^N(\mathbf{A})^2}$ is also a convex function. Note also that

$$\lambda_{\max}^N(\mathbf{A})^2 \leq (\lambda_{\max}^N(\mathbf{A} - \mathbf{M}) + \lambda_{\max}^N(\mathbf{M}))^2 \leq 2\lambda_{\max}^N(\mathbf{A} - \mathbf{M})^2 + 2\lambda_{\max}^N(\mathbf{M})^2,$$

where $\mathbf{M} = \mathbb{E}_{\mu_V^N}[\mathbf{A}]$ is mean of the tuple $\mathbf{A}$. By the proof of [GM07, Lemma 2.1] there is some $C_1 > 0$ for which $\lambda_{\max}^N(\mathbf{M}) < C_1$ for all N . Furthermore, by the generalised version of the Brascamp-Lieb inequality

$$\int e^{sN\lambda_{\max}^N(\mathbf{A}-\mathbf{M})^2} d\mu_V^N \leq \int e^{sN\lambda_{\max}^N(\mathbf{A})^2} d\mu_c^N \leq C_2^N,$$

for all suitably small s and some $C_2 > 0$, and where μ_c^N is the law of the GUE ensemble with variance (cN^{-1}) . The final inequality follows from the sub-Gaussian nature of the tails for the GUE and Lemma A.1.1. We conclude by observing that for some small $s > 0$,

$$\int e^{sN\lambda_{\max}^N(\mathbf{A})^2} d\mu_V^N \leq e^{2\lambda_{\max}^N(\mathbf{M})^2} \int e^{2sN\lambda_{\max}^N(\mathbf{A}-\mathbf{M})^2} d\mu_V^N \leq e^{C_1} C_2^N,$$

and applying Lemma A.1.1. □

APPENDIX B

APPENDIX TO LÉVYGAN

We collect here various results related to [Chapter 5](#).

B.1 Some properties of the characteristic function distance

This short subsection summarises two key properties of the distance CFD_ν . Namely, the following proposition shows that the distance between the empirical characteristic function and true characteristic function converges to zero almost surely as the number of observations tends to infinity. The second result demonstrates that by carefully choosing ν , convergence in CFD_ν implies weak convergence.

Proposition B.1.1. *Let $\{X_i\}_{i=1}^\infty$ be a collection of i.i.d. $\mathbb{R}^n$ valued random variables, then*

$$\lim_{n \rightarrow \infty} \mathbb{E}_{\Lambda \sim \nu} \left[\left| \hat{\Phi}_X^n(\Lambda) - \Phi_X(\Lambda) \right| \right] \rightarrow 0 \text{ almost-surely.}$$

Proof. By [\[FM77; Csö81a; Csö81b\]](#), there exists a sequence of real numbers $\{T_n\} \uparrow \infty$ such that almost-surely for every $\varepsilon > 0$ there exists an N such that for every $n \geq N$

$$\sup_{|\Lambda| \leq T_n} \left| \hat{\Phi}_X^n(\Lambda) - \Phi_X(\Lambda) \right| < \varepsilon.$$

It follows almost-surely for every $\varepsilon > 0$ that

$$\begin{aligned}
\mathbb{E}_{\Lambda \sim \nu} \left[\left| \hat{\Phi}_X^n(\Lambda) - \Phi_X(\Lambda) \right| \right] &= \int \left| \hat{\Phi}_X^n(\Lambda) - \Phi_X(\Lambda) \right| d\nu(\Lambda) \\
&\leq \int_{|\Lambda| \leq T_n} \left| \hat{\Phi}_X^n(\Lambda) - \Phi_X(\Lambda) \right| d\nu(\Lambda) + 2\nu(\{|\Lambda| \geq T_n\}) \\
&\leq \varepsilon \nu(\{|\Lambda| \leq T_n\}) + 2\nu(\{|\Lambda| \geq T_n\}) \\
&\rightarrow \varepsilon.
\end{aligned}$$

□

Proposition B.1.2. *Let ν be the Cauchy distribution on $\mathbb{R}^n$ with location parameter 0 scale parameter γ and independent coordinates, then CFD_ν metrizes the topology of weak convergence on $\mathcal{P}(\mathbb{R}^n)$.*

Proof. We consider the case $n = 1$; the general case is a straightforward extension. If the law of a random variable T is that of a Cauchy distribution with location parameter 0 and scale parameter γ , then it is well known that its characteristic function is given by

$$\Phi_T(t) = e^{-\gamma|t|},$$

and has a density given by

$$f_T(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} \Phi_T(t) dt = \frac{1}{\pi\gamma \left(1 + \left(\frac{x}{\gamma}\right)^2\right)}.$$

It is clear that both Φ_T and f_T are in $L^1(\mathbb{R})$, since

$$\int_{\mathbb{R}} \Phi_T(t) dt = \frac{2}{\gamma}, \text{ and } \int_{\mathbb{R}} f_T(x) dx = 1.$$

We also observe that

$$\int_{\mathbb{R}} \frac{1}{f_T(x)(1+|x|^4)} dx \leq 2\pi\gamma \int_0^\infty \frac{\left(1 + \frac{|x|}{\gamma}\right)^2}{(1+|x|)^4} dx < \infty.$$

An application of [Sri+10], the uniform boundedness of characteristic functions in $L^\infty(\nu)$, and a standard interpolation argument yields that CFD_ν metrizes weak convergence. □

B.2 Characteristic function and joint moments of a Brownian increment and Lévy area

Here we briefly give the definition of the joint characteristic function and Lévy area, and also provide a proof of [Proposition 5.2.7](#). Part of the proof of the second result comes as a corollary of the form of the characteristic function found in [Theorem B.2.1](#) and [Lemma B.2.2](#).

Theorem B.2.1 (Characteristic function of Brownian motion and Levy area). *Let $W_t = (W_t^{(1)}, \dots, W_t^{(d)})$ be a d -dimensional Brownian motion and let $A_t^{(j_1, j_2)}$ be the corresponding Lévy area of $W_t^{(j_1)}$ and $W_t^{(j_2)}$. Let $\mu \in \mathbb{R}^d$ and $\Lambda = \{\Lambda_{i,j}\}_{1 \leq i < j \leq d} \in \mathbb{R}^{\frac{d(d-1)}{2}}$. The joint characteristic function of coupled Brownian motion W_t and the Levy area A_t at time t*

$$\Phi_{(W,A)}(t, \mu, \Lambda) := \mathbb{E} \left[\exp \left(i \sum_{i=1}^d \mu_i W_{0,t}^{(i)} + i \sum_{1 \leq j_1 < j_2 \leq d} \Lambda_{j_1, j_2} A_t^{(j_1, j_2)} \right) \right]$$

admits the following formula

$$\begin{aligned} \Phi_{(W,A)}(t, \mu, \Lambda) = & \left(\prod_{i=1}^{d_1} \frac{1}{\cosh(\frac{\eta_i}{2}t)} \right) \exp \left(\left[\sum_{i=1}^{d_1} -\frac{1}{\eta_i} ((R\mu)_{2i-1}^2 \right. \right. \\ & \left. \left. + (R\mu)_{2i}^2) \tanh\left(\frac{\eta_i}{2}t\right) \right] - \frac{1}{2}t \sum_{i=1}^{d_0} (R\mu)_{2d_1+i}^2 \right) \end{aligned}$$

where R, η, d_0, d_1 are defined in [Lemma B.2.2](#).

Proof. The result was first proved in [\[HS83\]](#) using a probabilistic approach. Recently, [\[LNT24\]](#) have given a proof using rough path theory. $\square$

Lemma B.2.2 (Decomposition of anti-symmetric matrix). *For any $d \times d$ anti-symmetric real-valued matrix Λ , let $(\pm\eta_1 i, \dots, \pm\eta_{d_1} i)$ be the set of non-zero conjugate eigenvalue pairs of Λ with $\eta_i > 0$ and $\eta_1 \geq \dots \geq \eta_{d_1}$. Let $d_0 = d - 2d_1$ be the algebraic multiplicity of the eigenvalue 0 of Λ (if Λ does not have zero eigenvalues, then $d_0 = 0$). Then there exists an orthogonal matrix R , such that the following decomposition of Λ holds:*

$$\Lambda = R^T \Sigma R,$$

where Σ is of the form

$$\Sigma = \begin{pmatrix} \Sigma_0, & \mathbf{0}_{d-d_0, d_0} \\ \mathbf{0}_{d_0, d-d_0} & \mathbf{0}_{d_0, d_0} \end{pmatrix}$$

$$:= \begin{pmatrix} 0 & -\eta_1 & 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ \eta_1 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & -\eta_2 & \dots & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & \eta_2 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & 0 & \dots & 0 & -\eta_{d_1} & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & \dots & \eta_{d_1} & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 & 0 & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & \dots & 0 \end{pmatrix},$$

where Σ_0 is the block diagonal matrix with all non-zero η_i .

Using the above, we are now able to give a proof of [Proposition 5.2.7](#), concerning the joint moments of our generator.

Proof of Proposition 5.2.7. To see item (1), fix $\mu \in \mathbb{R}^d$ and $\lambda = \{\lambda_{i,j}\}_{1 \leq i < j \leq d} \in \mathbb{R}^{\frac{d(d-1)}{2}}$. We claim that

$$\Phi_{(W, \mathbf{A})}(t, \mu, \lambda) = \Phi_{(W, \mathbf{A})}(t, -\mu, -\lambda) = \Phi_{(W, \mathbf{A})}(t, \mu, -\lambda). \quad (\text{B.1})$$

Let Λ stand for the $d \times d$ anti-symmetric matrix with the above diagonal elements given by λ . Let Λ have a decomposition $\Lambda = R^T \Sigma R$ as in [Lemma B.2.2](#). Then the corresponding decomposition for $-\Lambda$ is given by $\tilde{R}^T \Sigma \tilde{R}$, where for $1 \leq i \leq d_1$

$$\tilde{R}_{2i,\cdot} := R_{2i-1,\cdot} \quad \text{and} \quad \tilde{R}_{2i-1,\cdot} := R_{2i,\cdot}.$$

It is clear from the form of the characteristic function, that pairwise swapping the first $2d_1$ rows of R leaves the final value unchanged. Since all the terms involving μ are squared, the characteristic function is also invariant under taking the negative of μ . Item (1) now follows from [Eq. \(B.1\)](#). For item (2), [Eq. \(5.5\)](#) follows immediately from item (1), and [Eq. \(5.7\)](#) follows from the fact that the conditional characteristic function of Lévy area given $W_{0,t} = \mathbf{w}$ is purely real [[Wik01](#)]. [Eq. \(5.8\)](#) follows by independence of ξ_0 and the observation that its power will be odd. We finally turn our attention to [Eq. \(5.6\)](#), where for simplicity we take $t = 1$. We recall that

$\tilde{A}_{0,1} \sim \mathbb{P}_{\text{BF}}^\theta$ is defined by $\text{BF}(W_{0,1}, H_{0,1}, \tilde{b}_{0,1}, \xi, \mathbf{\xi})$, where the only dependence is between $H_{0,1}$ and $\tilde{b}_{0,1}$. Indeed, by expanding each $\tilde{A}^{(i)}$ we see that it is enough to show that

$$\mathbb{E} \left[\prod_{i=1}^d (W^{(i)})^{n_i} \left(\prod_{1 \leq i < j \leq d} (\xi_0)^{n_{ij}} (\xi_i H^{(i)} W^{(j)})^{k_{ij}} (\xi_j H^{(j)} W^{(i)})^{l_{ij}} (\xi_i \xi_j \tilde{b}^{(i,j)})^{p_{ij}} \right) \right] = 0, \quad (\text{B.2})$$

where $k_{ij} + l_{ij} + p_{ij} = n_{ij}$. We consider three exhaustive, but overlapping cases.

- i) $\sum_{1 \leq i < j \leq d} n_{ij}$ is odd.
- ii) For some $1 \leq i \leq d$, the power of $W^{(i)}$ or ξ_i is odd.
- iii) For every $1 \leq i \leq d$, the power of $W^{(i)}$ and ξ_i is even.

If i) holds, then the power of ξ_0 is odd and by independence Eq. (B.2) is true. Item ii) is similar. To conclude, we will show that if iii) holds, then i) must also hold. We observe that the power of $W^{(i)}$ is given by

$$n_i + \sum_{1 \leq j < i} k_{ji} + \sum_{i < j \leq d} l_{ij},$$

and the power of ξ_i is given by

$$\sum_{1 \leq j < i} l_{ji} + p_{ji} + \sum_{i < j \leq d} k_{ij} + p_{ij}.$$

Under the assumption that all these powers are even, we may add them together to see that

$$\sum_{i=1}^d n_i + \left(\sum_{1 \leq j < i} k_{ji} + l_{ji} + p_{ji} + \sum_{i < j \leq d} k_{ij} + l_{ij} + p_{ij} \right) = \sum_{i=1}^d n_i + 2 \left(\sum_{1 \leq i < j \leq d} n_{ij} \right)$$

is even. As the sum over all n_i and n_{ij} is odd by assumption, it must be the case that

$$\sum_{1 \leq i < j \leq d} n_{ij}$$

is odd. □

Lemma B.2.3. *Let W be a d -dimensional Brownian motion on $[0, 1]$ and let B, H, b be the corresponding derived processes from Definition 5.2.1. Fix some $\xi \in \{-1, 1\}^d$, and let H' , and b' be the space-time and space-space Lévy area processes associated with the process*

$\xi \odot B = \{\xi \odot B_t\}_{t \in [0,t]}$, where $\odot$ denotes the Hadamard (coordinate-wise) product. Then

$$\{B_t\}_{t \in [0,1]} \stackrel{d}{=} \{\xi \odot B_t\}_{t \in [0,1]}, \quad H' = \xi \odot H, \quad b' = (\xi \otimes \xi) \odot b, \quad \text{and} \quad (H', b') \stackrel{d}{=} (H, b).$$

Proof. By definition of H and b (in vector form)

$$\begin{aligned} H'_{0,t} &= \frac{1}{t} \int_0^t \xi \odot B_{0,u} \, du = \xi \odot H_{0,t} \\ b'_{0,t} &= \int_0^t (\xi \odot B_{0,u}) \otimes d(\xi \odot B_u) \\ &= (\xi \otimes \xi) \odot \int_0^t B_{0,u} \otimes dB_u = (\xi \otimes \xi) \odot b_{0,t}. \end{aligned}$$

□

B.3 Test metrics

We list in this section the test metrics used to assess the performance of our generative model. Throughout the section let X, Y be d -dimensional random variables on the metric space (M, m) and let $\mu := \mathbb{P}_X, \nu := \mathbb{P}_Y$ be the induced probability measure. We denote by μ_i and ν_i the measure induced by the marginal distributions of $X^{(i)}$ and $Y^{(i)}$ respectively.

- 1) Marginal 2-Wasserstein metric. For $1 \leq i \leq d$, The 2-Wasserstein distance between $X^{(i)}$ and $Y^{(i)}$ is given by

$$\mathcal{W}_2(\mu, \nu) = \left(\inf_{\gamma \in \Gamma(\mu, \nu)} \int m(x, y)^2 d\gamma(x, y) \right)^{\frac{1}{2}}$$

where $\Gamma(\mu, \nu)$ is the set of all joint measures on $M \times M$ with marginals μ and ν . The Wasserstein metric is one way of assessing the difference between two distributions, however, in practice the estimation is often intractable when d is high as one needs to compute $m(x, y)$ for all samples x and y . To accommodate this issue, we compute the $\mathcal{W}_2$ distance between the marginal distribution $X^{(i)}$ and $Y^{(i)}$ only.

- 2) Cross moment metric. We compare the difference of the fourth moments between the real and generated Lévy area. Let $\{\mathbf{X}_i\}_{i=1}^N$ be samples of the random variable X , for any $(i_1, i_2, i_3, i_4) \in \{1, \dots, d\}^4$ we estimate the cross

moment $\mathbb{E}[X^{(i_1)} X^{(i_2)} X^{(i_3)} X^{(i_4)}]$ by

$$\frac{1}{N^4} \sum_{1 \leq j_1, j_2, j_3, j_4 \leq N} \mathbf{X}_{j_1}^{(i_1)} \mathbf{X}_{j_2}^{(i_2)} \mathbf{X}_{j_3}^{(i_3)} \mathbf{X}_{j_4}^{(i_4)}$$

If X happens to be the Lévy area of Brownian motions we do know the analytical form of each cross-moment. One can derive it using the basic tools of rough path theory, in this case the expected signature of Brownian motion. This metric is used to assess the joint fitting of generated Lévy area without using estimations of real data.

- 3) Characteristic Function Distance and Maximum Mean Discrepancy distance (MMD). In [Sri+10] it is shown that the characteristic function distance CFD_ν when using an L^2 norm instead of an L^1 norm is equivalent to an MMD distance with a certain kernel. For example, if ν is a Gaussian measure, then the corresponding kernel is the Gaussian kernel. For our tests we use the MMD distance with both a Gaussian and polynomial kernel, with the Fourier series expansion with a high truncation level taken to be the ground truth.
- 4) Empirical Unitary Characteristic Function Distance. We parametrize the measure $\mathcal{M}$ on the space of linear transformations $\mathbb{R}^{d+a}$ onto $\mathfrak{g}_m$ by M linear maps, where d, a and m denotes the Brownian dimension, Lévy dimension, and unitary Lie degree respectively. Then, we train $\mathcal{M}$ to maximize the EUCFD distance between real and generated data. In practice, we set $m = 8$, $M = 128$, and train $\mathcal{M}$ for 2000 iterations, the EUCFD computed on an independent test set is used for model assessment.

B.4 Training procedure and hyperparameter tuning

The training procedure was performed the training procedure as illustrated in [Algorithm 4](#). On the generator side, we used a feed-forward neural network, with the activation function is chosen to be LeakyRelu. On the discriminator side, we parameterise 128 linear maps onto the Lie algebra of degree 3 to mimic the empirical distribution used to compute UCFD mentioned in [Section 5.3.1](#). The total number of training iterations is set to be 2500, which was approximately the point that we observed convergence in the marginal 2-Wasserstein metric to reference samples. We optimised both the generator and discriminator using stochastic gradient descent with the Adam optimiser. We set the batch size to 2^{13} and the learning rate for generator/discriminator was chosen to be 0.001/0.01 respectively.

We also employed a learning rate scheduler, with both learning rates decay every 500 iterations. Finally, we set iter_d to be 3.

Regarding the choice of hyperparameters, we conducted a hyperparameter grid search for training the LévyGAN. Tuning is done on both the generator and discriminator sides. The model selection was based on the marginal 2-Wasserstein metric. We provide here a grid for interested readers.

Table B.1: Grid for hyperparameter tuning

Hyperparameter	Grid values
Generator hidden layers	2, 3, 5
Generator hidden dimension	8, 16
Generator noise size	4, 8
Generator slope of LeakyRelu	0.01, 0.2
Discriminator Lie degree	2, 3, 5
Discriminator batch size	16, 64, 128

B.5 Foster's moment-matching method

The current state-of-the-art weak estimator for Lévy area is Foster's method. We now detail its construction for the interested reader.

Definition B.5.1 (Foster's approximation of Lévy area). *For any $0 \leq s < t$ and $d \geq 2$ we define the antisymmetric tensor $\tilde{A}_{s,t}$ with entries*

$$\tilde{A}_{s,t}^{(i,j)} = H_{s,t}^{(i)} W_{s,t}^{(j)} - W_{s,t}^{(i)} H_{s,t}^{(j)} + 12 \left(K_{s,t}^{(i)} H_{s,t}^{(j)} - H_{s,t}^{(i)} K_{s,t}^{(j)} \right) + \tilde{a}_{s,t}^{(i,j)}, \quad \text{for } 1 \leq i, j \leq d$$

where

- W is a d -dimensional Brownian motion and H is a space-time bridge Lévy area on $[s, t]$.
- $K_{s,t} \in \mathbb{R}^d$ is the space-time-time Lévy area of the Brownian bridge between s and t , is distributed as $K_{s,t} \sim \mathcal{N}^d(0, \frac{1}{720}(t-s))$, with K and (W, H) being independent.
- $\tilde{a}$ is an approximation of the Brownian arch Lévy area (see [Fos20], definition 4.1.14) constructed as shown below.

$$\tilde{a}_{s,t}^{(i,j)} = \begin{cases} \sigma_{s,t}^{(i,j)} \xi_{s,t}^{(i,j)} & \text{if } i < j \\ -\sigma_{s,t}^{(i,j)} \xi_{s,t}^{(i,j)} & \text{if } j < i \\ 0 & \text{if } i = j \end{cases}$$

with the independent random variables $\sigma_{s,t}^{(i,j)}$ and $\xi_{s,t}^{(i,j)}$ defined for $1 \leq i < j \leq d$ according to

$$\xi_{s,t}^{(i,j)} \sim \begin{cases} \text{Uni}[-\sqrt{3}, \sqrt{3}] & \text{with probability } p \\ \text{Rad}(1/2) & \text{with probability } 1 - p \end{cases} \quad \text{where } p := \frac{21130}{25621}$$

and

$$\sigma_{s,t}^{(i,j)} = \sqrt{\frac{3}{28} (C^{(i)} + c) (C^{(j)} + c) (t - s)^2 + \frac{1}{28} (t - s) \left((12K_{s,t}^{(i)})^2 + (12K_{s,t}^{(j)})^2 \right)},$$

where the $C^{(i)} \sim \text{Exp}(\frac{15}{8})$, $(1 \leq i \leq d)$ are i.i.d. random variables and $c := \frac{1}{\sqrt{3}} - \frac{8}{15}$.

Theorem B.5.2. Let $d \in \{2, 3\}$ and let $\tilde{A}$ be as in [Definition B.5.1](#). Then $\tilde{A}$ matches all the fifth and lower moments of the Lévy area A conditional on W and H . That is, for any $n_1, n_2, n_3 \geq 0$ with $n_1 + n_2 + n_3 \leq 5$

$$\begin{aligned} \mathbb{E} \left[\left(\tilde{A}_{s,t}^{(1,2)} \right)^{n_1} \left(\tilde{A}_{s,t}^{(2,3)} \right)^{n_2} \left(\tilde{A}_{s,t}^{(3,1)} \right)^{n_3} \mid W_{s,t}, H_{s,t} \right] = \\ \mathbb{E} \left[\left(A_{s,t}^{(1,2)} \right)^{n_1} \left(A_{s,t}^{(2,3)} \right)^{n_2} \left(A_{s,t}^{(3,1)} \right)^{n_3} \mid W_{s,t}, H_{s,t} \right]. \end{aligned}$$

Proof. See [\[Fos20\]](#), page 190. □

B.6 Numerical methods for SDEs

This section details the various numerical schemes with which we perform our tests. We start by recalling the definition of weak and strong errors and discuss why we are only able to analyse the former.

B.6.1 Weak vs strong error

Strong error measures the discrepancy between the true sample paths of the process and the approximate sample paths, while weak error measures the similarity between the distributions of the true and approximate solutions. Since our simulation methods are not exact, nor do they correspond to some approximation of a true sample of Lévy area (in comparison to a truncated Fourier series expansion for example), we cannot measure the strong error for schemes involving fake Lévy area. We can, however, measure the weak error of the discretisation schemes. Concretely, a discretisation scheme $\{\hat{X}_k\}_{0 \leq k \leq N}$ is said to converge weakly with

order α if for any polynomial p

$$\left| \mathbb{E}[p(\hat{X}_N) - p(X_T)] \right| \leq C_p h^\alpha, \quad (\text{B.3})$$

for some constant $C_p > 0$ and for all sufficiently small step sizes $h := \frac{T}{N}$. Convergence with weak order α will be denoted by $O(h^\alpha)$. The following subsections introduce three discretisation schemes that will be used for our numerical results.

B.6.2 Discretisation schemes

Here we recall the definitions of the numerical schemes used for the results presented in [Section 5.5.2](#).

Milstein's Method

Milstein's method is derived from the second order (Itô) Taylor expansion of the SDE in [Eq. \(5.1\)](#). It is defined as follows: fix $N \geq 1$, set $\hat{X}_0 = x_0$, and for all $0 \leq k \leq N - 1$ construct $\hat{X}_{k+1}$ recursively via the relation

$$\begin{aligned} \hat{X}_{k+1} = & \hat{X}_k + f(\hat{X}_k)h + \sum_{i=1}^d g_i(\hat{X}_k) \Delta W_k^{(i)} \\ & + \sum_{i,j=1}^d g'_i(\hat{X}_k) g_j(\hat{X}_k) \left(\frac{1}{2} (\Delta W_k^{(i)}) (\Delta W_k^{(j)}) + A_k^{(i,j)} - \frac{1}{2} \delta_{ij} h \right), \end{aligned} \quad (\text{B.4})$$

where $h := \frac{T}{N}$, $\Delta W_k^{(i)} := W_{t_{k+1}}^{(i)} - W_{t_k}^{(i)}$, δ_{ij} is the Kronecker delta, and

$$A_k^{(i,j)} := \frac{1}{2} \left(\int_{t_k}^{t_{k+1}} (W_s^{(i)} - W_{t_k}^{(i)}) dW_s^{(j)} - \int_{t_k}^{t_{k+1}} (W_s^{(j)} - W_{t_k}^{(j)}) dW_s^{(i)} \right) \quad (\text{B.5})$$

is the Lévy area. Milstein's method, under certain conditions, is known to converge with both weak and strong order $O(h)$. Meanwhile, removing the Lévy area terms leaves the weak order unchanged, but reduces the strong order to $O(\sqrt{h})$. Since we are concerned exclusively with weak convergence, we will consider Milstein's method without Lévy area.

Strang log-ODE

The Strang log-ODE method, introduced in [\[FdRS24\]](#), is a higher order method that incorporates a term involving fake Lévy area. To define this scheme we must

first convert the SDE in Eq. (5.1) into Stratonovich form, that is

$$dX_t = \bar{f}(X_t)dt + \sum_{i=1}^d g_i(X_t) \circ dW_t^{(i)}, \quad X_0 = x_0, \quad (\text{B.6})$$

where the Stratonovich drift $\bar{f}$ is given by

$$\bar{f}(x) := f(x) - \sum_{i=1}^d g'_i(x)g_i(x).$$

As in the Milstein scheme we fix $\hat{X}_0 = x_0$, and recursively define

$$\hat{X}_{k+1} = \exp\left(\frac{1}{2}h\bar{f}\right) \exp\left(\sum_{i=1}^d g_i \Delta W_k^{(i)} + \sum_{i < j} (g'_j g_i - g_j g'_i) A_k^{(i,j)}\right) \exp\left(\frac{1}{2}h\bar{f}\right) (\hat{X}_k), \quad (\text{B.7})$$

where $\exp(Cf)(x)$ denotes the time 1 solution to the following ordinary differential equation

$$\frac{dy}{dt} = Cf(y), \quad y_0 = x.$$

For the Lévy area term we will consider three different weak approximations: a Rademacher random variable which matches the variance of Lévy area, Foster's method, and our generative modelling method. The first method is based on the use of Rademacher random variables in Talay's scheme [Tal90].

B.6.3 Antithetic MLMC

Normally, in order to achieve optimal computational complexity of the MLMC it is required that the underlying discretisation scheme has strong order $O(h)$, see [Gil08], which requires strong simulation of Lévy area when the SDE is more than one dimensional. However, there exists clever reformulation of the standard MLMC which achieves optimal computational complexity using the no area Milstein scheme, despite only having a strong convergence rate of $O(\sqrt{h})$. Antithetic MLMC, proposed in [GS14a], was originally motivated by the difficulty of simulating Lévy area exactly. Now, as mentioned in the previous subsection, on each level the coarse and fine path are generated using the same underlying Brownian motion. It is not in fact necessary to use the same estimator on both the coarse and fine paths. It suffices that $Y_l^{:,l} \stackrel{d}{=} Y_l^{:,l+1}$. For the antithetic method, the paper [GS14a] uses two different paths on the fine level: the fine path itself and an antithetic twin. The antithetic twin is obtained by pairwise swapping the increments of the fine path; essentially this means that $\hat{X}^{(f)} + \hat{X}^{(a)} \approx 2\hat{X}^{(c)}$, where

$\hat{X}^{(f)}$ denotes the fine path, $\hat{X}^{(a)}$ the antithetic path, and $\hat{X}^{(c)}$ the coarse path. The resulting effect is that the variance of $\frac{1}{2}(\varphi(\hat{X}^{(f)}) + \varphi(\hat{X}^{(a)})) - \varphi(\hat{X}^{(c)})$ should be small. For more details, we refer the reader to [GS14a].

B.6.4 The log-Heston SDE

We recall the definition of the log-Heston model as the following two-dimensional SDE

$$\begin{aligned} dU_t &= \left(r - \frac{1}{2}V_t\right) dt + \sqrt{V_t}dW_t^{(1)}, \quad U_0 \in \mathbb{R} \\ dV_t &= \kappa(\theta - V_t)dt + \sigma\sqrt{V_t}dW_t^{(2)}, \quad V_0 > 0. \end{aligned} \quad (\text{B.8})$$

The discounted payoff of a European call option for a price S with $d \log(S) := U$ is

$$\varphi(S) := e^{-rT} \left(e^{U_T} - K\right)^+.$$

The price at $t = 0$ of this option may be written as

$$C_0 := \mathbb{E} [\varphi(S)] = S_0 \Pi_0 - e^{-rT} K \Pi_1, \quad (\text{B.9})$$

where the factors Π_0 and Π_1 are given by

$$\begin{aligned} \Pi_0 &= \frac{1}{2} + \frac{1}{\pi} \int_0^\infty \Re \left[\frac{e^{i\omega K} \Psi_{U_T}(\omega - i)}{i\omega \Psi_{U_T}(-i)} \right] d\omega \\ \Pi_1 &= \frac{1}{2} + \frac{1}{\pi} \int_0^\infty \Re \left[\frac{e^{i\omega K} \Psi_{U_T}(\omega)}{i\omega} \right] d\omega. \end{aligned} \quad (\text{B.10})$$

Here $\Psi_{\log Y_T}$ denotes the characteristic function of U_T . The characteristic function itself has an analytic form given by

$$\Psi_{U_T}(\omega) = \exp \left\{ C(\omega)\theta + D(\omega)V_0 + i\omega \log \left(S_0 e^{rT} \right) \right\},$$

where the functions $C(\cdot)$ and $D(\cdot)$ are defined as

$$\begin{aligned} C(\omega) &:= \kappa \left[b_1 T - \frac{2}{\sigma^2} \log \left(\frac{1 - b_2 e^{-aT}}{1 - b_2} \right) \right] \\ D(\omega) &:= b_1 \frac{1 - e^{-aT}}{1 - b_2 e^{-aT}}, \end{aligned}$$

with constants a, b_1 and b_2 given by

$$\begin{aligned} a &:= \sqrt{\kappa^2 + \sigma^2 \omega(\omega - i)} \\ b_1 &:= \frac{\kappa - a}{\sigma^2} \\ b_2 &:= \frac{\kappa - a}{\kappa_a}. \end{aligned}$$

For full details of the derivation of this formula, we refer the reader to [Cri15]. While the preceding expression may appear complicated, it requires only the evaluation of the deterministic integrals in Eq. (B.10), which are computed in practice using quadrature.

What remains to full specify the numerical schemes is to compute the one step recursion for the various numerical schemes applied log-Heston SDE from Eq. (B.8). Full details and derivations of the vector field derivatives can be found in [Ter22]. The no-area Milstein update is given by

$$\begin{aligned} \hat{U}_{k+1} &= \hat{U}_k + (r - \tfrac{1}{2}\hat{V}_k)h + \sqrt{\hat{V}_k}\Delta W_k^{(1)} + \tfrac{1}{4}\sigma\Delta W_k^{(1)}\Delta W_k^{(2)} \\ \hat{V}_{k+1} &= \hat{V}_k + \kappa(\theta - \hat{V}_k)h + \sigma\sqrt{\hat{V}_k}\Delta W_k^{(2)} + \tfrac{1}{4}\sigma^2\left((\Delta W_k^{(2)})^2 - h\right). \end{aligned} \quad (\text{B.11})$$

The Strang log-ODE recursion is more involved as it involves solving two ordinary differential equations: one for the Stratonovich drift and one involving diffusion terms. Fortunately, these are analytically solvable, with the resulting scheme given by

$$\begin{aligned} \tilde{V}_{k+1}^1 &= (\hat{V}_k + \xi)e^{-\frac{\kappa h}{2}} + \xi \\ \tilde{U}_{k+1}^1 &= \hat{U}_k + \frac{1}{2\kappa}(\hat{V}_k - \xi)(e^{-\frac{\kappa h}{2}} - 1) + \frac{h}{2}(r - \frac{\xi}{2}) \\ \tilde{V}_{k+1}^2 &= \left(\sqrt{\tilde{V}_{k+1}^1} + \frac{\sigma}{2}\Delta W_k^{(2)}\right)^2 \end{aligned} \quad (\text{B.12})$$

$$\begin{aligned} \tilde{U}_{k+1}^2 &= \tilde{U}_{k+1}^1 + \sqrt{\tilde{V}_{k+1}^1}\Delta W_k^{(1)} + \frac{\sigma}{4}\Delta W_k^{(1)}\Delta W_k^{(2)} - \frac{\sigma}{2}\tilde{A}_k^{(1,2)} \\ \hat{V}_{k+1} &= (\tilde{V}_{k+1}^2 - \xi)e^{-\frac{\kappa h}{2}} + \xi \\ \hat{U}_{k+1} &= \tilde{U}_{k+1} + \frac{1}{2\kappa}(\tilde{V}_{k+1}^2 - \xi)(e^{-\frac{\kappa h}{2}} - 1) + \frac{h}{2}(r - \frac{\xi}{2}), \end{aligned}$$

where $\xi := \theta - \frac{\sigma^2}{4\kappa}$ and $\tilde{A}$ denotes the fake Lévy area.