

# QUANTUM PATH SIGNATURES

Samuel Crew<sup>1,2,\*</sup>, Cristopher Salvi<sup>1,\*</sup>, William F. Turner<sup>1,\*</sup>, Thomas Cass<sup>1</sup>, Antoine Jacquier<sup>1</sup>

<sup>1</sup>Department of Mathematics, Imperial College London, UK.

<sup>2</sup>Department of Physics, National Tsing Hua University, Taiwan.

August 8, 2025

## Abstract

We elucidate physical aspects of path signatures by formulating randomised path developments within the framework of matrix models in quantum field theory. Using tools from physics, we introduce a new family of randomised path developments and derive corresponding loop equations. We then interpret unitary randomised path developments as time evolution operators on a Hilbert space of qubits. This leads to a definition of a quantum path signature feature map and associated quantum signature kernel through a quantum circuit construction. In the case of the Gaussian matrix model, we study a random ensemble of Pauli strings and formulate a quantum algorithm to compute such kernel.

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Background</b>	<b>5</b>
2.1	Quantum field theory . . . . .	5
2.2	Wilson lines . . . . .	6
2.3	Path signatures . . . . .	7
2.4	Signature kernels . . . . .	9
<b>3</b>	<b>Universal Limit of Matrix Model Unitary Developments</b>	<b>11</b>
3.1	Non-commutative laws and Schwinger-Dyson equations . . . . .	12
3.2	Path loop equations . . . . .	15
<b>4</b>	<b>Quantum algorithm</b>	<b>18</b>
4.1	Preliminaries . . . . .	19
4.2	Quantum circuit . . . . .	19
4.3	Quantum signature kernel . . . . .	23
<b>5</b>	<b>Conclusion</b>	<b>26</b>
	<b>References</b>	<b>27</b>
<b>A</b>	<b>Technical proofs</b>	<b>30</b>
<b>B</b>	<b>Proof of Theorem 4.8</b>	<b>32</b>
<b>C</b>	<b>Proofs of Theorem 4.11 and Theorem 4.12</b>	<b>41</b>
C.1	Quantum algorithm: Proof of Theorem 4.11 . . . . .	41
C.2	Classical Monte Carlo algorithm: Proof of Theorem 4.12 . . . . .	42

<sup>1</sup>These authors contributed equally to this work.

# I Introduction

On sufficiently fine time scales, many types of sequential data, such as audio, video, time series, or text, can naturally be modelled as paths  $\gamma : [0, 1] \rightarrow \mathbb{R}^d$ . The idea that such paths can be characterised (up to reparameterisation) by their *path signature* was first introduced by Chen [Che57] for smooth paths, and later extended to more irregular paths in the context of *rough path theory* [Ly098; HL10; Boe+16]. In this work, we elucidate a connection between *quantum field theory* (QFT) and the theory of path signatures. Adopting the language of QFT, refined later in Section 2, we treat a path  $\gamma$  as a Wilson line and interpret its signature as arising from parallel transport in gauge theory. This perspective allows us to draw on tools from high-energy theoretical physics to study a class of differential equations driven by random vector fields sampled from so-called matrix model distributions, which develop the path into a unitary group. An important instance of this framework is the Gaussian unitary development, whose physical interpretation naturally leads to a representation of the control system as a quantum circuit acting on a Hilbert space of qubits. This construction, in turn, gives rise to a quantum algorithm whose output approximates the associated quantum signature kernels.

The path signature  $\mathcal{S}(\gamma)$  is formally defined as the solution to the tensor differential equation  $dy = y \otimes d\gamma$  in the free tensor algebra  $T((\mathbb{R}^d)) := \prod_{n=0}^{\infty} (\mathbb{R}^d)^{\otimes n}$ , and thereby can be seen as a non-commutative analogue of the exponential function. By applying Picard iteration, one recovers the familiar formulation of the signature as the sequence of iterated integrals  $\left( \int_{0 < t_1 < \dots < t_n < 1} d\gamma_{t_1} \otimes \dots \otimes d\gamma_{t_n} \right)_{n \in \mathbb{N}}$ . It is a classical result that linear functionals on  $T((\mathbb{R}^d))$ , when restricted to the image of the signature map, form a unital algebra that separates points. Consequently, the classical Stone–Weierstrass theorem ensures that linear functionals on signatures are dense in the space of continuous real-valued functions defined on compact subsets of unparameterised paths [CT24b]. This makes the signature a powerful representation, enabling the approximation of arbitrary path-dependent functionals using linear models [Sal+23]. Indeed, signature methods have seen a rapid rise in popularity in recent years, finding applications across a broad range of domains in data science, including deep learning for sequence modelling [Kid+19; Mor+21; SLG22; Hog+23; Muç+24; Iss+24; BHS24; MS25a], quantitative finance [ASS20; Sal+21b; Hor+23; PS24; MS25b], cybersecurity [Coc+21], and computational neuroscience [HS24], among many others. The interested reader is referred to [CS24] for a detailed and pedagogical account on the subject and to [Fer+23] for an overview of recent applications.

In practice, the paths encountered in data-driven applications are typically constructed via piecewise linear interpolation of discrete time series. By combining Chen’s identity with the fact that the signature of a linear segment is given by the tensor exponential of its increment, one obtains a concise formula for the signature of a piecewise linear path  $\gamma = \gamma^1 * \dots * \gamma^L$  as  $\mathcal{S}(\gamma) = \exp(\Delta\gamma^1) \otimes \dots \otimes \exp(\Delta\gamma^L)$ . This expression underlies the implementation of signature computations in standard Python libraries such as `esig` [Lar10], `iisignature` [RG18], and `signatory` [KL21]. It enables an efficient and exact evaluation of the signature, with time complexity  $\mathcal{O}(Ld^n)$ , where  $n \in \mathbb{N}$  denotes the truncation level. However, due to the exponential growth in the dimension  $d$ , the method becomes computationally prohibitive as  $n$  increases. A commonly-taken approach solution to this dimensionality issue making use of signature kernel methods has become popular [Lem+21b; Sal21; Lem+21a; Man+24; SS24]. These allow for efficient computation of kernels of the form  $\langle \mathcal{S}(\gamma), \mathcal{S}(\sigma) \rangle$ , for various choices of inner product  $\langle \cdot, \cdot \rangle$  on  $T((\mathbb{R}^d))$ , without the need for computing  $\mathcal{S}(\gamma)$  directly, for example leveraging the recent result that the signature kernel satisfies a Goursat PDE [Sal+21a].

Another promising recent alternative to these ideas is a proposal to use the development of  $\gamma$  into a matrix Lie group [LLN22]. In this formulation, the central object becomes not the signature but related representations obtained by the evolution

$$dU(\gamma; M)_t = U(\gamma; M)_t \cdot M(d\gamma_t) = \sum_{j=d}^d U(\gamma; M)_t A_j d\gamma^j, \quad U(\gamma; M)_0 = I_N, \quad (1)$$

where  $U(\gamma; M)$  belongs to a matrix Lie group  $G_N$ ,  $M$  is a map in  $\text{Hom}(\mathbb{R}^d, \mathfrak{g}_N)$  such that  $M(v) = \sum_{j=1}^d A_j v^j$  for  $A_1, \dots, A_d \in \mathfrak{g}_N$ , with  $\mathfrak{g}_N$  denoting the associated Lie algebra, the latter being isomorphic to the space of left-invariant vector fields on  $G_N$ , and in which  $\cdot$  denotes matrix multiplication. In

this work we consider the physical and geometrical interpretation of the development of  $\gamma$  as a parallel transport operator on a  $G_N$  bundle. The solution to (1) can be expressed in terms of the signature as

$$U(\gamma; M)_t = \sum_{w \in \mathcal{W}_d} i^{|w|} A_w \mathcal{S}_{0,t}^w(\gamma), \quad (2)$$

where the sum is taken over all words in  $d$  letters. An important example of this framework is when  $G_N = U(N)$  is the unitary group, whose corresponding Lie algebra  $\mathfrak{g}_N = \mathfrak{u}(N)$  is the set of anti-Hermitian matrices. We note that 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. A promising approach recently proposed in the literature [CT24a; MLS23] consists of randomising the choice of vector fields  $M = (A_1, \dots, A_d) \sim \mu_V^N$  over a suitable class of (Borel) probability measures  $\mu_V^N$  on the Lie algebra and study the limiting behaviour of (inner products of such) path developments as  $N$  tends to infinity.

The physical perspective adopted in this work interprets path development as parallel transport in gauge theory. This viewpoint leads to an interpretation of vector field randomisation as a zero-dimensional quantum field theory, where integration over gauge fields is realised via a matrix-valued path integral. Conversely, we introduce path developments as a novel class of observables in matrix model theory, interpolating between zero-dimensional quantum field theory and Yang–Mills theory. More precisely, we will be interested in characterising the large- $N$  behaviour of the law of the random variables

$$\langle U(\gamma; M), U(\sigma; M) \rangle_{\text{HS}} := \lim_{N \rightarrow \infty} \frac{1}{N} \text{Tr}(U(\gamma; M) U(\sigma; \cdot)^\dagger) \quad \text{under } M \sim \mu_V^N, \quad (3)$$

where, importing tools from theoretical physics, we consider  $\mu_V^N$  a matrix model measure on  $\mathfrak{u}(N)$  given by [GM06; GM07]

$$d\mu_V^N(M) \propto \exp(-N \text{tr}(V(M))) dM, \quad (4)$$

where  $\text{tr}(\cdot)$  denotes the normalised trace defined by  $\text{tr}(\cdot) := \frac{1}{N} \text{Tr}(\cdot)$  and  $V(A)$  denotes a Gaussian-perturbed polynomial potential

$$V(A) = V(A_1, \dots, A_d) = \frac{1}{2} \sum_{i=1}^d A_i^2 + W, \quad (5)$$

where  $W$  is a polynomial satisfying some suitable convexity conditions. By combining techniques from non-commutative probability theory [Gui19] and rough path theory [CS24], we can define the limit

$$k^{\mu_V^\infty}(\gamma, \sigma) := \lim_{N \rightarrow \infty} \mathbb{E}_{M \sim \mu_V^N} [\text{tr}(U(\gamma; M) U(\sigma; M)^\dagger)], \quad (6)$$

and moreover understand  $k^{\mu_V^\infty}(\gamma, \sigma)$  in terms of the signature of the path  $y = \gamma * \overleftarrow{\sigma}$ :

$$k^{\mu_V^\infty}(\gamma, \sigma) = \sum_{w \in \mathcal{W}_d} i^{|w|} \lim_{N \rightarrow \infty} \mathbb{E}_{\mu_V^N} [\text{tr}(A_w)] \mathcal{S}^w(y) \quad (7)$$

$$:= \sum_{w \in \mathcal{W}_d} i^{|w|} \tau_V(A_w) \mathcal{S}^w(y), \quad (8)$$

where  $\tau_V(A_w)$  is a non-commutative law satisfying the Schwinger-Dyson equation (Theorem 3.9). We use this structure to prove that  $k^{\mu_V^\infty}(\gamma, \sigma) = k^y(1, 1)$ , where  $k^y : [0, 1]^2 \rightarrow \mathbb{R}$  satisfies an integro-differential equation of the form (Theorem 3.14)

$$\langle \gamma_{s,t} \rangle_{\mu_V^\infty} = 1 - \int_{s \leq u \leq v \leq t} \langle \gamma_{s,u} \rangle_{\mu_V^\infty} \langle \gamma_{u,v} \rangle_{\mu_V^\infty} \langle d\gamma_u, d\gamma_v \rangle - \sum_{k=1}^d \int_s^t \mathcal{D}_V^k \langle \gamma_{s,u} \rangle_{\mu_V^\infty} d\gamma_u^k, \quad (9)$$

with the boundary condition  $\langle \gamma_{s,s} \rangle_{\mu_V^\infty} = 1$  for all  $s \in [0, T]$ , and where the pathwise derivative  $\mathcal{D}_V^k$  will be rigorously defined in [Section 3](#).

This result generalises the findings in [\[CT24a\]](#), where the special case of quadratic potential  $V = \frac{1}{2} \sum_{i=1}^d X_i^2$  was considered along with variants<sup>1</sup> leading to the same limiting PDE, for which  $\mu_V^N$  corresponds to the classical Gaussian unitary ensemble (GUE). We show in addition that the integro-differential equation (9) has a unique solution when the solution space is restricted to (infinite) linear functionals on the signature.

In [Section 4](#) we turn our attention to the GUE ensemble and piecewise linear paths  $\gamma : [0, T] \rightarrow \mathbb{R}^d$  given as  $L$  specified increment vectors  $\{\Delta_l^1, \Delta_l^2, \dots, \Delta_l^d\}_{l=1}^L$ . We consider the path development as a unitary operator on  $\mathcal{H} = (\mathbb{C}^2)^{\otimes n}$ , a Hilbert space of  $n$  qubits. A linear basis of generators of unitary operators on qubits is given by Pauli strings: length  $n$  tensor products of the Pauli matrices  $P = \{\sigma_I, \sigma_X, \sigma_Y, \sigma_Z\}$ . We define a random ensemble of linear combinations of Pauli strings

$$A_\nu = \sum_{\mathbf{w} \in P} \alpha_\nu^{\mathbf{w}} \sigma_{\mathbf{w}}, \quad \nu = 1, \dots, d, \quad (10)$$

where the coefficients are randomly distributed as defined precisely in [Definition 4.6](#).

We first demonstrate that the GUE randomised path development may be approximated by a development driven by this sparse ensemble of random Pauli strings ([Theorem 4.8](#)). We then consider a Trotterisation of this path development into  $K$  subdivisions and thereby define a random quantum circuit approximating the unitary development. The circuit depends on the parameters:  $K$  Trotter subdivisions,  $m$  densities of Pauli ensemble and  $n$  qubits, and is given by

$$U_\gamma^Q(\alpha(m), n, K) = \prod_{l=1}^L \left[ \prod_{\nu=1}^d \prod_{i=1}^m P_{\mathbf{w}_{i,\nu}}(\Delta_l^\nu \alpha_\nu^{\mathbf{w}_i} / K) \right]^K, \quad (11)$$

where  $P_{\mathbf{w}}(\theta)$  is a Pauli rotation of angle  $\theta$  associated to the Pauli string  $\sigma_{\mathbf{w}}$ . Finally, we use this circuit to propose an efficient quantum algorithm ([Theorem 4.11](#)) in the one clean qubit model of quantum computation to approximate the GUE path development and the associated kernel.

**Outline and summary.** This work lies at the intersection of matrix models in quantum field theory (QFT) and the theory of path signatures, with a particular focus on their applications in machine learning and data science. Our aim is to make this work accessible to researchers from both communities. To that end, we provide a concise introduction to the relevant background and establish our notation in [Section 2](#). In [Section 3](#), we apply classical tools from non-commutative probability theory to derive the integro-differential equation (9) satisfied by  $\langle \gamma_{s,t} \rangle_{\mu_V^\infty}$  under small perturbations of the Gaussian potential defined by matrix models. Finally, in [Section 4](#), we introduce quantum analogues of the classical path signature and signature kernel. The main results of this section demonstrate that the path development driven by the random Pauli ensemble approximates well the unitary path development and that the associated quantum algorithm is efficient.

**Acknowledgements.** SC, AJ and CS are supported by the Innovate UK "Quantum Machine Learning for Financial Data Streams" (Project ref 10073285), in collaboration with Standard Chartered and Rigetti Computing. SC, AJ and CS are grateful to Marco Painsi (Rigetti), Ernesto Palidda (Rigetti) and Mattia Fiorentini (Rigetti) for many interesting discussions. AJ is also supported by the EPSRC Grant EP/W032643/1. TC is supported by the EPSRC Programme Grant EP/S026347/1. SC is supported by National Science and Technology Council of Taiwan (NSTC 113-2112-M-007-019) and is grateful to Masazumi Honda (RIKEN) and Heng-Yu Chen (NTU) for many interesting discussions and RIKEN for hospitality whilst part of this work was completed. WFT has been supported by the EPSRC Centre for Doctoral Training in Mathematics of Random Systems: Analysis, Modelling and Simulation (EP/S023925/1). SC, WFT and CS are grateful to Pochung Cheng and Po-Yao Chang for their organisation of the workshop 'Random matrices and quantum machine

<sup>1</sup>The work [\[CT24a\]](#) shows, in addition, stability of the limit beyond the setting of i.i.d. Gaussian entries.

learning' at National Tsing Hua University, Taiwan where part of this work was finalised. For the purpose of open access, the authors have applied for a creative commons attribution (CC BY) licence to any author accepted manuscript version arising.

## 2 Background

In this section, we provide a brief introduction to quantum field theory (QFT) in zero dimensions and path signatures and set the notation for the rest of the work. We also introduce a novel geometric-physical interpretation of path signatures and signature kernels as matrix model Wilson line correlation functions.

### 2.1 Quantum field theory

The physical context of our work is that of zero-dimensional QFT in the form of Hermitian matrix models. A thorough introduction to zero-dimensional QFT aimed at a mathematical audience may be found in [Horo3] and we refer the reader to [EKR15] for a comprehensive review of matrix model theory. The theory of matrix models is a classical subject in theoretical Physics [tH93] with deep connections to enumerative geometry via topological recursion [EO07] and quantum gravity in two dimensions [GM90]. We provide a cursory overview of the subject and focus on reviewing the basic aspects of loop operators that the present work generalises.

We consider matrix-valued fields fluctuating on a zero-dimensional spacetime. Let us consider the unitary matrix Lie group

$$U_N = \{U \in \mathcal{M}_N(\mathbb{C}) : U^\dagger U = I_N\}, \quad (12)$$

where  $\mathcal{M}_N(\mathbb{C})$  denotes  $N \times N$  complex matrices, and the corresponding Lie algebra

$$\mathfrak{u}_N = \{A \in \mathcal{M}_N(\mathbb{C}) : A^\dagger = -A\}. \quad (13)$$

We now consider a connection (or gauge field in the Physics terminology); a Lie algebra-valued one-form which, in coordinates, may be expressed as an  $N \times N$  matrix  $A \in \mathfrak{u}_N$ . Global gauge transformations act on these matrices by the adjoint action  $g \cdot A = g^{-1} A g$  with  $g \in U_N$ . In QFT, the Feynman path integral averages over fluctuating field configurations weighted by a gauge-invariant action. In the present context we integrate over anti-Hermitian  $N \times N$  matrices and the measure we choose is a so-called matrix model measure

$$d\mu_V(A) = \exp\left(-N \operatorname{tr}(V(A))\right) dA.$$

We study interacting polynomial potentials of the form

$$V(A, \{g_k\}) = \frac{1}{2} A^2 + \sum_{k=3}^m \frac{g_k}{k!} A^k,$$

depending on a set of coupling constants  $\{g_k\}_{k \geq 3}$  and  $A \in \mathfrak{u}_N$ . We will provide a more mathematically precise exposition of such matrix model measures in the context of non-commutative probability theory in Section 3. Finally, we should specify the set of (gauge-invariant) observables we are interested in computing. We write expectation values as

$$\left\langle \operatorname{tr}(f(A)) \right\rangle := \frac{1}{Z} \int d\mu_V(A) \operatorname{tr}(f(A)), \quad (14)$$

for a matrix function  $f(A)$  and where  $Z := \langle 1 \rangle$  denotes the partition function. Solving the matrix model corresponds to determining the moment generating function. In the matrix model context, this object is the multi-trace correlation function<sup>2</sup>

$$W_n(x_1, \dots, x_n; \{g_k\}) := \left\langle \operatorname{tr}\left(\frac{1}{x_1 I - A}\right) \cdots \operatorname{tr}\left(\frac{1}{x_n I - A}\right) \right\rangle_{\mu_V, c} \quad (15)$$

<sup>2</sup>The right hand side of this expression is shorthand for the formal generating function of the single trace correlators  $\langle \operatorname{tr}(A^{l_1}) \cdots \operatorname{tr}(A^{l_n}) \rangle_c$ .

where the subscript  $c$  denotes the connected part of the correlator. Correlation function with  $n$  insertions enjoy a formal asymptotic expansion, the so-called genus expansion, of the form

$$W_n(x_1, \dots, x_n; \{g_k\}) = \sum_{g=0}^{\infty} N^{2-2g-n} W_{g,n}(x_1, \dots, x_n; \{g_k\}) \quad \text{as } N \rightarrow \infty. \quad (16)$$

The single trace function  $W_{0,1}$  satisfies an algebraic equation, known as the spectral curve of the matrix model. Its Laplace transform, the so-called loop operator denoted  $\widehat{W}_{0,1}$ , may be expressed as

$$\widehat{W}_{0,1}(t) = \lim_{N \rightarrow \infty} \frac{1}{N} \left\langle \text{tr}(e^{tA}) \right\rangle, \quad (17)$$

and solves an integral equation known as the loop equation<sup>3</sup>

$$V' \left( \frac{d}{dt} \right) \cdot \widehat{W}_{0,1}(t) = \left( \widehat{W}_{0,1} \star \widehat{W}_{0,1} \right)(t). \quad (18)$$

In the present work we develop a generalisation of this matrix model. We study a  $d$ -dimensional Hermitian matrix model together with a smooth path  $\gamma : [0, 1] \rightarrow \mathbb{R}^d$  and replace the exponentials in the loop operator with path-ordered exponentials

$$\widehat{W}_{0,1}(t) = \lim_{N \rightarrow \infty} \left\langle \text{tr}(e^{tA}) \right\rangle \longrightarrow \widehat{W}_{0,\gamma}(t) = \lim_{N \rightarrow \infty} \left\langle \text{tr} \left( \text{P exp} \int_{\gamma} A \right) \right\rangle. \quad (19)$$

In the following we derive corresponding loop equations generalising (18) for these path loop operators.

## 2.2 Wilson lines

We now turn to the study of smooth paths  $\gamma : [0, T] \rightarrow \mathbb{R}^d$ . In the context of learning and kernel methods, such paths could be, for example, trajectories of the price a collection of financial instruments or patient heart rate data. In the present context we think of them as the world line of a charged particle moving through a gauge field—a so-called *Wilson line* in the physical terminology—we now explain this construction.

We consider  $V = \mathbb{R}^d$  as a smooth Riemannian manifold with the Euclidean metric.<sup>4</sup> We consider a (topologically trivial) principal  $G = U_N$  bundle  $P = V \times G \xrightarrow{\pi} V$ . A connection on  $P$  is then uniquely specified by a Lie algebra valued one-form, written  $A \in \Omega^1(V, \mathfrak{g})$ , and we consider the case where  $A$  is constant on  $V$ . A path  $\gamma : [0, T] \rightarrow V$  then has a unique horizontal lift<sup>5</sup>  $\tilde{\gamma} : [0, T] \rightarrow P$  where  $\tilde{\gamma}(t) = (\gamma(t), g(t))$  is determined by the solution to the ODE

$$\dot{g}(t) = g(t) \cdot A(\dot{\gamma}(t)), \quad g(0) = e. \quad (20)$$

We now consider a finite-dimensional  $G$ -representation  $\mathcal{H}$ . We will take throughout  $G = U_N$  with  $\mathcal{H} = \mathbb{C}^N$  the fundamental representation and therefore, in a minor abuse of notation, conflate  $A$  with its representation. Following the Dirac notation convention, vectors in the associated bundle are written as  $|\psi\rangle \in \mathcal{H}$ . We then define the topologically trivial associated bundle by  $\mathbb{H} = V \times \mathcal{H}$  and write the fibre at  $x \in V$  as  $\mathcal{H}_x$ . The parallel transport of a vector  $|\psi\rangle \in \mathcal{H}_{\gamma(0)}$  along  $\gamma$  from  $\gamma(0) \in V$  to  $\gamma(T) \in V$  is then determined by the solution to (20). In coordinates this equation reads

$$dU_{0,t}(\gamma) = U_{0,t}(\gamma) \cdot A_{\mu} d\gamma^{\mu}, \quad U_{0,0} = I_N,$$

and the solution to this equation defines the path-ordered exponential

$$U_{0,T}(\gamma) = \text{P exp} \left( \sum_{\mu=1}^d \int_0^T A_{\mu} d\gamma^{\mu} \right). \quad (21)$$

<sup>3</sup> $\star$  denotes convolution.  $(f \star g)(t) = \int_0^t ds f(s)g(t-s)$ .

<sup>4</sup>The following construction applies more generally to parallelisable manifolds.

<sup>5</sup>A horizontal lift of a path  $\gamma \in V$  is a path  $\tilde{\gamma} \in P$  such that  $\pi(\tilde{\gamma}) = \gamma$  and  $\dot{\tilde{\gamma}}$  lies in the horizontal subspace  $V \cong TV$ .



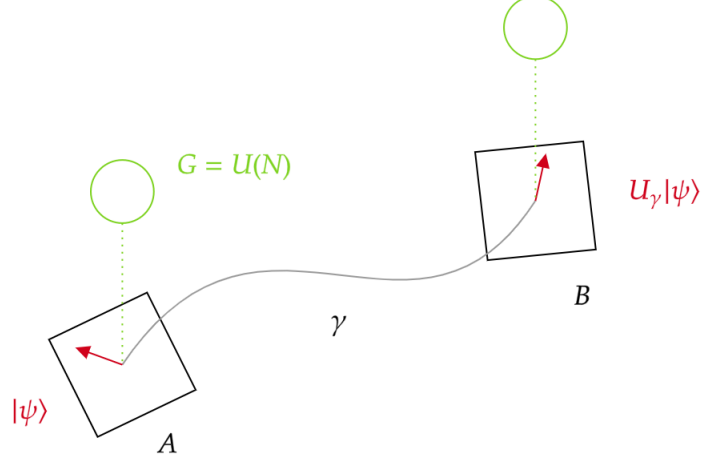


Figure 1: The path development as parallel transport in a  $G = U(N)$  bundle along a Wilson line.

The setup is illustrated in Fig. 1. Later, in Section 4, the associated bundle will be further interpreted physically as a  $N = 2^n$  dimensional Hilbert space of qubits  $\mathcal{H} = (\mathbb{C}^2)^{\otimes n}$  and  $|\psi\rangle$  will represent the state of a register of qubits. The development  $U_{0,T}(\gamma)$  is then a time-dependent many-body time evolution operator.

In the physical terminology  $\gamma : [0, T] \rightarrow V$  describes the worldline of a charged particle moving from  $\gamma(0)$  to  $\gamma(T)$  through a (constant) electromagnetic field described by the connection  $A$ , that is a Wilson line. The charge of the particle is described by a vector in  $\mathcal{H}$  and the parallel transport operator (21) describes how the charge changes throughout the motion. The constant connection has non-trivial curvature and thus introduces a path-dependence in this charge evolution. In the data science context, this vector may thus be used as a feature associated to the path.

In the context of defining kernels on path space, we will later be interested in the expectation values of the trace of this quantity,  $\text{tr}(U_{0,T}(\gamma))$ . In summary, we are interested in integrating over the choice of connection using  $d$ -dimensional generalisations of the matrix model integrals discussed in the previous subsection.

**Remark 2.1.** *The quantum theory we define interpolates between zero-dimensional QFT and  $d$ -dimensional Yang-Mills theory. The unitary development  $U_{0,T}(\gamma)$  may be considered more generally as a Wilson line in Yang-Mills theory on  $\mathbb{R}^d$  where  $A_\mu$  now varies over spacetime. Instead, we simplify the setup and consider a constant off-shell gauge field on a parallel manifold averaged over a matrix model.*

### 2.3 Path signatures

When analysing the path-ordered exponentials appearing in (21), it will be convenient to consider power series expansions. The terms involving the path  $\gamma$  will be given by its iterated integrals, the collection of which is called its *path signature*.<sup>6</sup> Arising in the work of K.T. Chen in the 1950s [Che54; Che57] and forming a central part of Lyons' rough path theory [Ly098], the signature has more recently found success both as a theoretical and as a practical tool in modern machine learning. The signature forms the analogue of the collection of monomials for vector valued data: under certain conditions, linear functionals of the signature are dense in continuous functions on compact subsets of (unparameterised) path space [LLN16; CT24b]. Before we give the definition of a signature we require some basic definitions.

**Definition 2.2** (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 base field  $\mathbb{C}$  and multiplication defined by

$$wu := w_1 \cdots w_n u_1 \cdots u_m,$$

<sup>6</sup>In fact, the path signature itself fits into the geometric setup of Section 2.2 as a parallel transport on a principal  $T((V))$  bundle. The corresponding Lie algebra  $\mathcal{L}(V)$  has a universal property meaning its  $\mathfrak{g}$ -representations are determined by homomorphisms  $\phi : V \rightarrow \mathfrak{g}$ . Hence connections on  $G$ -bundles determine representations of  $T((V))$  uniquely.

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.

**Definition 2.3** (Time-Simplex). For any  $n \in \mathbb{N}$ , define the  $n$ -simplex over an interval  $[a, b]$  by

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

Throughout, we will take  $V \equiv \mathbb{R}^d$  equipped with the standard Euclidean inner product. The signature of a path  $\gamma : [0, T] \rightarrow V$  is then defined as follows.

**Definition 2.4** (Paths and Signatures). Let  $\mathcal{X} := C^{1\text{-var}}([0, T], V)$  be the space of continuous 1-variation paths  $\gamma : [0, T] \rightarrow \mathbb{R}^d$  with  $\gamma_0 = 0$ . For  $\gamma \in \mathcal{X}$ , the signature of  $\gamma$  is a function

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

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

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

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

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

where  $\mathcal{S}_{s,t}^{\emptyset}(\gamma) := \mathcal{S}_{s,t}(\gamma)^0 = 1$ . We will commonly write  $\mathcal{S}(\gamma) := \mathcal{S}_{0,T}(\gamma)$  for the signature over the entire interval.

**Proposition 2.5** (Chen's identities). *We define the concatenation of two paths  $\gamma, \sigma \in \mathcal{X}$  by*

$$(\gamma \star \sigma)_t := \begin{cases} \gamma_{2t}, & t \in [0, T/2], \\ \gamma_T + \sigma_{2(t-T/2)}, & t \in [T/2, T]. \end{cases}$$

*Then*

$$\mathcal{S}(\gamma \star \sigma) = \mathcal{S}(\gamma) \otimes \mathcal{S}(\sigma).$$

*Additionally, we may define the reversal of the path  $\gamma$  by*

$$\overleftarrow{\gamma}_t := \gamma_{T-t} - \gamma_T.$$

*Then*

$$\mathcal{S}(\gamma \star \overleftarrow{\gamma}) = \mathcal{S}(\gamma) \otimes \mathcal{S}(\overleftarrow{\gamma}) = \mathbf{1}.$$

Chen's identities imply that the space of signatures  $\mathcal{S}(\mathcal{X}) \subset T((V))$  forms a group with multiplication given by path concatenation and inverse given by path reversal. Using the signature, we may expand (21) as

$$U_{0,T}(\gamma) = \text{P exp} \left( \sum_{i=1}^d \int_0^T A_i d\gamma_u^i \right) = \sum_{\mathbf{w} \in \mathcal{W}_d} i^{|\mathbf{w}|} A_{\mathbf{w}} \mathcal{S}_{a,b}^{\mathbf{w}}(\gamma), \quad (22)$$

where  $i$  denotes the imaginary unit. In this work, particularly later in the context of kernels, we are interested in computing the expectation value of this quantity over the matrix models in the large  $N$  limit discussed in the previous subsection. Provided the quantities exist, we introduce the notations



$$\langle \gamma_{0,T} \rangle_{\mu^N} := \sum_{\mathbf{w} \in \mathcal{W}_d} i^{|\mathbf{w}|} \mathbb{E}_{\mu^N} [\text{tr}(A_{\mathbf{w}})] \mathcal{S}_{a,b}^{\mathbf{w}}(\gamma), \quad (23)$$

$$\langle \gamma_{0,T} \rangle_{\mu^\infty} := \sum_{\mathbf{w} \in \mathcal{W}_d} i^{|\mathbf{w}|} \lim_{N \rightarrow \infty} \mathbb{E}_{\mu^N} [\text{tr}(A_{\mathbf{w}})] \mathcal{S}_{a,b}^{\mathbf{w}}(\gamma). \quad (24)$$

Finally, we note that in the case  $d = 1$  when  $\gamma_t = t$  is a straight line, the signature becomes  $\mathcal{S}_{0,T}(\gamma)^n = T^n/n!$ . The right-hand side of (23) then becomes the matrix model moment generating function of trace correlators discussed in the previous section

$$\langle 1_{0,T} \rangle_{\mu^N} = \sum_{n=0}^{\infty} \langle \text{tr}(A^n) \rangle_{\mu^N} \frac{T^n}{n!} = W_N(T).$$

The present article may then be considered a generalisation of traditional Physics matrix models to the case of  $d$  multi matrices and the observable is upgraded to a Wilson line.

## 2.4 Signature kernels

We now discuss kernels on path space which are important from an applications perspective. Typically, one considers a feature map  $\varphi : \mathcal{X} \rightarrow \mathcal{H}$  that embeds the data  $\mathcal{X}$  into a typically high-dimensional (possibly infinite) Hilbert space  $\mathcal{H}$  where linear methods may be applied. A kernel  $k : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$  is simply the inner product of the feature map, namely

$$k(x, y) := \langle \varphi(x), \varphi(y) \rangle_{\mathcal{H}}. \quad (25)$$

If one can access the Gram matrix  $G_{ij} = k(x_i, x_j)$  for all pairs of points in the dataset, then one can perform a variety of linear separation techniques without ever evaluating  $\varphi$ . Such an ability is known as a “kernel trick”. Conversely, every kernel corresponds to some feature map into some Hilbert space, known as the reproducing kernel Hilbert space (RKHS). Indeed let  $\mathcal{H}^0 = \text{span}\{k(x, \cdot) : x \in \mathcal{X}\}$ , and for functions

$$f = \sum_{i=1}^n \alpha_i k(x_i, \cdot) \quad \text{and} \quad g = \sum_{j=1}^m k(y_j, \cdot),$$

define

$$\langle f, g \rangle_{\mathcal{H}^0} := \sum_{i=1}^n \sum_{j=1}^m \alpha_i \beta_j k(x_i, y_j),$$

then  $\mathcal{H} := \overline{\mathcal{H}^0}$  with respect to this inner product. This space satisfies the reproducing property

$$\langle f, k(x, \cdot) \rangle_{\mathcal{H}} = f(x), \quad \text{for all } f \in \mathcal{H} \text{ and } x \in \mathcal{X},$$

and the feature map  $\varphi$  is simply given by the embedding  $x \mapsto k(x, \cdot)$ . By equipping (an appropriate subspace of)  $T((V))$  with the standard Hilbert-Schmidt inner product, we arrive at the ordinary signature kernel defined by

$$k_{s,t}^{\text{sig}}(\gamma, \sigma) := \langle \mathcal{S}_{0,s}(\gamma), \mathcal{S}_{0,t}(\sigma) \rangle := \sum_{n=0}^{\infty} \langle \mathcal{S}_{0,s}(\gamma)^n, \mathcal{S}_{0,t}(\sigma)^n \rangle_{V^{\otimes n}},$$

which is of the form (25) with feature map  $\varphi(\cdot) = \mathcal{S}(\cdot)$ . The ordinary signature kernel comes with a kernel trick, being that  $k_{s,t}^{\text{sig}}$  solves the integral equation [Sal+21a]

$$k_{s,t}^{\text{sig}} = 1 + \int_0^s \int_0^t k_{u,v}^{\text{sig}} \langle d\gamma_u, d\gamma_v \rangle. \quad (26)$$

A variety of algorithms [LL24; CPP25; Sal+21a] have been proposed to solve (26), however, they all suffer from quadratic complexity in the length of the input time series. An alternative approach is to compute the inner product between low-dimensional random projections of the signature that approximately preserve their inner products. Indeed, consider the differential equation

$$dZ_t^\gamma = \sum_{i=1}^d Z_t^\gamma A_i d\gamma_t^i, \quad Z_0^\gamma = I_N,$$

where each  $A_i \in \mathbb{R}^{N \times N}$  is a random matrix with i.i.d.  $\mathcal{N}(0, \frac{1}{N})$  entries and  $N$  is large. Then  $Z^\gamma$  is a  $\text{GL}_N(\mathbb{R})$ -valued random variable where

$$\langle Z_s^\gamma, Z_t^\sigma \rangle := \frac{1}{N} \text{Tr}((Z_s^\gamma)^\top Z_t^\sigma) \xrightarrow{N \rightarrow \infty} k_{s,t}^{\text{sig}}(\gamma, \sigma),$$

see for example [MLS23; CT24a]. Solving for  $Z^\gamma$  is then only linear in the length of the input time series. Supposing that  $\gamma = \gamma_1 \star \dots \star \gamma_n$  is a piecewise linear paths valued in  $\mathbb{R}^d$  with linear increments  $\gamma_i$  for  $i = 1, \dots, n$ , then  $Z_T^\gamma$  is given exactly by the matrix product

$$Z_T^\gamma = \exp\left(\sum_{i=1}^d A_i \gamma_1^i\right) \cdots \exp\left(\sum_{i=1}^d A_i \gamma_n^i\right).$$

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  $A_i$ . However, even for sparse  $A_i$ , each matrix exponential is typically dense and so the products are still an obstacle. Supposing however, that the matrices  $A_i$  where instead anti-Hermitian, then each matrix exponential would be a unitary matrix with sparse representation. Such a setting naturally leads one to consider whether quantum algorithms are applicable, which is the study of Section 4.3.

In alignment with the previous discussion, in the present work we are interested in kernels on paths induced by unitary matrix path developments, that is the Wilson line observables introduced above averaged over matrix models with potential  $V$ . The Lie group  $U_N$  has a natural unitary inner product (the Hilbert-Schmidt inner product) and we may define

$$k^{\mu_V^N}(\gamma, \sigma) := \langle \text{tr}(U_\gamma U_\sigma^\dagger) \rangle_{\mu_V^N} := \frac{1}{N} \langle \text{Tr}(U_\gamma U_\sigma^\dagger) \rangle_{\mu_V^N}, \quad (27)$$

where we have replaced  $Z^\gamma$  with  $U_\gamma$  to reflect the unitary nature of the developments. The particular case where  $\mu_V$  is the free Gaussian multi-matrix model, i.e. the Gaussian Unitary Ensemble (GUE), will be the focus of Section 4 where we will discuss its interpretation as the kernel associated to a quantum feature map built from an efficient quantum circuit. Further, in the case that  $\mu_V$  is a free Gaussian multi-matrix model, a governing integral for

$$k^{\text{GUE}}(\gamma, \sigma) := \lim_{N \rightarrow \infty} k^{\mu_V^N}(\gamma, \sigma) = \lim_{N \rightarrow \infty} \langle \text{tr}(U_\gamma U_\sigma^\dagger) \rangle_{\mu_V^N}$$

was derived in [CT24a]. The derivation of the limiting functional equation relied heavily on the Schwinger-Dyson equation that is asymptotically satisfied by the GUE ensemble. Section 3 is devoted to finding the corresponding functional equations in the case of a small perturbation of the Gaussian law. Such perturbations asymptotically satisfy Schwinger-Dyson equations that are perturbations of the one asymptotically satisfied by the GUE ensemble. Kernels like  $k^{\text{GUE}}$  are examples of kernels where the feature map is not known explicitly, and whose existence is only guaranteed through the construction of its RKHS.

**Geometric interpretation.** The kernel also enjoys a geometric interpretation in the context of Section 2.2, the setup is illustrated in Fig. 2. We begin with a charge vector  $|\psi\rangle$  living in the associated  $N$  dimensional  $U_N$  bundle  $\mathcal{H}$ . In Euclidean space  $\mathbb{R}^d$ , we may identify  $\mathcal{H}_{\tau(0)}$  with  $\mathcal{H}_{\sigma(0)}$  and  $\mathcal{H}_{\tau(T)}$  with  $\mathcal{H}_{\sigma(T)}$ . The kernel is

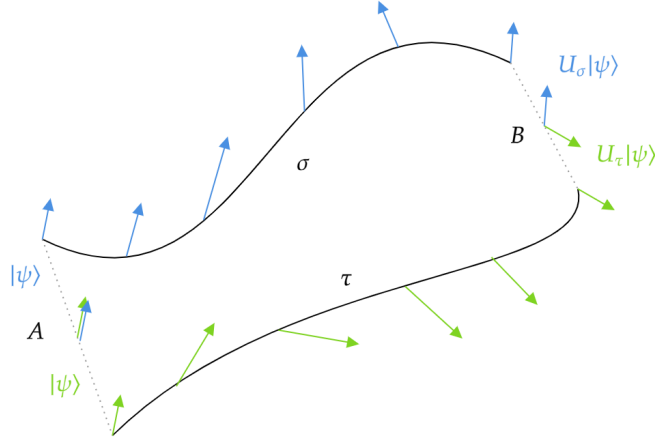


Figure 2: The kernel as parallel transport along  $\tau$  compared with parallel transport along  $\sigma$ .

then a measure of the difference between a charge vector  $|\psi\rangle$  being transported along  $\tau$ , compared with the same vector being transported along  $\sigma$ . That is we may compare the angles of  $U_\tau|\psi\rangle$  with  $U_\sigma|\psi\rangle$  by taking the Hermitian inner product. If we average the resulting product uniformly over a basis  $\{|\psi_i\rangle\}$  of  $\mathcal{H}$  then we find

$$\frac{1}{N} \sum_{i=1}^N \langle \psi_i | U_\sigma^\dagger U_\tau | \psi_i \rangle = \frac{1}{N} \text{tr}(U_\sigma^\dagger U_\tau),$$

the right-hand side is the definition of the kernel as the Hilbert-Schmidt inner product and thus we have a geometrical interpretation of the path development signature kernel as the average change in angle between charge vectors under two Wilson lines.

### 3 Universal Limit of Matrix Model Unitary Developments

This section is dedicated to deriving integro-differential equations satisfied by  $\langle \gamma_{s,t} \rangle_{\mu_V^\infty}$  for small perturbations of the Gaussian potential. We start by relating this to the quantity  $k^{\mu\nu}(\tau, \sigma)$ . By standard properties of path developments,  $U_\tau U_\sigma = U_{\tau \star \sigma}$  and  $U_\sigma^{-1} = U_{\overleftarrow{\sigma}}$ . Since the inverse of a unitary matrix is its conjugate transpose, it holds that

$$k^{\mu\nu}(\tau, \sigma) = \langle \text{tr}(U_{\tau \star \overleftarrow{\sigma}}) \rangle_{\mu_V}$$

As such, it is enough to understand the limiting functional  $\langle \gamma_{s,t} \rangle_{\mu_V^\infty}$  for a single path  $\gamma$  and replace it by  $\gamma = \tau \star \overleftarrow{\sigma}$  when conducting a kernel evaluation, which is what we do in this section. The case where  $V$  corresponds to the Gaussian potential  $\frac{1}{2} \sum_{i=1}^d X_i^2$  was considered in [CT24a] along with variants leading to the same limit. In the present work we extend these results to the matrix models discussed above in Section 2. Recall that  $\mathfrak{u}_N = i\mathfrak{h}_N$ , where  $\mathfrak{h}_N$  is the set of  $N \times N$  Hermitian matrices and that for a path  $\gamma \in \mathcal{X}$ , the unitary path development of path segment  $\gamma_{s,t}$  with development map  $M \in \text{Hom}(V, \mathfrak{u}_n)$  is the time  $t$ -solution to the  $\mathcal{M}_N(\mathbb{C})$ -valued controlled differential equation

$$dU_{s,u} = U_{s,u} \cdot M(d\gamma_u), \quad Z_{s,s} = I_N, \quad (28)$$

where  $\cdot$  denotes matrix multiplication. Fixing a basis for  $V \cong \mathbb{R}^d$  and noting that  $M(v) = i \sum_{j=1}^d A_j v^j$  for matrices  $A_j \in \mathfrak{h}_N$ , the differential equation reads

$$dU_{s,u} = \sum_{j=1}^d U_{s,u} A_j d\gamma_u^j, \quad (29)$$

whose solution is given by

$$U_{s,t} = \text{P exp} \left( i \sum_{j=1}^d \int_s^t A_j d\gamma_u^j \right) = \sum_{w \in \mathcal{W}_d} i^{|w|} A_w \mathcal{S}_{s,t}^w(\gamma),$$

where  $P \exp$  is the path-ordered exponential along the curve  $\gamma$ . We may randomise the solution to (28) via the randomisation of the collection  $(A_1, \dots, A_d) \sim \mu_V^N$ . In particular, our quantity of interest is  $\lim_{N \rightarrow \infty} \mathbb{E}[\text{tr}(U_{s,t})]$ .

### 3.1 Non-commutative laws and Schwinger-Dyson equations

In this section, we gather various notions both from the world of non-commutative probability theory and from rough path theory, in particular following [Gui19] for the non-commutative definitions. We shall use  $\bar{\cdot}$  to denote complex conjugation

**Definition 3.1** ( $C^*$  algebra). A  $C^*$  algebra  $\mathcal{A}$  is a Banach algebra together with an involution  $*$  :  $\mathcal{A} \rightarrow \mathcal{A}$  satisfying the following properties

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

**Definition 3.2.** 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  $\mathbf{w} \in \mathcal{W}_d$  we will write  $X_{\mathbf{w}} := X_{w_1} \cdots X_{w_n}$ , which may be extended by linearity to all  $\mathbf{w} \in \mathcal{A}_d$  with  $X_{\emptyset} = 1$ . This induces an algebra isomorphism  $\mathcal{A}_d \rightarrow \mathcal{C}_d$ .

**Definition 3.3** (Non-commutative law). A non-commutative law  $\tau \in \mathcal{C}_d^*$  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  $\tau(P P^*) \geq 0$ .
- 2) Mass:  $\tau(1) = 1$ .
- 3) Trace: for all  $P, Q \in \mathbb{C}\langle X_1, \dots, X_d \rangle$ ,  $\tau(PQ) = \tau(QP)$ .

A general non-commutative probability space is a pair  $(\mathcal{A}, \varphi)$  of a unital  $C^*$  algebra  $\mathcal{A}$  equipped with a state  $\varphi$ . However, given  $d$  non-commuting random variables  $a_1, \dots, a_d$  living in  $(\mathcal{A}, \varphi)$ , there exists a unique non-commutative law  $\tau$  on the free  $*$ -algebra  $\mathcal{C}_d$  such that  $\tau(P(X_1, \dots, X_d)) = \varphi(P(a_1, \dots, a_d))$  for any polynomial  $P$ . As such, Definition 3.3 is enough for our purposes.

**Definition 3.4.** 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 the 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

$$\mathbb{E}[X_i f(X_1, \dots, X_d)] = \mathbb{E}[\partial_i f(X_1, \dots, X_d)]. \quad (30)$$

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

$$\mathbb{E}[X_i^{n+1} X_{\mathbf{w}}] = n \mathbb{E}[X_i^{n-1} X_{\mathbf{w}}]. \quad (31)$$

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 (31) 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, (30), 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 3.5** (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

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

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

$$\partial_i X_w = \sum_{w=uv} X_u \otimes X_v. \quad (32)$$

**Definition 3.6** (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  $\partial_i$  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 3.7** (Schwinger-Dyson Equations). Let  $\tau : \mathcal{C}_d \rightarrow \mathbb{C}$  be a non-commutative law. We say that  $\tau$  satisfies a Schwinger-Dyson equation with conjugates  $P_i \in \mathcal{C}_d$  if, for all  $1 \leq i \leq d$ ,

$$\tau \otimes \tau(\partial_i P) = \tau(P P_i), \quad (33)$$

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  $\tau$ , then  $\tau(X_1 X_2 X_1 X_2) = 0$ , whereas 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 3.8** (Semicircular Variables). The law  $\tau$  of  $d$ -free semicircular random variables satisfies the Schwinger-Dyson equation

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

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

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

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  $\text{NC}_2^n$  is the set of *non-crossing* pair partitions of  $n$  letters.

An important class of Schwinger-Dyson equations, including [Example 3.8](#), are those where the conjugate polynomials  $P_i$  are given by the cyclic derivatives of some polynomial  $V \in \mathcal{C}_d$ . In this case, we have

$$\tau \otimes \tau(\partial_i P) = \tau(PD_i V). \quad (34)$$

The non-commutative laws satisfying the Schwinger-Dyson equation of type (34) arise as the asymptotic limit of the non-commutative laws of interacting random matrix ensembles. For a self-adjoint polynomial  $V \in \mathcal{C}_d$ , consider the law  $\mu_V^N$  on  $N \times N$  Hermitian matrices given by

$$d\mu_V^N(X_1^N, \dots, X_d^N) = \frac{1}{Z_V^N} \exp \left\{ -N \text{Tr}(V(X_1^N, \dots, X_d^N)) \right\} dX^N,$$

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  $Z_V^N$  is a normalising constant. In order to ensure the partition function  $Z_V^N$  is finite for all  $N$ , some restrictions on  $V$  are required. Assume that  $V$  takes the form  $V = \frac{1}{2} \sum_{i=1}^d X_i^2 + W$  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

$$(X_\nu^N(jk))_{\substack{1 \leq j \leq k \leq N \\ 1 \leq \nu \leq d}} \mapsto \text{Tr}(W(X^N)) + \frac{1-c}{2} \sum_{i=1}^d \text{Tr}(X_i^2) \quad (35)$$

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 (35) holds. For  $\varepsilon > 0$  we also write  $B_\varepsilon$  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 summarise now a collection of results from [\[GMo6; GMo7\]](#), see also [\[Gui9, Chapter 7\]](#).

**Theorem 3.9** ([\[GMo6; GMo7\]](#)). *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  $\mathbf{g} \in U_W \cap B_\varepsilon$*

$$\lim_{N \rightarrow \infty} \mathbb{E}_{\mu_V^N} [\text{tr}(X_{\mathbf{w}}^N)] = \tau_V(X_{\mathbf{w}}),$$

for every  $\mathbf{w} \in \mathcal{W}_d$  where  $\tau_V$  solves the Schwinger-Dyson equation (34). Moreover, there is exactly one solution  $\tau_V \in \mathcal{C}_d^*$  to (34) and it satisfies

$$|\tau_V(X_{\mathbf{w}})| \leq R^{|\mathbf{w}|} \text{ for all } \mathbf{w} \in \mathcal{W}_d.$$

Note that when  $W = 0$ , the distribution  $\mu_V^N$  is the Gaussian Unitary Ensemble (GUE) and the limiting law  $\tau_V$  is the law of  $d$  free semicircular random variables. From now on, we will only consider those  $V$  satisfying the assumptions of [Theorem 3.9](#). 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 3.10** (Sub-Gaussian tail of largest eigenvalue). *Let  $V$  be a potential satisfying the conditions of [Theorem 3.9](#), and for each  $N$  suppose that  $(A_i)_{i=1}^d \sim \mu_V^N$ . Then there exist  $\alpha > 0$  and  $t_0 < \infty$  such that, for every  $t \geq t_0$  and  $N$ ,*

$$\mathbb{P}_{\mu_V^N} [\lambda_{\max}^N(\mathbf{A}) \geq t] \leq e^{-\alpha N t^2}.$$

*Proof.* See [Appendix A](#). □

### 3.2 Path loop equations

We now use the Schwinger-Dyson equations to derive governing equations for the trace of the path development in the large  $N$  limit. 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]. This notion of derivative has been popular in deriving Taylor-like expansions of functions on path space.

**Definition 3.11** (Vertical derivative). Let  $f : \mathcal{X} \rightarrow \mathbb{R}$  be a suitably regular functional,  $1 \leq i \leq d$ , and  $h^i \in \mathcal{X}$  the path defined by  $t \mapsto \frac{ht}{T} e^i$  with  $\{e^i\}_{i=1}^d$  basis vectors in  $\mathbb{R}^d$ . The vertical derivative of  $f$  at a path  $\gamma$  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  $\mathbf{w} = w_1 \cdots w_k \in \mathcal{W}_d$  of length  $k$ , define

$$\nabla^{\mathbf{w}} := \nabla^{w_1} \cdots \nabla^{w_k} f(\gamma).$$

**Proposition 3.12** (Proposition 3.8 [DT23]). *For any path  $\gamma \in \mathcal{X}$  and  $\mathbf{v}, \mathbf{w} \in \mathcal{W}_d$ , the vertical derivative satisfies*

$$\nabla^{\mathbf{v}} \mathcal{S}_{s,t}^{\mathbf{w}}(\gamma) = \begin{cases} \mathcal{S}_{s,t}^{\mathbf{u}}(\gamma), & \mathbf{w} = \mathbf{u}\mathbf{v}, \\ 0, & \text{otherwise.} \end{cases}$$

Consider now  $V \in \mathcal{C}_d$  given by

$$V = \frac{1}{2} \sum_{i=1}^d X_i^2 + W = \frac{1}{2} \sum_{i=1}^d X_i^2 + \sum_{j=1}^m g_j X_{v_j},$$

satisfying the conditions of Theorem 3.9 with associated non-commutative law  $\tau_V$ . Also associate to  $W$  the path-dependent differential operators

$$\mathcal{D}_W^k := - \sum_{j=1}^m (-i)^{|v_j|} g_j \nabla^{D_k v_j}, \quad (36)$$

where  $D_k$  is the cyclic derivative from Definition 3.6. Recall now the definitions

$$\begin{aligned} \langle \gamma_{a,b} \rangle_{\mu_V^N} &:= \sum_{\mathbf{w} \in \mathcal{W}_d} i^{|\mathbf{w}|} \mathbb{E}_{\mu_V^N} [\text{tr}(A_{\mathbf{w}})] \mathcal{S}_{a,b}^{\mathbf{w}}(\gamma), \\ \langle \gamma_{a,b} \rangle_{\mu_V^\infty} &:= \sum_{\mathbf{w} \in \mathcal{W}_d} i^{|\mathbf{w}|} \lim_{N \rightarrow \infty} \mathbb{E}_{\mu_V^N} [\text{tr}(A_{\mathbf{w}})] \mathcal{S}_{a,b}^{\mathbf{w}}(\gamma). \end{aligned}$$

**Proposition 3.13.** *Suppose that  $V$  satisfies the assumptions of Theorem 3.9 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  $(A_1^N, \dots, A_d^N) \sim \mu_V^N$ . Then

$$\mathbb{E}_{\mu_V^N} [\text{tr}(U_{s,t})] = \langle \gamma_{s,t} \rangle_{\mu_V^N} < \infty \quad \text{and} \quad \lim_{N \rightarrow \infty} \mathbb{E}_{\mu_V^N} [\text{tr}(U_{s,t})] = \langle \gamma_{s,t} \rangle_{\mu_V^\infty} < \infty. \quad (37)$$

The content of the above is simply a statement that limits and expectations may be interchanged with the sum over words.



*Proof.* The expansion of the path ordered exponential in terms of the signature provides that

$$\mathrm{tr}(U_{s,t}^N) = \sum_{\mathbf{w} \in \mathcal{W}_d} i^{|\mathbf{w}|} \mathrm{tr}(A_{\mathbf{w}}^N) \mathcal{S}_{s,t}^{\mathbf{w}}(\gamma).$$

To attain both parts of (37), 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

$$\sum_{\mathbf{w} \in \mathcal{W}_d} \mathbb{E}_{\mu_V^N} \left[ \left| \mathrm{tr}(A_{\mathbf{w}}^N) \right| \right] |\mathcal{S}_{s,t}^{\mathbf{w}}(\gamma)|$$

is uniformly bounded in  $N$ . The sub-Gaussian nature of the tail behaviour of  $\lambda_{\max}^N(\mathbf{A})$ , Lemma 3.10, yields the following bound

$$\mathbb{E}_{\mu_V^N} \left[ \left| \mathrm{tr}(A_{\mathbf{w}}^N) \right| \right] \leq \mathbb{E}_{\mu_V^N} \left[ \lambda_{\max}^N(\mathbf{A})^{|\mathbf{w}|} \right] \leq \kappa^{|\mathbf{w}|} \Gamma\left(\frac{|\mathbf{w}|}{2} + 1\right), \quad (38)$$

for some  $\kappa > 0$ , where  $\Gamma$  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

$$\sum_{\mathbf{w} \in \mathcal{W}_d} \mathbb{E}_{\mu_V^N} \left[ \left| \mathrm{tr}(A_{\mathbf{w}}^N) \right| \right] |\mathcal{S}_{s,t}^{\mathbf{w}}(\gamma)| \leq \sum_{\mathbf{w} \in \mathcal{W}_d} \kappa^{|\mathbf{w}|} \Gamma\left(\frac{|\mathbf{w}|}{2} + 1\right) |\mathcal{S}_{s,t}^{\mathbf{w}}(\gamma)| \leq \sum_{n=0}^{\infty} \frac{\tilde{\kappa}^n \Gamma\left(\frac{n}{2} + 1\right)}{\Gamma(n+1)} < \infty,$$

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 3.14.** *The functional  $\langle \gamma_{s,t} \rangle_{\mu_V^\infty}$  satisfies the path-dependent integro-differential equation*

$$\langle \gamma_{s,t} \rangle_{\mu_V^\infty} = 1 - \int_{s \leq u \leq v \leq t} \langle \gamma_{s,u} \rangle_{\mu_V^\infty} \langle \gamma_{u,v} \rangle_{\mu_V^\infty} \langle d\gamma_u, d\gamma_v \rangle - \sum_{k=1}^d \int_s^t \mathcal{D}_W^k \langle \gamma_{s,u} \rangle_{\mu_V^\infty} d\gamma_u^k. \quad (39)$$

with the boundary condition  $\langle \gamma_{s,s} \rangle_{\mu_V^\infty} = 1$  for all  $s \in [0, T]$ .

*Proof.* Recall from Theorem 3.9 that

$$\lim_{N \rightarrow \infty} \mathbb{E}_{\mu_V^N} [\mathrm{tr}(A_{\mathbf{w}})] = \tau_V(X_{\mathbf{w}}),$$

where  $\tau_V$  is the unique non-commutative law associated with  $V$  solving the corresponding Schwinger-Dyson equation. Consider now the following expansion, where we use (32) for the first equality.

$$\begin{aligned} \sum_{\mathbf{w} \in \mathcal{W}_d} i^{|\mathbf{w}|+1} \tau_V \otimes \tau_V (\partial_k X_{\mathbf{w}}) \mathcal{S}_{s,t}^{\mathbf{w}}(\gamma) &= \sum_{\mathbf{u}, \mathbf{v} \in \mathcal{W}_d} i^{|\mathbf{u}|+|\mathbf{v}|+2} \tau_V(X_{\mathbf{u}}) \tau_V(X_{\mathbf{v}}) \mathcal{S}_{s,t}^{\mathbf{u}k\mathbf{v}}(\gamma) \\ &= - \sum_{\mathbf{u}, \mathbf{v} \in \mathcal{W}_d} i^{|\mathbf{u}|+|\mathbf{v}|} \tau_V(X_{\mathbf{u}}) \tau_V(X_{\mathbf{v}}) \int_s^t \mathcal{S}_{s,u}^{\mathbf{u}}(\gamma) \mathcal{S}_{u,t}^{\mathbf{v}}(\gamma) d\gamma_u^k \\ &= - \int_s^t \left( \sum_{\mathbf{u} \in \mathcal{W}_d} i^{|\mathbf{u}|} \tau_V(X_{\mathbf{u}}) \mathcal{S}_{s,u}^{\mathbf{u}}(\gamma) \right) \left( \sum_{\mathbf{v} \in \mathcal{W}_d} i^{|\mathbf{v}|} \tau_V(X_{\mathbf{v}})(\gamma) \mathcal{S}_{u,t}^{\mathbf{v}} \right) d\gamma_u^k \\ &= - \int_s^t \langle \gamma_{s,u} \rangle_{\mu_V^\infty} \langle \gamma_{u,t} \rangle_{\mu_V^\infty} d\gamma_u^k. \end{aligned}$$

We now expand the left-hand-side using the Schwinger-Dyson equation

$$\begin{aligned}
\sum_{\mathbf{w} \in \mathcal{W}_d} i^{|\mathbf{w}|+1} \tau_V \otimes \tau_V(\partial_k X_{\mathbf{w}}) \mathcal{S}_{s,t}^{\mathbf{w}}(\gamma) &= \sum_{\mathbf{w} \in \mathcal{W}_d} i^{|\mathbf{w}|+1} \tau_V(X_{\mathbf{w}} D_k V) \mathcal{S}_{s,t}^{\mathbf{w}}(\gamma) \\
&= \sum_{\mathbf{w} \in \mathcal{W}_d} i^{|\mathbf{w}|+1} \left( \tau(X_{\mathbf{w}} X_k) + \sum_{j=1}^m g_j \tau_V(X_{\mathbf{w}} X_{D_k v_j}) \right) \mathcal{S}_{s,t}^{\mathbf{w}}(\gamma) \\
&= \sum_{\mathbf{w} \in \mathcal{W}_d} i^{|\mathbf{w}|} \tau(X_{\mathbf{w}}) \nabla^k \mathcal{S}_{s,t}^{\mathbf{w}}(\gamma) + \sum_{j=1}^m (-i)^{|v_j|-2} g_j \sum_{\mathbf{w} \in \mathcal{W}_d} \tau_V(X_{\mathbf{w}}) i^{|\mathbf{w}|} \nabla^{D_k v_j} \mathcal{S}_{s,t}^{\mathbf{w}}(\gamma) \\
&= \nabla^k \sum_{\mathbf{w} \in \mathcal{W}_d} i^{|\mathbf{w}|} \tau(X_{\mathbf{w}}) \mathcal{S}_{s,t}^{\mathbf{w}}(\gamma) + \sum_{j=1}^m (-i)^{|v_j|-2} g_j \nabla^{D_k v_j} \sum_{\mathbf{w} \in \mathcal{W}_d} \tau_V(X_{\mathbf{w}}) \mathcal{S}_{s,t}^{\mathbf{w}}(\gamma) \\
&= \nabla^k \langle \gamma_{s,t} \rangle_{\mu_V^\infty} + \sum_{j=1}^m (-i)^{|v_j|-2} g_j \nabla^{D_k v_j} \langle \gamma_{s,t} \rangle_{\mu_V^\infty} \\
&= (\nabla^k + \mathcal{D}_W^k) \langle \gamma_{s,t} \rangle_{\mu_V^\infty}.
\end{aligned}$$

Equating these identities and integrating both sides with respect to  $d\gamma^k$  and then summing over  $k$  results in (39):

$$\langle \gamma_{s,t} \rangle_{\mu_V^\infty} = 1 - \int_{s \leq u \leq v \leq t} \langle \gamma_{s,u} \rangle_{\mu_V^\infty} \langle \gamma_{u,v} \rangle_{\mu_V^\infty} \langle d\gamma_u, d\gamma_v \rangle - \sum_{k=1}^d \int_s^t \mathcal{D}_W^k \langle \gamma_{s,u} \rangle_{\mu_V^\infty} d\gamma_u^k. \quad (40)$$

□

**Example 3.15** (Straight Lines). Fix  $T = 1$  and consider the case of a straight line  $\gamma = tv$ , for  $v \in \mathbb{R}^d$ . We note that  $d\gamma_t^k = v^k dt$  and that  $\langle \gamma_{s,t} \rangle_{\mu_V^\infty}$  is now a function of only  $(t - s)$ , so that (39) may be written as

$$\mathcal{D}_V \langle \gamma \rangle_t = - \int_0^t \langle \gamma \rangle_u \langle \gamma \rangle_{t-u} v^k dt,$$

for  $t \in [0, T]$ . We thus recover the classical result (18).

**Example 3.16** (GUE). Consider the case of the quadratic potential  $V = \frac{1}{2} \sum_{i=1}^d X_i^2$ , for which  $\mu_V^N$  corresponds to the GUE. Then each  $\mathcal{D}_V^k$  is given by  $\Delta_k$ . We may then integrate both sides of the equation

$$\mathcal{D}_V^k \langle \gamma_{s,t} \rangle_{\mu_V^\infty} = - \int_s^t \langle \gamma_{s,u} \rangle_{\mu_V^\infty} \langle \gamma_{u,t} \rangle_{\mu_V^\infty} d\gamma_u^k,$$

with respect to  $\gamma^k$  to obtain

$$\sum_{\mathbf{w} \in \mathcal{W}_d} i^{|\mathbf{w}|+1} \tau_V(X_{\mathbf{w}k}) \mathcal{S}_{s,t}^{\mathbf{w}k}(\gamma) = \int_s^t \mathcal{D}_V^k \langle \gamma_{s,r} \rangle_{\mu_V^\infty} d\gamma_r^k = - \int_{s < u < r < t} \langle \gamma_{s,u} \rangle_{\mu_V^\infty} \langle \gamma_{u,r} \rangle_{\mu_V^\infty} d\gamma_u^k d\gamma_r^k$$

Summing over  $k$ , we see that

$$\begin{aligned}
- \int_{s < u < r < t} \langle \gamma_{s,u} \rangle_{\mu_V^\infty} \langle \gamma_{u,r} \rangle_{\mu_V^\infty} \langle d\gamma_u, d\gamma_r \rangle &= \sum_{k=1}^d \int_{s < u < r < t} \langle \gamma_{s,u} \rangle_{\mu_V^\infty} \langle \gamma_{u,r} \rangle_{\mu_V^\infty} d\gamma_u^k d\gamma_r^k \\
&= - \sum_{k=1}^d \sum_{\mathbf{w} \in \mathcal{W}_d} i^{|\mathbf{w}|+1} \tau_V(X_{\mathbf{w}k}) \mathcal{S}_{s,t}^{\mathbf{w}k}(\gamma) \\
&= -1 + \sum_{\mathbf{w} \in \mathcal{W}_d} i^{|\mathbf{w}|} \tau_V(X_{\mathbf{w}}) \mathcal{S}_{s,t}^{\mathbf{w}}(\gamma) \\
&= \langle \gamma_{s,t} \rangle_{\mu_V^\infty} - 1,
\end{aligned}$$

which is the equation obtained in [CT24a].

To conclude this subsection, we consider the uniqueness of solutions to (39). 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 3.17.** *Consider the set of infinite linear functionals,  $\ell : \mathcal{X} \rightarrow \mathbb{C}$ , 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  $|a_{\mathbf{w}}| \leq C^{|\mathbf{w}|}$  for some  $C > 0$ . Suppose that

$$\ell(\gamma_{s,t}) = 1 - \int_{s \leq u \leq v \leq t} \ell(\gamma_{s,u}) \ell(\gamma_{u,v}) \langle d\gamma_u, d\gamma_v \rangle - \sum_{k=1}^d \int_s^t \mathcal{D}_W^k \ell(\gamma_{s,u}) d\gamma_u^k, \quad (41)$$

holds for all  $\gamma \in \mathcal{X}$ , then  $a_{\mathbf{w}} = \tau_V(X_{\mathbf{w}})$  for all  $\mathbf{w} \in \mathcal{W}_d$ .

*Proof.* Expanding (41) gives

$$\begin{aligned} \sum_{\mathbf{w} \in \mathcal{W}_d} i^{|\mathbf{w}|} a_{\mathbf{w}} \mathcal{S}_{s,t}^{\mathbf{w}}(\gamma) &= 1 + \sum_{k=1}^d \sum_{\mathbf{u}, \mathbf{v} \in \mathcal{W}_d} i^{|\mathbf{u}|+|\mathbf{v}|+2} a_{\mathbf{u}} a_{\mathbf{v}} \mathcal{S}_{s,t}^{\mathbf{u} \mathbf{v} k}(\gamma) \\ &\quad + \sum_{k=1}^d \sum_{\mathbf{u} \in \mathcal{W}_d} \sum_{j=1}^m i^{|\mathbf{u}|+|\mathbf{v}_j|-2-|\mathbf{v}_j|} g_j a_{\mathbf{u}} D_k v_j \mathcal{S}_{s,t}^{\mathbf{u} k}(\gamma) \end{aligned}$$

Noting that the left-hand-side may be expanded as

$$\sum_{\mathbf{w} \in \mathcal{W}_d} i^{|\mathbf{w}|} a_{\mathbf{w}} \mathcal{S}_{s,t}^{\mathbf{w}}(\gamma) = 1 + \sum_{k=1}^d \sum_{\mathbf{w} \in \mathcal{W}_d} i^{|\mathbf{w}|+1} a_{\mathbf{w}k} \mathcal{S}_{s,t}^{\mathbf{w}k}(\gamma),$$

means that we may equate the coefficients of  $\mathcal{S}^{\mathbf{w}k}$  to obtain the system of equations

$$a_{\mathbf{w}k} = \sum_{\mathbf{w}=\mathbf{u}k\mathbf{v}} a_{\mathbf{u}} a_{\mathbf{v}} - \sum_{j=1}^m g_j a_{\mathbf{w}} D_k v_j,$$

which after rearranging yields

$$\sum_{\mathbf{w}=\mathbf{u}k\mathbf{v}} a_{\mathbf{u}} a_{\mathbf{v}} = a_{\mathbf{w}k} + \sum_{j=1}^m g_j a_{\mathbf{w}} D_k v_j,$$

which implies that the extension of  $\ell$  to a linear functional  $\tau_a : \mathcal{C}_d \rightarrow \mathbb{C}$  satisfies the Schwinger-Dyson equation  $\tau_a \otimes \tau_a(\partial_k X_{\mathbf{w}}) = \tau_a(X_{\mathbf{w}} D_k V)$ . By the uniqueness of solutions to the Schwinger-Dyson equation, Theorem 3.9, we must have  $\tau_a = \tau_V$ .  $\square$

## 4 Quantum algorithm

In this section we introduce quantum analogues of the classical path signature and signature kernel. We will study three main objects. Firstly, the *quantum path development*—this is a random quantum circuit that we denote by  $U_{\gamma}^Q(\alpha(m), n, K)$ . The circuit acts on  $n$  qubits, dependent on  $m$  random Pauli strings labelled by  $\alpha$ , an approximation quality  $K$ , and a path  $\gamma$ . We will show (Theorem 4.8) that the expected value of this circuit converges to the Gaussian unitary path developments studied in the previous section. Secondly, we define the *quantum path signature*  $\mathcal{S}_{\epsilon}^Q$  to be the quantum feature map associated to the quantum circuit. Finally, in Section 4.3, we discuss a *quantum signature kernel*, denoted  $k^Q(\sigma, \tau)$ , which is realised as the output of a quantum algorithm  $\mathcal{A}$  built using the quantum path development  $U_{\gamma}^Q(\alpha(m), n, K)$ . One of the main results of this section is Theorem 4.11 demonstrating the efficiency of this algorithm. Along the way, we give theoretical guarantees for an analogous classical algorithm (Theorem 4.12) to compute the randomised path development.

## 4.1 Preliminaries

In the following, we review basic definitions and concepts from quantum computing. Whilst this material will likely be familiar to readers with a background in quantum information theory, we include it to ensure the work remains accessible to researchers from the path signature and machine learning communities, in line with our aim of connecting ideas across these fields. For a more comprehensive introduction, we refer the reader to the standard text [NCro].

**Quantum states.** We consider the Hilbert space of  $n$  qubits

$$\mathcal{H} = \underbrace{\mathbb{C}^2 \otimes \cdots \otimes \mathbb{C}^2}_{n\text{-times}}, \quad (42)$$

which has dimension  $N = 2^n$ . We write  $|\psi\rangle$  for vectors, henceforth referred to as (pure) quantum states, in  $\mathcal{H}$ . We may write the basis of tensor product states of  $\mathcal{H}$  as  $|x\rangle = |x_1\rangle \otimes \cdots \otimes |x_n\rangle$  where each  $x_i \in \{0, 1\}$ . Each state may then be identified with a binary string  $x \in \{0, 1\}^n$  which defines an isomorphism with  $\mathbb{C}^{N=2^n}$ . We define  $|\mathbf{0}\rangle := |0\rangle \otimes \cdots \otimes |0\rangle \in \mathcal{H}$ .

More general quantum states, referred to as *mixed states*, are represented by a *density operator*  $\rho$  acting on  $\mathcal{H}$ , which is a positive semi-definite Hermitian operator satisfying  $\text{Tr}(\rho) = 1$ . If, in addition,  $\text{Tr}(\rho^2) = 1$ , then  $\rho$  corresponds to a pure state and can be written in the form  $\rho = |\psi\rangle\langle\psi|$ , considered as an element of  $\mathcal{H} \otimes \mathcal{H}^*$ , for some state  $|\psi\rangle \in \mathcal{H}$ . In the general case,  $\rho$  describes a classical probabilistic mixture of pure states and can be expressed as a convex combination:

$$\rho = \sum_{i=1}^k p_i |\psi_i\rangle\langle\psi_i|, \quad (43)$$

where  $p_i \geq 0$ ,  $\sum_i p_i = 1$ , and each  $|\psi_i\rangle \in \mathcal{H}$  is a pure state. We denote this (Hilbert) space of general quantum states by  $S(\mathcal{H})$ . A quantum circuit is a unitary matrix  $U$  acting on  $\mathcal{H} = (\mathbb{C}^2)^{\otimes n}$  built from a finite sequence of elementary unitaries chosen from a *universal gate set* [Kit97].

**Measurements.** In the present context, the measurement of a state  $\rho$  will be described by a set of orthogonal projectors  $\{\Pi_i\}_{i=1}^k$  on  $\mathcal{H}$  satisfying  $\sum_{i=1}^k \Pi_i = I_N$  and  $\Pi_i \Pi_j = \delta_{ij} \Pi_i$ . The measurement describes the possible outcomes of a discrete random variable  $\Pi$  taking values in  $\{1, \dots, k\}$ , with associated probabilities

$$\mathbb{P}(\Pi = i) = \text{tr}(\rho \Pi_i). \quad (44)$$

This is known as Born's rule.

**Quantum feature map.** Let  $\mathcal{X}$  be a set of data; a quantum feature map [SK19] is an embedding

$$\rho : \mathcal{X} \rightarrow S(\mathcal{H}), \quad (45)$$

where the feature space is the space of quantum states on  $\mathcal{H}$ , i.e. we map  $x \in \mathcal{X}$  to a density operator  $\rho(x)$ .

## 4.2 Quantum circuit

We now turn to a quantum circuit implementation of the unitary path development. The Hilbert space of qubits  $\mathcal{H}$  has dimension  $N = 2^n$ . In the geometrical setup outlined in Section 2.2, this Hilbert space now plays the role of the associated  $G = U_N$  bundle in the fundamental representation and the quantum circuit will play the role of the parallel transport operator between representation vectors.

We assume throughout this section that paths  $\gamma$  are continuous piecewise linear  $\gamma : [0, T] \rightarrow \mathbb{R}^d$ . It is natural to consider this class of paths as any smooth path can be approximated in an appropriate topology by sequences of piecewise linear paths. The time interval  $[0, T]$  is divided into intervals  $0 = t_0 < t_1 < \dots < t_L = T$  and we write

$$\gamma = \gamma_1 \star \cdots \star \gamma_L, \quad (46)$$

where each  $\gamma_l$  is a linear path on  $[t_{l-1}, t_l]$ . The data specifying the path is then  $L$  vector increments, denoted  $\Delta_l^\nu := \gamma^\nu(t_l) - \gamma^\nu(t_{l-1})$  for  $l = 1, \dots, L$  and  $\nu = 1, \dots, d$ . For piecewise linear paths, the length is equal to the total one-variation which we denote by  $\Delta_\gamma = \sum_{l=1}^L \|\Delta_l\|$ .

As we will see more precisely later, the computational difficulty of computing the path development is controlled by the  $U_N$  dimension  $N = 2^n$ , the total one-variation  $\Delta_\gamma$  (*i.e.* the length of the path) and the quality of the approximation  $\epsilon$  arising due to the discretisation as a quantum circuit.

**Definition 4.1** (Pauli strings). A set of Hermitian operators on  $\mathbb{C}^2$  given by

$$\sigma_I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \sigma_X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (47)$$

A (real) linear basis for  $\mathbf{u}_N$  is then given by tensor products of the Pauli matrices

$$\sigma_{\mathbf{w}} = \sigma_{w_1} \otimes \dots \otimes \sigma_{w_n}, \quad (48)$$

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  $\mathcal{P}$ .

We now consider the  $N \times N$  unitaries  $U_\gamma$  discussed in previous sections as operations on qubits. We will show that these may be decomposed into quantum gates. Let us first consider the  $N$ -dimensional path development for a fixed set of  $d$  Hermitian matrices  $A_1, \dots, A_d$ , that is the solution to<sup>7</sup>

$$dU_\gamma(t) = iU_\gamma(t)A_\nu \dot{\gamma}_t^\nu dt, \quad U_\gamma(0) = I, \quad (49)$$

at  $t = T$ . We denote the solution by  $U_\gamma(A, N)$ . We now turn to how to implement this unitary as a quantum circuit acting on the Hilbert space  $\mathcal{H} = (\mathbb{C}^2)^{\otimes n}$ .

**Trotterisation.** Let us discuss the solution to the path development equation (49) for piecewise linear paths (46) and for a fixed set of  $N \times N$  matrices  $\{A_\nu\}_{\nu=1}^d$ . From the fact  $U_{\gamma_1 \star \gamma_2} = U_{\gamma_1} U_{\gamma_2}$  we have that

$$U_\gamma(A, N) = U_{\gamma_1}(A, N) U_{\gamma_2}(A, N) \dots U_{\gamma_L}(A, N), \quad (50)$$

where  $U_{\gamma_l}(A, N)$  denotes the unitary time evolution operator/path development for the straight line segment on  $[t_l, t_{l+1}]$ . Each  $U_{\gamma_l}(A, N)$  is then given by

$$U_{\gamma_l}(A; N) = \exp \left( i \sum_{\nu=1}^d \Delta_l^\nu A_\nu \right), \quad (51)$$

where  $\Delta_l^\nu = \gamma^\nu(t_l) - \gamma^\nu(t_{l-1})$  are the (vector) increments of the path on each time step. Recall (Definition 4.1) that each Hermitian matrix  $A_\nu$  may be expressed in the Pauli string basis. We write the expansion as

$$A_\nu = \sum_{\mathbf{w} \in \mathcal{P}} \alpha_\nu^{\mathbf{w}} \sigma_{\mathbf{w}}, \quad \nu = 1, \dots, d, \quad (52)$$

with  $\{\alpha_\nu^{\mathbf{w}}\}_{\mathbf{w}}$  a set of  $4^n$  real coefficients for each matrix and often write the development as  $U_\gamma(\alpha, N)$  emphasising the dependence on these coefficients. We now consider the Trotterisation of this unitary to define a quantum circuit.

<sup>7</sup>In this equation and in the following we use the summation convention that repeated upper and lower indices are summed. E.g.  $X_\nu X^\nu := \sum_{\nu=1}^d X_\nu X^\nu$ .

**Definition 4.2.** The first-order Trotterisation with  $K$  subdivisions of each straight line interval  $\gamma_l$ , written  $U_{\gamma_l}(\alpha, n, K)$ , is defined by

$$U_{\gamma_i}(\alpha; n, K) := \left[ \prod_{\nu=1}^d \prod_{\mathbf{w} \in P} \exp\left(i \frac{\Delta_i^\nu \alpha_\nu^\mathbf{w}}{K} \sigma_\mathbf{w}\right) \right]^K = \left[ \prod_{\nu=1}^d \prod_{\mathbf{w} \in P} P_\mathbf{w}\left(\frac{\Delta_i^\nu \alpha_\nu^\mathbf{w}}{K}\right) \right]^K, \quad (53)$$

where we write the shorthand  $\alpha$  for the set of coefficients  $\{\alpha_\nu^\mathbf{w}\}$  that specify  $A$ . In the second equality, we have re-written the exponentials as Pauli rotations  $P_\mathbf{w}(\theta) := \exp(i\theta\sigma_\mathbf{w})$  with small angles. Finally, we define the Trotterised path development of  $\gamma$  by

$$U_\gamma(\alpha; n, K) := U_{\gamma_L}(\alpha; n, K) U_{\gamma_{L-1}}(\alpha; n, K) \cdots U_{\gamma_1}(\alpha; n, K). \quad (54)$$

We have thus constructed a quantum circuit from repeated application of multi-qubit Pauli rotations with small angles that approximates the path development with a fixed set of Hermitian matrices determined by coefficients  $\{\alpha_\nu^\mathbf{w}\}$ . Intuitively, one may view the Trotterisation as dividing each interval  $[t_i, t_{i+1}]$  further into  $K$  subdivisions and one recovers the full development in the limit  $K \rightarrow \infty$ . The following lemma quantifies the quality of this approximation.

**Lemma 4.3.** *We consider the Trotterisation into  $K$  subdivisions of the path development  $U_\gamma$  for a piecewise linear path  $\gamma$  defined by a fixed gauge field specified by coefficients  $\{\alpha_\nu^\mathbf{w} : |\alpha_\nu^\mathbf{w}| \leq 1/m\}$  in the Pauli basis. Then*

$$\|U_\gamma(\alpha, n, K) - U_\gamma(\alpha, n)\| \leq \frac{|P|_\alpha}{K} \Delta_\gamma^2 + \mathcal{O}\left(\frac{1}{K^2}\right), \quad (55)$$

where  $|P|_\alpha$  denotes the maximum number (over  $\nu$ ) of non-zero coefficients in the Pauli string expansion and  $\Delta_\gamma$  denotes the total one-variation of the path  $\gamma$ .

**Remark 4.4.** *When  $\gamma(t) = t$  is a one-dimensional linear path segment on  $[0, t]$ , the lemma recovers the familiar Trotter error bound from Hamiltonian simulation theory that is quadratic  $\mathcal{O}(t^2)$  in the time interval (length of the path) and  $\mathcal{O}(1/K)$  in the Trotter subdivision.*

*Proof.* First, recall the standard Trotterisation error bound [Suz76]. Suppose a Hamiltonian is given as a sum of bounded operators  $H = \sum_{j=1}^N H_j$  then we have, in operator norm,

$$\left\| \left( \prod_{j=1}^N \exp\left\{\frac{iH_j}{K}\right\} \right)^K - e^{iH} \right\| \leq \frac{1}{2K} \sum_{i < j} \|[H_i, H_j]\| + \mathcal{O}\left(\frac{1}{K^2}\right). \quad (56)$$

Consider now the Hamiltonian  $H = \sum_{\nu=1}^d \sum_{\mathbf{w} \in P} H_{\nu, \mathbf{w}}$  with  $H_{\nu, \mathbf{w}} = \Delta_i^\nu \alpha_\nu^\mathbf{w} \sigma_\mathbf{w}$ . From the Lie algebra structure of the Pauli matrices, then  $\|[\alpha_\nu^\mathbf{w} \sigma_\mathbf{w}, \alpha_{\nu'}^{\mathbf{w}'} \sigma_{\mathbf{w}'}]\| \leq 2$ , and (56) yields

$$\|U_{\gamma_i}(\alpha; n, K) - U_{\gamma_i}(\alpha; n)\| \leq \frac{|P|_{\alpha_\nu}}{K} \sum_{\nu=1}^d |\Delta_i^\nu|^2 + \mathcal{O}\left(\frac{1}{K^2}\right). \quad (57)$$

where  $|P|_{\alpha_\nu}$  denotes the number of  $\mathbf{w}$  for which  $\alpha_\nu^\mathbf{w}$  is non-zero. Now we would like to consider a bound on

$$\|U_\gamma(\alpha, n, K) - U_\gamma(\alpha, n)\| = \left\| \prod_{i=1}^L U_{\gamma_i}(\alpha, n, K) - \prod_{i=1}^L U_{\gamma_i}(\alpha, n) \right\|. \quad (58)$$

Telescoping the product and using the fact that the operators are unitary we simply obtain the sum of the individual bounds between the product components

$$\begin{aligned} \|U_\gamma(\alpha, n, K) - U_\gamma(\alpha, n)\| &\leq \frac{|P|_\alpha}{K} \sum_{i=1}^L \sum_{\nu=1}^d |\Delta_i^\nu|^2 + \mathcal{O}\left(\frac{1}{K^2}\right) \\ &= \frac{|P|_\alpha}{K} \Delta_{\gamma,2}^2 + \mathcal{O}\left(\frac{1}{K^2}\right), \\ &\leq \frac{|P|_\alpha}{K} \Delta_\gamma^2 + \mathcal{O}\left(\frac{1}{K^2}\right), \end{aligned} \quad (59)$$

where in the second line  $\Delta_{\gamma,2}^2$  denotes the square of the total two-variation of the path and in the third line we have used that for piecewise linear paths the two-variation is bounded by the one-variation. In the above  $|P|_\alpha := \max_\nu |P|_{\alpha_\nu}$ .  $\square$

**Remark 4.5.** The circuit  $U_\gamma(\alpha, n, K)$  may be further decomposed into a fundamental gate set, for example using the methods described in [Sri+23].

**Randomisation and sparsification.** We now discuss how to randomise over the Hermitian matrices  $A_\nu$  (or equivalently the coefficients  $\alpha$  in the Pauli string expansion of  $A$ ). We sample  $A_\nu$  from the  $N = 2^n$  dimensional Gaussian unitary ensemble (GUE) by sampling coefficients of the Pauli strings in the circuit (54). In recent work [Che+24], the authors define a sparse ensemble of random Pauli strings. They also show that ground states of the associated Hamiltonian are efficient to prepare. In the following, we use the same matrix ensemble in a different context that allows us to consider a sparsification of Pauli strings in the path development circuit (54). We first recall the definition of the random Pauli ensemble and then argue that the first moment of this random Pauli ensemble converges to the semi-circular law with a sufficiently fast rate.

**Definition 4.6.** (Random Pauli ensemble) Let  $N = 2^n$ . 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  $r_i \stackrel{\text{i.i.d.}}{\sim} \text{Rademacher}(1/2)$  and  $\sigma^i = \sigma(1)^{i_1} \otimes \cdots \otimes \sigma(n)^{i_n}$ , where  $\sigma(j)^{i_j} \stackrel{\text{i.i.d.}}{\sim} \mathcal{P}$ , where  $\sigma \sim \mathcal{P}$  denotes a random variable uniformly distributed on the set of Pauli matrices  $\mathcal{P}$ .

The distribution of  $\nu = 1, \dots, d$  matrices from the random Pauli ensemble of Definition 4.6, may be equivalently described by a random variable  $\alpha$  for the coefficients in the expansion (52) where for each  $\nu = 1, \dots, d$ , the coefficient  $\alpha_\nu^\mathbf{w}$  is non-zero for only  $m$  uniformly chosen  $\mathbf{w}$  where it takes the two values  $\{-1/\sqrt{m}, 1/\sqrt{m}\}$  with equal probability. We henceforth denote this random variable by  $\alpha(m) = \alpha(m)_\nu^\mathbf{w}$ . Accordingly, we write the path development associated to  $\alpha(m)$  as

$$dU_{\gamma_{s,t}}^{\text{SP}}(\alpha(m), n) = iU_{\gamma_{s,t}}^{\text{SP}}(\alpha(m), n) A_\nu^m \dot{\gamma}_t^\nu dt, \quad U_{\gamma_{s,t}}^{\text{SP}}(\alpha(m), n) = I_{2^n}, \quad (60)$$

where the matrices  $A_\nu^m$  have expansions  $A_\nu^m = \sum_{\mathbf{w} \in P} \alpha(m)_\nu^\mathbf{w} \sigma_\mathbf{w}$  with  $m$  non-zero terms. We write the solution as  $U_{\gamma_{s,t}}^{\text{SP}}(\alpha(m), n)$ .

Next we present the main result of this subsection, which states that the path development driven by random Pauli ensemble matrices approximates well the unitary path development. The proof is given in Appendix B and relies crucially on the following proposition about the moments of the Pauli ensemble. For completeness, and because it may be of independent interest, we restate this proposition below.

**Proposition 4.7.** For each  $n$ , let  $A_1^n, \dots, A_d^n$  be i.i.d. matrices from the random Pauli string ensemble of dimension  $N = 2^n$  with  $m_n$  strings each. Then

$$\left| \mathbb{E}[\text{tr}(A_\mathbf{w}^n)] - \tau(X_\mathbf{w}) \right| \leq (4e)^p p! \left( m_n^{-1} + 2^{-2n} \right),$$

where  $\tau$  is the law of  $d$ -free semicircular random variables and  $\mathbf{w} \in \mathcal{W}_d$  has length  $2p$  for some  $p \in \mathbb{N}$ . In the above  $A_\mathbf{w}$  denotes the product  $A_{w_1} \dots A_{w_{2p}}$ .

**Theorem 4.8.** Fix  $n > 0$  and let  $\alpha(m)$  be the random variable defined above. Let

$$\langle \gamma_{s,t} \rangle_{\mu_V^\infty} = \sum_{\mathbf{w} \in \mathcal{W}_d} i^{|\mathbf{w}|} \tau(X_\mathbf{w}) \mathcal{S}_{s,t}^\mathbf{w}(\gamma),$$

be the GUE randomised path development, then there exists a constant  $C > 0$  depending only on the length of  $\gamma$  and the dimension  $d$  of the path target space, for which

$$\mathbb{E} \left[ \left( \text{tr}(U_{\gamma_{s,t}}^{\text{SP}}(\alpha(m), n)) - \langle \gamma_{s,t} \rangle_{\mu_V^\infty} \right)^2 \right] < C(m^{-1} + 4^{-n}). \quad (61)$$



We now combine the previous arguments to define a quantum path development. We consider a piecewise linear path  $\gamma : [a, b] \rightarrow \mathbb{R}^d$  given as  $L$  specified increment vectors  $\{\Delta_l^1, \Delta_l^2, \dots, \Delta_l^d\}_{l=1}^L$ . We define the *quantum path development* which combines (1) the sparse Pauli approximation of the large  $N$  GUE development of [Theorem 4.8](#) and (2) the Trotterisation approximation of the sparse Pauli development of [Lemma 4.3](#).

**Definition 4.9** (Quantum path development). The *quantum path development* is the random quantum circuit  $U_\gamma^Q(\alpha(m), n, K)$  defined in terms of Pauli rotations, following [\(54\)](#):

$$U_\gamma^Q(\alpha(m), n, K) = \prod_{l=1}^L \left[ \prod_{\nu=1}^d \prod_{i=1}^m P_{\mathbf{w}_{i,\nu}}(\Delta_l^\nu \alpha_\nu^{\mathbf{w}_i} / K) \right]^K, \quad (62)$$

where  $\mathbf{w}_{i,\nu}$  denotes the  $i = 1, \dots, m$  non-zero coefficients of  $\alpha_\nu^{\mathbf{w}}$  for each  $\nu$ . The circuit depends on parameters  $n$  (number of qubits), Trotter subdivisions  $K$  and  $m$  random Pauli strings. The circuit consists of  $L \times K \times d \times m$  Pauli rotations.

In the following section we demonstrate, as part of the proof of [Theorem 4.11](#), that the quantum path development well-approximates the unitary development of a path  $\gamma$ .

**Quantum path signatures.** We define the quantum path signature as a quantum feature map, in the sense discussed at the beginning of this section [\(45\)](#), that allows one to embed a path  $\gamma$  into the space of states on qubits. In the present case,  $\mathcal{X}$  is a space of paths and we would like to embed the path  $\gamma$  into the Hilbert space of a large number of qubits  $n$ , that is we would like to consider a mean embedding of the form

$$\int U_\gamma(A) |\mathbf{0}\rangle \langle \mathbf{0}| U_\gamma^\dagger(A) d\mu_V^N(A), \quad (63)$$

where  $N$  is large and  $\mu_V^N$  is the Gaussian measure. To make this well-defined, based on the previous discussion, we see that we may work with an approximate embedding into a truncated Hilbert space of  $n$  qubits, with Trotterised time evolution into  $K$  Trotter subdivisions and a sparse Pauli ensemble controlled by an integer  $m$ .

**Definition 4.10** (Quantum path signature). We now define the quantum path signature as a quantum feature map  $\mathcal{S}^Q(\gamma) : \mathcal{X} \rightarrow S(\mathcal{H}_n)$  embedding paths  $\gamma$  into a quantum state (density matrix) by the circuit of [Definition 4.9](#), namely

$$\mathcal{S}^Q(\gamma) := \mathbb{E}_{\alpha(m)} \left[ U_\gamma^Q(\alpha(m), n, K) |\mathbf{0}\rangle \langle \mathbf{0}| U_\gamma^Q(\alpha(m), n, K)^\dagger \right], \quad (64)$$

In applications, we propose that the parameters  $n$ ,  $K$  and  $m$  should be chosen to scale as in [Theorem 4.11](#).

### 4.3 Quantum signature kernel

We now introduce a quantum algorithm, making use of the quantum path development circuit, to compute the trace of the GUE randomised path development<sup>8</sup>

$$\langle \gamma \rangle := \lim_{N \rightarrow \infty} \mathbb{E} \left[ \text{tr}(U_\gamma(A, N)) \right]. \quad (65)$$

In the subsection that follows, we discuss how the algorithm may be further used to approximate the GUE kernel on path space  $k^{\text{GUE}} : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$  defined by

$$k^{\text{GUE}}(\sigma, \tau) := \lim_{N \rightarrow \infty} \mathbb{E} \left[ \text{tr}(U_\sigma(A, N) U_\tau(A, N)^\dagger) \right]. \quad (66)$$

as discussed in [Section 2.4](#). We define the *quantum signature kernel* as the output of this quantum algorithm.

<sup>8</sup>We drop the  $\mu_\infty$  subscript compared with [Definition \(24\)](#) since we only consider here the large  $N$  GUE path development.

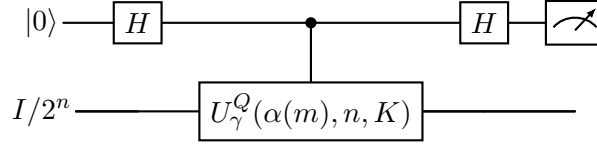


Figure 3: One clean qubit (DCQI) circuit with the quantum path development.

**Problem statement.** We consider a continuous piecewise linear path  $\gamma$  on a partitioned time interval  $0 = t_0 < t_1 < \dots < t_L = T$  mapping into  $\mathbb{R}^d$ . The input data is thus the set of increments  $\{\Delta_l^\nu\}$  with  $l = 1, \dots, L$  and  $\nu = 1, \dots, d$ . The computational problem is then to output a quantity  $Q$  that is  $\epsilon$ -additively close to the path development  $\langle \gamma \rangle$  with small failure probability  $\delta$ .

**Computation.** We now define a quantum algorithm  $\mathcal{A}(\{\Delta_l^\nu\}, \epsilon, \delta)$  in the one-clean-qubit model of quantum computation [KL98] with random output  $Q$ . The algorithm takes as input the increments  $\{\Delta_l^\nu\}$  together with an additive error  $\epsilon$  and a probability of failure  $\delta$ . We first take a sample  $\tilde{\alpha}(m)$  of the random variable  $\alpha(m)$  described below Definition 4.6. We then embed the random quantum circuit  $U_\gamma^Q(\tilde{\alpha}(m), n, K)$  from Definition 4.9 as the controlled unitary in the one-clean-qubit circuit illustrated in Fig. 3. The initial state of the system is

$$\rho_0 = |0\rangle\langle 0| \otimes \frac{I}{2^n}, \quad (67)$$

and, after applying a Hadamard gate to the first qubit, a sample of the quantum path development circuit and an additional Hadamard gate, the state of the system is

$$\rho_1 = \frac{1}{2^{n+1}} \left[ (|0\rangle\langle 0| + |1\rangle\langle 1|) \otimes I + |0\rangle\langle 1| \otimes U_\gamma^Q(\tilde{\alpha}(m), n, K)^\dagger + |1\rangle\langle 0| \otimes U_\gamma^Q(\tilde{\alpha}(m), n, K) \right]. \quad (68)$$

Measuring the first qubit in the computational basis then yields the probability of observing "1" by Born's rule (44),<sup>9</sup>

$$\text{tr}(|1\rangle\langle 1| \rho_1) = \frac{1}{2} \left( 1 - \text{tr}(U_\gamma^Q(\tilde{\alpha}(m), n, K)) \right). \quad (69)$$

Finally, we define the output of the algorithm to be the estimator of  $\text{tr}(U_\gamma^Q(\tilde{\alpha}(m), n, K))$ . Namely

$$Q_\Delta(M, m, n, K) := 1 - \frac{2T_\Delta(M, m, n, K)}{M}, \quad (70)$$

where  $T_\Delta(M, m, n, K)$  denotes the number of ones observed when initialising, running and measuring the circuit  $M$  times. The output depends on the number of qubits  $n$ ; the Trotter approximation  $K$ ; the number of Pauli strings  $m$ ; and the number of times  $M$  to run the circuit of Figure 3. Theorem 4.11 explains how to scale these parameters as functions of  $\epsilon$  and  $\delta$ . The algorithm may be summarised as follows:<sup>10</sup>

<sup>9</sup>Recall that the unitary path development is real valued.

<sup>10</sup>This algorithm is named **QSigKer** because of its direct application to computing a kernel in the following subsection.

---

**Algorithm 1** QSigKer – Quantum Signature Kernel Estimator  $\mathcal{A}(\{\Delta_l^\nu\}, \epsilon, \delta)$ 


---

**Input:** Data of increments  $\{\Delta_l^\nu\}$  and  $\epsilon, \delta > 0$ . Parameters: integers  $m, n, K, M > 0$  for Pauli string number, qubit number, Trotter subdivisions and number of circuit runs respectively. These are selected according to [Theorem 4.11](#).

**Output:** Approximation  $Q_\Delta(M, m, n, K)$  of  $\langle \gamma \rangle$ .

- 1: Set counter  $T \leftarrow 0$
  - 2: **for**  $i = 1$  to  $M$  **do**
  - 3:   Sample  $\tilde{\alpha}(m)$  from random variable  $\alpha(m)$
  - 4:   Initialise system in state  $\rho_0 = |0\rangle\langle 0| \otimes \frac{I}{2^n}$
  - 5:   Apply  $H$  to first qubit. Apply unitary  $U_\gamma^Q(\tilde{\alpha}(m), n, K)$  of definition 4.9. Apply  $H$  to first qubit.
  - 6:   Measure first qubit in computational basis obtaining outcome  $x \in \{0, 1\}$
  - 7:   **if**  $x = 1$  **then**
  - 8:      $T \leftarrow T + 1$
  - 9: **return**  $Q_\Delta(M, m, n, K) \leftarrow 1 - \frac{2T}{M}$
- 

**Theorem 4.11.** *The positive integer parameters  $m, n, K$  and  $M$  (Pauli string number, qubit number, Trotter subdivisions, number of circuit runs) may be chosen such that the output of  $\mathcal{A}(\{\Delta_l^\nu\}, \epsilon, \delta)$  estimates the unitary path development  $\langle \gamma \rangle$  to within additive error  $\epsilon$  with high probability  $1 - \delta$ , namely such that*

$$\mathbb{P}(|Q_\Delta(M, m, n, K) - \langle \gamma \rangle| \geq \epsilon) < \delta. \quad (71)$$

*The circuit requires  $\log(1/\epsilon, 1/\delta)$  qubits and  $\text{poly}(1/\epsilon, 1/\delta)$  Pauli rotations.*

*Proof.* See Appendix C.1 □

**Classical algorithm.** Let us now define an analogous classical algorithm and perform the complexity analysis for comparison. The input data remains unchanged—the set of path increments  $\{\Delta_l^1, \Delta_l^2, \dots, \Delta_l^d\}_{l=1}^L$ . We define a probabilistic classical algorithm,  $\mathcal{A}^c(\{\Delta_l^\nu\}, \epsilon, \delta)$ , which, given the input data, an additive error  $\epsilon$  and a probability of failure  $\delta$ , produces a random output that is  $\epsilon$  close to the path development with high probability  $1 - \delta$ . We follow a standard Monte Carlo approach, defined precisely as:<sup>11</sup>

---

**Algorithm 2** Classical Sampling  $\mathcal{A}^c$ 


---

**Input:** Data: path increments  $\{\Delta_l^1, \dots, \Delta_l^d\}_{l=1}^L$  and  $\epsilon, \delta > 0$ . Parameters: integer  $N > 0, M > 0$  and  $K > 0$  are chosen dependent on  $\epsilon$  and  $\delta$  according to [Theorem 4.12](#).

**for**  $m = 1$  to  $M$  **do**

    Draw  $d$  independent  $N \times N$  GUE matrices  $\{A_\nu\}_{\nu=1}^d$

**for**  $l = 1$  to  $L$  **do**

        Compute  $A_l := \sum_{\nu=1}^d \Delta_l^\nu A_\nu$

        Compute  $U_l^K := (1 + A_l/K)^K$  using a truncated approximation dependent on  $K > 0$ .

    Compute  $Q_m := \text{Tr}(U_1^K U_2^K \dots U_L^K)$

**return**  $Q_\Delta(M, N, K) := \frac{1}{M} \sum_{m=1}^M Q_m$

---

**Theorem 4.12.** *The output of  $\mathcal{A}^c(\{\Delta_l^\nu\})$  satisfies*

$$\mathbb{P}(|Q_\Delta(N, M, K) - \langle \gamma \rangle| > \epsilon) < \delta, \quad (72)$$

*for small  $\epsilon > 0$  provided that  $M > 2/\epsilon^2 \log(2/\delta)$  samples are taken, the dimension  $N$  of matrices satisfies  $N > e^{2\Delta_\gamma}/\epsilon^2$  and the matrix exponential is truncated to  $K > \Delta_\gamma e^{2\Delta_\gamma}/\epsilon$ .*

<sup>11</sup>The truncated product is a poor practical approach to compute a matrix exponential, in particular it is numerically unstable (see for example [MVo3] for a thorough discussion of numerical matrix exponentiation methods) but for our purpose of discussing separation between classical and quantum approaches it is sufficient.

*Proof.* See Appendix C.2 □

In Theorem 4.11 we argue that the resources required to approximate the development of a path with one-variation to accuracy  $\epsilon$  with the quantum algorithm are polynomial in  $1/\epsilon$  in time and logarithmic in  $1/\epsilon$  in space. In comparison, the classical algorithm outlined above requires  $L \times M \times K$  matrix multiplications and hence, according to Theorem 4.12, requires polynomial resources in  $\epsilon$  in both time and space.

**QSigKer.** We now consider the computation of the GUE signature kernel using the quantum algorithm  $\mathcal{A}$ . In the notation of the previous section, the GUE kernel of two paths  $\sigma$  and  $\tau$  may be expressed as

$$k^{\text{GUE}}(\sigma, \tau) = \langle \gamma \rangle, \quad (73)$$

where  $\gamma$  denotes the concatenation  $\gamma = \sigma \star \overleftarrow{\tau}$ . This follows from the property that for unitary path developments we have  $U_\gamma = U_\sigma U_\tau^\dagger$ . We conclude the work with the following definition:

**Definition 4.13.** (QSigKer) The probabilistic quantum signature kernel  $k_{\epsilon, \delta}^Q : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$  is defined by

$$k_{\epsilon, \delta}^Q(\sigma, \tau) := 1 - \frac{2T_\Delta(M, m, n, K)}{M}, \quad (74)$$

where  $T_\Delta(M, m, n, K)$  is the output of the algorithm of theorem 4.11. In applications, the hyperparameters should be chosen to scale according to Theorem 4.11.

## 5 Conclusion

In this work, we have developed a novel bridge between the theory of path signatures and matrix models in quantum field theory, uncovering structural connections with implications for both machine learning and quantum computing. By interpreting randomised path developments through the lens of non-commutative probability and unitary matrix models, we derived a class of loop equations governing their evolution. This formalism naturally led us to define the quantum path signature and the quantum signature kernel as quantum analogues of their classical counterparts, realised via quantum circuits acting on qubits. In particular, we showed that in the Gaussian case, these constructions admit an efficient quantum implementation using random ensembles of Pauli strings. This provides a concrete quantum algorithm for approximating signature kernels, with potential applications in quantum-enhanced time series analysis.

**Future directions.** This work represents an initial step toward importing tools from theoretical physics and quantum computing into the study of path signatures in machine learning. As such, we believe our work opens several avenues for future research, including the exploration of other matrix model potentials, deeper study of universality phenomena in quantum path signatures, and the practical integration of these quantum kernels into hybrid classical-quantum machine learning pipelines.

On the matrix model side, the theory of *topological recursion* [EO07] offers a powerful framework for computing  $1/N$  corrections to matrix model observables. A natural extension would be to explore the “loop” version of our model by applying topological recursion to correlators of the form

$$\begin{aligned} W_n(\gamma_1, \dots, \gamma_n; \{g_k\}) &= \left\langle \text{tr} \left( \text{P exp} \int_{\gamma_1} A \right) \cdots \text{tr} \left( \text{P exp} \int_{\gamma_n} A \right) \right\rangle \\ &= \sum_{g \geq 0} N^{2-2g-n} W_{g,n}(\gamma_1, \dots, \gamma_n; \{g_k\}), \end{aligned} \quad (75)$$

where the insertions correspond to a collection of paths  $\gamma_1, \dots, \gamma_n$ . This approach would enable a systematic investigation of finite- $N$  corrections to unitary path developments and their induced kernels.

On the machine learning side, many open questions remain concerning the statistical and computational properties of the interacting path kernels introduced in Section 3. In particular, it would be valuable to establish

theoretical guarantees such as generalisation bounds or sample complexity estimates. Extending these results to the quantum path signature defined in Definition 4.10 is especially compelling, as it could provide a rigorous foundation for the use of quantum path features in downstream learning tasks.

## References

- [ASS20] I. P. Arribas, C. Salvi, and L. Szpruch. “Sig-SDEs model for quantitative finance”. In: *ACM International Conference on AI in Finance*. 2020.
- [BDG01] G. Ben Arous, A. Dembo, and A. Guionnet. “Aging of spherical spin glasses”. In: *Probab. Theory Related Fields* 120.1 (2001), pp. 1–67. ISSN: 0178-8051,1432-2064. DOI: [10.1007/PL00008774](https://doi.org/10.1007/PL00008774).
- [BG97] G. Ben Arous and A. Guionnet. “Large deviations for Wigner’s law and Voiculescu’s non-commutative entropy”. In: *Probab. Theory Related Fields* 108.4 (1997), pp. 517–542. ISSN: 0178-8051,1432-2064. DOI: [10.1007/s004400050119](https://doi.org/10.1007/s004400050119).
- [BHS24] B. Barancikova, Z. Huang, and C. Salvi. *SigDiffusions: Score-Based Diffusion Models for Long Time Series via Log-Signature Embeddings*. 2024. arXiv: [2406.10354](https://arxiv.org/abs/2406.10354).
- [Boe+16] H. Boedihardjo, X. Geng, T. Lyons, and D. Yang. “The signature of a rough path: uniqueness”. In: *Advances in Mathematics* 293 (2016), pp. 720–737.
- [Che+24] C.-F. Chen, A. M. Dalzell, M. Berta, F. G. Brandão, and J. A. Tropp. “Sparse random Hamiltonians are quantumly easy”. In: *Physical Review X* 14.1 (2024), p. 011014.
- [Che54] K.-T. Chen. “Iterated integrals and exponential homomorphisms”. In: *Proc. London Math. Soc.* (3) 4 (1954), pp. 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”. In: *Annals of Mathematics* (1957), pp. 163–178.
- [Coc+21] T. Cochrane, P. Foster, V. Chhabra, M. Lemerrier, T. Lyons, and C. Salvi. “SK-Tree: a systematic malware detection algorithm on streaming trees via the signature kernel”. In: *2021 IEEE International Conference on Cyber Security and Resilience (CSR)*. IEEE. 2021, pp. 35–40.
- [CPP25] T. Cass, F. Piatti, and J. Pei. *Numerical schemes for signature kernels*. 2025. arXiv: [2502.08470](https://arxiv.org/abs/2502.08470).
- [CS24] T. Cass and C. Salvi. *Lecture Notes on Rough Paths and Applications to Machine Learning*. 2024. arXiv: [2404.06583](https://arxiv.org/abs/2404.06583).
- [CT24a] T. Cass and W. F. Turner. *Free probability, path developments and signature kernels as universal scaling limits*. 2024. arXiv: [2402.12311](https://arxiv.org/abs/2402.12311).
- [CT24b] T. Cass and W. F. Turner. “Topologies on unparameterised path space”. In: *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).
- [DT23] B. Dupire and V. Tissot-Daguette. *Functional Expansions*. 2023. arXiv: [2212.13628](https://arxiv.org/abs/2212.13628).
- [EKR15] B. Eynard, T. Kimura, and S. Ribault. *Random matrices*. 2015. arXiv: [1510.04430](https://arxiv.org/abs/1510.04430).
- [EO07] B. Eynard and N. Orantin. *Invariants of algebraic curves and topological expansion*. 2007. arXiv: [math-ph/0702045](https://arxiv.org/abs/math-ph/0702045).
- [Fer+23] A. Fermanian, T. Lyons, J. Morrill, and C. Salvi. “New directions in the applications of rough path theory”. In: *IEEE BITS the Information Theory Magazine* (2023).
- [GMO6] A. Guionnet and E. Maurel-Segala. “Combinatorial aspects of matrix models”. In: *ALEA Lat. Am. J. Probab. Math. Stat.* 1 (2006), pp. 241–279. ISSN: 1980-0436.
- [GMO7] A. Guionnet and E. Maurel-Segala. “Second order asymptotics for matrix models”. In: *Ann. Probab.* 35.6 (2007), pp. 2160–2212. ISSN: 0091-1798,2168-894X. DOI: [10.1214/009117907000000141](https://doi.org/10.1214/009117907000000141).

- [GM90] D. J. Gross and A. A. Migdal. “Nonperturbative two-dimensional quantum gravity”. In: *Physical Review Letters* 64.2 (1990), p. 127.
- [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).
- [HL10] B. Hambly and T. Lyons. “Uniqueness for the signature of a path of bounded variation and the reduced path group”. In: *Annals of Mathematics* (2010), pp. 109–167.
- [Hog+23] M. Hoglund, E. Ferrucci, C. Hernández, A. M. Gonzalez, C. Salvi, L. Sanchez-Betancourt, and Y. Zhang. *A Neural RDE approach for continuous-time non-Markovian stochastic control problems*. 2023. arXiv: [2306.14258](https://arxiv.org/abs/2306.14258).
- [Hor+23] B. Horvath, M. Lemerrier, C. Liu, T. Lyons, and C. Salvi. *Optimal Stopping via Distribution Regression: a Higher Rank Signature Approach*. 2023. arXiv: [2304.01479](https://arxiv.org/abs/2304.01479).
- [Horo3] K. Hori. *Mirror symmetry*. Vol. 1. American Mathematical Soc., 2003.
- [HS24] C. Holberg and C. Salvi. *Exact gradients for stochastic spiking neural networks driven by rough signals*. 2024. arXiv: [2405.13587](https://arxiv.org/abs/2405.13587).
- [Iss+24] Z. Issa, B. Horvath, M. Lemerrier, and C. Salvi. “Non-adversarial training of neural SDEs with signature kernel scores”. In: *Advances in Neural Information Processing Systems* 36 (2024).
- [Kid+19] P. Kidger, P. Bonnier, I. Perez Arribas, C. Salvi, and T. Lyons. “Deep signature transforms”. In: *Advances in Neural Information Processing Systems* 32 (2019).
- [Kit97] A. Y. Kitaev. “Quantum computations: algorithms and error correction”. In: *Russian Mathematical Surveys* 52.6 (1997), p. 1191.
- [KL21] P. Kidger and T. Lyons. *Signatory: differentiable computations of the signature and logsignature transforms, on both CPU and GPU*. 2021. arXiv: [2001.00706](https://arxiv.org/abs/2001.00706). URL: <https://github.com/patrick-kidger/signatory>.
- [KL98] E. Knill and R. Laflamme. “Power of one bit of quantum information”. In: *Physical Review Letters* 81.25 (1998), p. 5672.
- [Lai0] T. Lyons and al. *CoRoPa Computational Rough Paths (software library)*. 2010. URL: <https://coropa.sourceforge.io/>.
- [Lem+21a] M. Lemerrier, C. Salvi, T. Cass, E. V. Bonilla, T. Damoulas, and T. Lyons. “SigGPDE: Scaling Sparse Gaussian Processes on Sequential Data”. In: *International Conference on Machine Learning*. PMLR. 2021.
- [Lem+21b] M. Lemerrier, C. Salvi, T. Damoulas, E. Bonilla, and T. Lyons. “Distribution regression for sequential data”. In: *International Conference on Artificial Intelligence and Statistics*. PMLR. 2021, pp. 3754–3762.
- [LL24] M. Lemerrier and T. Lyons. *A high order solver for signature kernels*. 2024. arXiv: [2404.02926](https://arxiv.org/abs/2404.02926).
- [LLN16] D. Levin, T. Lyons, and H. Ni. *Learning from the past, predicting the statistics for the future, learning an evolving system*. 2016. arXiv: [1309.0260](https://arxiv.org/abs/1309.0260).
- [LLN22] H. Lou, S. Li, and H. Ni. *Path development network with finite-dimensional Lie group representation*. 2022. arXiv: [2204.00740](https://arxiv.org/abs/2204.00740).
- [Ly098] T. J. Lyons. “Differential equations driven by rough signals”. In: *Revista Matemática Iberoamericana* 14.2 (1998), pp. 215–310.
- [Man+24] G. Manten, C. Casolo, E. Ferrucci, S. W. Mogensen, C. Salvi, and N. Kilbertus. *Signature kernel conditional independence tests in causal discovery for stochastic processes*. 2024. arXiv: [2402.18477](https://arxiv.org/abs/2402.18477).



- [MLS23] N. Mua Cirone, M. Lemerrier, and 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. PMLR, 2023, pp. 25358–25425.
- [Mor+21] J. Morrill, C. Salvi, P. Kidger, and J. Foster. “Neural rough differential equations for long time series”. In: *International Conference on Machine Learning*. PMLR, 2021, pp. 7829–7838.
- [MS25a] N. Mua Cirone and C. Salvi. *ParallelFlow: Parallelizing Linear Transformers via Flow Discretization*. 2025. arXiv: [2504.00492](#).
- [MS25b] N. Mua Cirone and C. Salvi. *Rough kernel hedging*. 2025. arXiv: [2501.09683](#).
- [Mu+24] N. Mua Cirone, A. Orvieto, B. Walker, C. Salvi, and T. Lyons. *Theoretical foundations of deep selective state-space models*. 2024. arXiv: [2402.19047](#).
- [MV03] C. Moler and C. Van Loan. “Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later”. In: *SIAM review* 45.1 (2003), pp. 3–49.
- [NC10] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information*. Cambridge university press, 2010.
- [PS24] A. Pannier and C. Salvi. *A path-dependent PDE solver based on signature kernels*. 2024. arXiv: [2403.11738](#).
- [RG18] J. Reizenstein and B. Graham. *The iisignature library: efficient calculation of iterated-integral signatures and log signatures*. 2018. arXiv: [1802.08252](#).
- [Sal+21a] C. Salvi, T. Cass, J. Foster, T. Lyons, and W. Yang. “The signature kernel is the solution of a Goursat PDE”. In: *SIAM J. Math. Data Sci.* 3.3 (2021), pp. 873–899. ISSN: 2577-0187. DOI: [10.1137/20M1366794](#).
- [Sal+21b] C. Salvi, M. Lemerrier, C. Liu, B. Horvath, T. Damoulas, and T. Lyons. “Higher order kernel mean embeddings to capture filtrations of stochastic processes”. In: *Advances in Neural Information Processing Systems* 34 (2021), pp. 16635–16647.
- [Sal+23] C. Salvi, J. Diehl, T. Lyons, R. Preiss, and J. Reizenstein. “A structure theorem for streamed information”. In: *Journal of Algebra* 634 (2023), pp. 911–938.
- [Sal21] C. Salvi. “Rough paths, kernels, differential equations and an algebra of functions on streams”. PhD thesis. University of Oxford, 2021.
- [SK19] M. Schuld and N. Killoran. “Quantum machine learning in feature Hilbert spaces”. In: *Physical review letters* 122.4 (2019), p. 040504.
- [SLG22] C. Salvi, M. Lemerrier, and A. Gerasimovics. “Neural stochastic PDEs: Resolution-invariant learning of continuous spatiotemporal dynamics”. In: *Advances in Neural Information Processing Systems* 35 (2022), pp. 1333–1344.
- [Sri+23] P. Sriluckshmy, V. Pina-Canelles, M. Ponce, M. G. Algaba, F. Šimkovic IV, and M. Leib. “Optimal, hardware native decomposition of parameterized multi-qubit Pauli gates”. In: *Quantum Science and Technology* 8.4 (2023), p. 045029.
- [SS24] D. Shmelev and C. Salvi. *Sparse signature coefficient recovery via kernels*. 2024. arXiv: [2412.08579](#).
- [Suz76] M. Suzuki. “Generalized Trotter’s formula and systematic approximants of exponential operators and inner derivations with applications to many-body problems”. In: *Communications in Mathematical Physics* 51.2 (1976), pp. 183–190.
- [tH93] G. ’t Hooft. “A planar diagram theory for strong interactions”. In: *The Large N Expansion In Quantum Field Theory And Statistical Physics: From Spin Systems to 2-Dimensional Gravity*. World Scientific, 1993, pp. 80–92.
- [Ver18] R. Vershynin. *High-dimensional probability*. Vol. 47. Cambridge University Press, 2018.



## A Technical proofs

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.** *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  $\mathbb{E}[\exp(sN X_N^2)] \leq C^N$ .*

*Proof.* Assume (1), then

$$\begin{aligned} \mathbb{E}[\exp(sN X_N^2)] &= \int_0^\infty \mathbb{P}(\exp(sN X_N^2) \geq t) dt \\ &\leq 1 + 2 \int_0^\infty \mathbb{P}(X_N \geq u) sN u e^{sN u^2} du \\ &= 1 + \int_0^{t_0} 2sN u e^{sN u^2} du + \int_{t_0}^\infty 2sN u e^{-(\alpha-s)N u^2} du \\ &= 1 + e^{sN t_0^2} + \frac{s}{\alpha-s} \leq C^N, \end{aligned}$$

for  $C \geq 1 + e^{s t_0^2} + \frac{s}{\alpha-s}$  and  $s < \alpha$ . Conversely, assume (2), then Chebyshev's inequality yields

$$\mathbb{P}(X_N \geq t) \leq e^{-sN t^2} \mathbb{E}[\exp(sN X_N^2)] \leq e^{-sN t^2} C^N \leq e^{-\alpha N t^2},$$

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

$$\sigma^N(d\lambda) = \frac{1}{Z^N} \prod_{1 \leq i < j \leq N} |\lambda_i - \lambda_j|^2 \exp \left\{ -\frac{N}{2} \sum_{i=1}^N \lambda_i^2 \right\} d\lambda, \quad (76)$$

where  $Z^N$  is the normalisation constant given by

$$Z^N = (2\pi)^{\frac{N}{2}} N^{-\frac{N^2}{2}} \prod_{i=1}^N i!, \quad (77)$$

see [BG97, Property 3.1] with  $\beta = 2$ . The following pair of lemmas are key.

**Lemma A.2.** *There exists  $C > 0$  such that, for all  $N \geq 2$ ,  $\frac{Z^{N-1}}{Z^N} \leq e^{CN}$ .*

*Proof.* From (77), we have

$$\frac{Z^{N-1}}{Z^N} = \frac{(N-1)^{-\frac{(N-1)^2}{2}} N^{\frac{N^2}{2}}}{\sqrt{2\pi} \Gamma(N+1)} = \frac{\exp \left\{ \frac{1}{2} (N^2 \log(N) - (N-1)^2 \log(N-1)) \right\}}{\sqrt{2\pi} N!}.$$

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 [BDGo1, 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  $\sigma^N$ , as defined in (76).

**Lemma A.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*

$$\mathbb{P}_{\sigma_N} \left( \max_{i=1, \dots, N} |\lambda_i| \geq t \right) \leq e^{-\alpha N t^2}. \quad (78)$$

*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  $|x| \geq t \geq 8$ ,

$$\begin{aligned} \mathbb{P}_{\sigma_N}(|\lambda_1| \geq t) &\leq \frac{Z^{N-1}}{Z^N} e^{-\frac{1}{4} N t^2} \int_{|x| \geq t} e^{-\frac{x^2}{4}} \int \prod_{i=2}^N |x - \lambda_i|^2 e^{-\frac{\lambda_i^2}{2} - \frac{x^2}{4}} \sigma^{N-1}(\mathrm{d}\underline{\lambda}^{N-1}) \mathrm{d}x \\ &\leq \frac{Z^{N-1}}{Z^N} e^{-\frac{1}{4} N t^2} \int e^{-\frac{x^2}{4}} \int e^{\frac{x^2}{4} - \frac{x^2}{4}} \sigma^{N-1}(\mathrm{d}\underline{\lambda}^{N-1}) \mathrm{d}x \\ &\leq \frac{Z^{N-1}}{Z^N} e^{-\frac{1}{4} N t^2} \int e^{-\frac{x^2}{4}} \mathrm{d}x \\ &\leq e^{C N - \frac{1}{4} N t^2}, \end{aligned}$$

for some  $C > 0$ , where the last inequality follows from Lemma A.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

$$\mathbb{P}_{\sigma_N} \left( \max_{i=1, \dots, N} |\lambda_i| \geq t \right) \leq N \mathbb{P}_{\sigma_N}(|\lambda_1| \geq t) \leq e^{C N - \frac{1}{4} N t^2},$$

for some new constant  $C > 0$ . Then for any  $\alpha < \frac{1}{4}$  and  $t \geq t_0(\alpha)$ , (78) holds for all  $N$ .  $\square$

The proof of Lemma 3.10 now follows that of [GMo7, Lemma 2.2].

*Proof of Lemma 3.10.* 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  $e^{s N \lambda_{\max}^N(\mathbf{A})^2}$  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 [GMo7, 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

$$\int e^{s N \lambda_{\max}^N(\mathbf{A} - \mathbf{M})^2} \mathrm{d}\mu_V^N \leq \int e^{s N \lambda_{\max}^N(\mathbf{A})^2} \mathrm{d}\mu_c^N \leq C_2^N,$$

for all suitably small  $s$  and some  $C_2 > 0$ , and where  $\mu_c^N$  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. We conclude by observing that for some small  $s > 0$ ,

$$\int e^{s N \lambda_{\max}^N(\mathbf{A})^2} \mathrm{d}\mu_V^N \leq e^{2\lambda_{\max}^N(\mathbf{M})^2} \int e^{2s N \lambda_{\max}^N(\mathbf{A} - \mathbf{M})^2} \mathrm{d}\mu_V^N \leq e^{C_1} C_2^N,$$

and applying Lemma A.1.  $\square$

## B Proof of Theorem 4.8

**Definition B.1** (Pauli matrices). We denote by  $\mathbb{P} = \{\sigma_I, \sigma_X, \sigma_Y, \sigma_Z\}$  the set of Pauli matrices, introduced in Definition 4.1 and write  $\sigma \sim \mathcal{P}$  for a random variable  $\sigma$  uniformly distributed over  $\mathcal{P}$ .

We recall now some basic properties of the Pauli ensemble. Firstly, multiplication of Pauli matrices follows the following rules.

$\times$	$\sigma_I$	$\sigma_X$	$\sigma_Y$	$\sigma_Z$
$\sigma_I$	$\sigma_I$	$\sigma_X$	$\sigma_Y$	$\sigma_Z$
$\sigma_X$	$\sigma_X$	$\sigma_I$	$i\sigma_Z$	$-i\sigma_Y$
$\sigma_Y$	$\sigma_Y$	$-i\sigma_Z$	$\sigma_I$	$i\sigma_X$
$\sigma_Z$	$\sigma_Z$	$i\sigma_Y$	$-i\sigma_X$	$\sigma_I$

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

$$g = \phi(g)m(g), \quad \phi(g) \in \{1, i, -1, -i\} \quad \text{and} \quad m(g) \in \mathcal{P}. \quad (79)$$

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  $m(g) = m(h)$ ,  $m(g) = I$  or  $m(h) = I$  and  $gh = -hg$  if  $m(g) \neq m(h)$  and neither is the identity. Moreover,  $m(g) \in \mathcal{P}$  for any  $g \in G$ , it follows that  $|\text{tr}(g)| \leq 1$  for all  $g \in G$ .

**Lemma B.2.** 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 B.3.** 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 B.4.** 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  $\frac{1}{2}$ ,  $\mathbb{E}[XgX] = 0$ .  $\square$

We may now combine the above to conclude the following.

**Proposition B.5.** 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

$$\left| \mathbb{E}[\text{tr}(X_{w_1} \cdots X_{w_{2p}})] \right| \leq \frac{1}{4}. \quad (80)$$

*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 pair, resulting in a shorter word of  $2p - 2$  letters defining a non-crossing pair partition in  $p$ -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. We claim that there must be a subword of  $\mathbf{w}$  of the form  $X_j X_1 X_j$  for some word non-empty word  $\mathbf{l}$  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  $\mathbf{w} = \mathbf{k} j \mathbf{l} j \mathbf{m}$ , where  $\mathbf{k}, \mathbf{m}$  do not contain  $j$ . Hence

$$\begin{aligned}\mathbb{E}[\text{tr}(X_{\mathbf{w}})] &= \text{tr}(\mathbb{E}[X_{\mathbf{w}}]) \\ &= \text{tr}(\mathbb{E}[X_{\mathbf{k}} X_j X_1 X_j X_{\mathbf{m}}]) \\ &= \text{tr}(\mathbb{E}[X_{\mathbf{k}} \mathbb{E}[X_j X_1 X_j \mid X_i : i \in [p] \setminus j] X_{\mathbf{m}}]) \\ &= \text{tr}(\mathbb{E}[X_{\mathbf{k}} \mathbb{E}[X_j X_1 X_j \mid X_1] X_{\mathbf{m}}]) \\ &= \text{tr}(\mathbb{E}[X_{\mathbf{k}} \phi(X_1) \mathbb{1}_{\{m(X_1)=I\}} X_{\mathbf{m}}]) \\ &= \mathbb{E}[\phi(X_1) \mathbb{1}_{\{m(X_1)=I\}} \text{tr}(X_{\mathbf{k}} X_{\mathbf{m}})].\end{aligned}$$

The fifth equality used [Lemma B.4](#). Therefore

$$\left| \mathbb{E}[\text{tr}(X_{\mathbf{w}})] \right| \leq \mathbb{E}[\mathbb{1}_{\{m(X_1)=I\}} |\phi(X_1)| |\text{tr}(X_{\mathbf{k}} X_{\mathbf{m}})|] \leq \mathbb{P}(\{m(X_1) = I\}) = \frac{1}{4},$$

since  $|\phi(g)| = 1$  for all  $g \in G$  and  $|\text{tr}(g)| \leq 1$  for all  $g \in G$  and the final equality uses [Corollary B.3](#).  $\square$

**Lemma B.6.** 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

$$m(X_j) = \alpha^{-1} \quad \text{and} \quad m(X_j) = \beta^{-1} \tag{81}$$

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  $\sigma$ -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,

$$\mathbb{E} \left[ |\text{tr}(X_{\mathbf{u}_1} X_j X_{\mathbf{u}_2}) \text{tr}(X_{\mathbf{v}_1} X_j X_{\mathbf{v}_2})| \mid \mathcal{F} \right] = \mathbb{P}((81) \text{ holds} \mid \mathcal{F}) = \begin{cases} \frac{1}{4}, & \text{if } \alpha = \beta, \\ 0, & \text{if } \alpha \neq \beta. \end{cases}$$

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.  $\square$

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

$$\text{tr}(\sigma_{\mathbf{w}}) = \text{tr}(\sigma_{w_1}) \cdots \text{tr}(\sigma_{w_n}). \quad (82)$$

We recall again the definition of the random Pauli string ensemble.

**Definition B.7.** Let  $N = 2^n$ . 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  $r_i \stackrel{\text{i.i.d.}}{\sim} \text{Rad}(1/2)$  and  $\sigma^i = \sigma(1)^i \otimes \cdots \otimes \sigma(n)^i$ , where  $\sigma(j)^i \stackrel{\text{i.i.d.}}{\sim} \mathcal{P}$ .

**Definition B.8.** For any alphabet size  $m \geq 1$ , denoted by  $[m]$ , and integer  $p \geq 1$  define

- $W(m, p)$  : number of words of length  $2p$  where distinct letters appear an even number of times.
- $N(m, p)$  : number of words of length  $2p$  where distinct letters appear exactly twice.

**Lemma B.9.** For any  $m \geq 1$  and  $p \geq 1$ ,

$$W(m, p) = \frac{1}{2^m} \sum_{k=0}^m \binom{m}{k} (m - 2k)^{2p} = \sum_{\substack{k_1 + \cdots + k_m = p \\ k_i \geq 0}} \frac{(2p)!}{(2k_1)! \cdots (2k_m)!}. \quad (83)$$

*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  $\epsilon$ , 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

$$W(m, p) = \frac{1}{2^m} \sum_{\epsilon \in \{0,1\}^m} \left( \sum_{i=1}^m (-1)^{\epsilon_i} \right)^{2p} = \frac{1}{2^m} \sum_{t=0}^m \binom{m}{t} (m - 2t)^{2p}.$$

To see 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 B.10.** For any  $m \geq 1$  and  $p \geq 1$ ,

$$N(m, p) = \binom{m}{p} \frac{(2p)!}{2^p} = \sum_{\substack{k_1 + \dots + k_m = p \\ k_i \in \{0, 1\}}} \frac{(2p)!}{(2k_1)! \dots (2k_m)!}. \quad (84)$$

*Proof.* We have to choose which  $p$  letters out of the  $m$  possible letters will appear, and then arrange those  $p$  letters twice each in a word of length  $2p$ . 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

$$\frac{(2p)!}{(2k_1)! (2k_2)! \dots (2k_m)!} = \frac{(2p)!}{\underbrace{2! \dots 2!}_{p \text{ factors}} \underbrace{0! \dots 0!}_{m-p \text{ factors}}} = \frac{(2p)!}{(2!)^p} = \frac{(2p)!}{2^p}.$$

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 B.11.** 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  $k \geq 0$ ,  $(2k)! = 2^k k! (1 \cdot 3 \cdot \dots \cdot (2k - 1)) \geq 2^k k!$ . Hence, from (83) we obtain

$$W(m, p) \leq \frac{(2p)!}{2^p} \sum_{\substack{k_1 + \dots + k_m = p \\ k_i \geq 0}} \frac{1}{k_1! \dots k_m!} = \frac{(2p)!}{2^p p!} m^p$$

where we used the fact that

$$\sum_{\substack{k_1 + \dots + k_m = p \\ k_i \geq 0}} \frac{1}{k_1! \dots k_m!} = \frac{m^p}{p!}.$$

Finally the lemma follows since  $(2p)! = 2^p p! (1 \cdot 3 \cdot \dots \cdot (2p - 1)) \leq 2^p p! (2 \cdot 4 \cdot \dots \cdot 2p) = (2^p p!)^2$ .  $\square$

**Lemma B.12.** For any  $m \geq 1$  and  $p \geq 1$  we have

$$W(m, p) - N(m, p) \leq (4e)^p m^{p-1} p! \quad (85)$$

*Proof.* Using (83) and (84) we obtain

$$W(m, p) - N(m, p) = \sum_{\substack{k_1 + \dots + k_m = p \\ \text{at least one } k_i \geq 2}} \frac{(2p)!}{(2k_1)! \dots (2k_m)!}.$$

Using similar arguments as in the proof of [Lemma B.11](#),

$$\begin{aligned}
W(m, p) - N(m, p) &\leq \frac{(2p)!}{2^p} \sum_{\substack{k_1 + \dots + k_m = p \\ \text{at least one } k_i \geq 2}} \frac{1}{k_1! \dots k_m!} \\
&= \frac{(2p)!}{2^p} \left( \sum_{\substack{k_1 + \dots + k_m = p \\ k_i \geq 0}} \frac{1}{k_1! \dots k_m!} - \sum_{\substack{k_1 + \dots + k_m = p \\ k_i \in \{0, 1\}}} \frac{1}{k_1! \dots k_m!} \right) \\
&= \frac{(2p)!}{2^p} \left( \frac{m^p}{p!} - \binom{m}{p} \right) \\
&= \frac{(2p)!}{2^p p!} \left( m^p - \frac{m!}{(m-p)!} \right) \\
&= \frac{(2p)!}{2^p p!} m^{p-1} \left( m - m \left( 1 - \frac{1}{m} \right) \left( 1 - \frac{2}{m} \right) \dots \left( 1 - \frac{p-1}{m} \right) \right)
\end{aligned}$$

Now, for any  $j = 1, \dots, p-1$  then  $0 < \frac{j}{m} < 1$  and  $(1 - \alpha) \geq \exp\left(-\frac{\alpha}{1-\alpha}\right)$  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.  $\square$

**Proposition B.13.** *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  $\tau$  is the law of  $d$ -free semicircular random variables.

*Proof.* First note that 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  $m_n$ -letters of length  $|\mathbf{w}|$ . Since  $\mathbb{E}[r_{\mathbf{w}}^{\mathbf{u}}]$  disappears unless each letter in  $\mathbf{u}$  appears an even number of times, we need only consider the case where  $|\mathbf{w}| = 2p$  for  $p \in \mathbb{N}$ . In this case, we still require 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 (82) and the property of traces of the group  $G$ , then  $|\mathbb{E}[\text{tr}(\sigma_w^u)]| \leq 1$ . Combining these facts yields

$$\sum_{\mathbf{u}: \text{wt}(\mathbf{u}) \leq p-1} |\mathbb{E}[r_w^u] \mathbb{E}[\text{tr}(\sigma_w^u)]| \leq \sum_{t=1}^{p-1} m_n^t t^{2p} \leq m_n^{p-1} p^{2p+1}. \quad (86)$$

We turn our attention to words of weight  $p$ . First, we see that

$$\mathbb{E}[\text{tr}(\sigma_w^u)] = \mathbb{E}[\text{tr}(\sigma(1)_w^u) \cdots \text{tr}(\sigma(n)_w^u)] = \mathbb{E}[\text{tr}(\sigma(1)_w^u)]^n,$$

by the i.i.d. construction of the Pauli strings. Consider first case where pairing equal letters in  $\mathbf{u}$  defines a crossing pair partition, then by Proposition B.5 we see that know that

$$|\mathbb{E}[\text{tr}(\sigma(1)_w^u)]^n| \leq 2^{-2n}.$$

The number of words of length  $2p$  with weight  $p$ , where distinct letters appear exactly twice may be bounded by  $m^p \frac{p^{2p}}{p!}$ . 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

$$\mathbb{E}[r_w^u] \mathbb{E}[\text{tr}(\sigma_w^u)] = \prod_{(a,b) \in \pi_u} \mathbb{1}_{w_a = w_b},$$

where  $\pi_u$  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

$$\prod_{(a,b) \in \pi_u} \mathbb{1}_{w_a = w_b} = 1.$$

Recalling the discussion in Section 3.1, this is one of the defining properties of  $\tau(X_w)$ . The number of words in each equivalence class is  $m_n(m_n - 1) \cdots (m_n - p + 1)$ . It follows that

$$\lim_{n \rightarrow \infty} \mathbb{E}[\text{tr}(A_w^n)] = \lim_{n \rightarrow \infty} \left\{ \tau(X_w) \frac{m_n(m_n - 1) \cdots (m_n - p + 1)}{m^p} + \mathcal{O}(2^{-2n}) + \mathcal{O}(m_n^{-1}) \right\} = \tau(X_w),$$

where the first term is the contribution of the non-crossing pair partitions, the second 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 B.14.** *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. Then*

$$\left| \mathbb{E}[\text{tr}(A_w^n)] - \tau(X_w) \right| \leq (4e)^p p! \left( m_n^{-1} + 2^{-2n} \right),$$

where  $\tau$  is the law of  $d$ -free semicircular random variables and  $w \in \mathcal{W}_d$  has length  $2p$  for some  $p \in \mathbb{N}$ .

*Proof.* As in the previous proposition, we note that

$$\mathbb{E}[\text{tr}(A_{\mathbf{w}}^n)] = m_n^{-|\mathbf{w}|/2} \sum_{\mathbf{u}} \mathbb{E}[r_{\mathbf{w}}^{\mathbf{u}}] \mathbb{E}[\text{tr}(\sigma_{\mathbf{w}}^{\mathbf{u}})],$$

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 B.12](#) we obtain directly

$$m_n^{-p} \sum_{\mathbf{u} : \text{wt}(\mathbf{u}) < p} \left| \mathbb{E}[r_{\mathbf{w}}^{\mathbf{u}}] \mathbb{E}[\text{tr}(\sigma_{\mathbf{w}}^{\mathbf{u}})] \right| = m_n^{-p} (W(m_n, p) - N(m_n, p)) \leq (4e)^p p! \frac{1}{m_n} \quad (87)$$

(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 B.5](#)

$$\mathbb{E}[\text{tr}(\sigma_{\mathbf{w}}^{\mathbf{u}})] = \mathbb{E}[\text{tr}(\sigma(1)_{\mathbf{w}}^{\mathbf{u}})]^n \leq 2^{-2n}.$$

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,

$$m_n^{-p} \sum_{\substack{\mathbf{u} : \text{wt}(\mathbf{u}) = p \\ \text{Case 1}}} \left| \mathbb{E}[r_{\mathbf{w}}^{\mathbf{u}}] \mathbb{E}[\text{tr}(\sigma_{\mathbf{w}}^{\mathbf{u}})] \right| \leq \frac{(2p)!}{2^p p!} 2^{-2n} \leq (4e)^p p! 2^{-2n}, \quad (88)$$

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

$$\left| \frac{m(m-1) \cdots (m-p+1)}{m^p} - 1 \right| \leq \frac{p^2}{2m}, \quad (89)$$

yields

$$\left| m_n^{-p} \sum_{\substack{\mathbf{u} : \text{wt}(\mathbf{u}) = p \\ \text{Case 2}}} \mathbb{E}[r_{\mathbf{w}}^{\mathbf{u}}] \mathbb{E}[\text{tr}(\sigma_{\mathbf{w}}^{\mathbf{u}})] - \tau(X_{\mathbf{w}}) \right| \leq \frac{p^2}{2m_n} |\tau(X_{\mathbf{w}})| \leq \frac{(4e)^p p!}{m_n}, \quad (90)$$

where we used that fact that  $|\tau(X_{\mathbf{w}})| \leq 4^p$ .

Combining the bounds (87), (88) and (90) yields

$$\left| \mathbb{E}[\text{tr}(A_{\mathbf{w}}^n)] - \tau(X_{\mathbf{w}}) \right| \leq (4e)^p p! \left( 2m_n^{-1} + 2^{-2n} \right),$$

□

**Proposition B.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. Then

$$\left| \mathbb{E} \left[ \text{tr}(A_{\mathbf{u}}^n) \text{tr}(A_{\mathbf{v}}^n) \right] - \tau(X_{\mathbf{u}}) \tau(X_{\mathbf{v}}) \right| \leq (4e)^p p! \left( 2m_n^{-1} + 2^{-2n} \right),$$

for all words  $\mathbf{u}, \mathbf{v} \in \mathcal{W}_d$  such that  $|\mathbf{u}| + |\mathbf{v}| = 2p$  for some  $p \in \mathbb{N}$ , with the left-hand-side being zero otherwise.

*Proof.* We first write

$$\mathbb{E} \left[ \text{tr}(A_{\mathbf{u}}^n) \text{tr}(A_{\mathbf{v}}^n) \right] = m_n^{-(|\mathbf{u}|+|\mathbf{v}|)/2} \sum_{\mathbf{a}, \mathbf{b}} \mathbb{E}[r_{\mathbf{u}}^{\mathbf{a}} r_{\mathbf{v}}^{\mathbf{b}}] \mathbb{E}[\text{tr}(\sigma_{\mathbf{u}}^{\mathbf{a}}) \text{tr}(\sigma_{\mathbf{v}}^{\mathbf{b}})],$$

where the sum runs over words  $\mathbf{a}$  and  $\mathbf{b}$  in the letters  $[m]$  of length  $|\mathbf{u}|$  and  $|\mathbf{v}|$  respectively. Due to the Rademacher random variables, we need only consider the case where  $|\mathbf{u}| + |\mathbf{v}| = 2p$  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 some into three parts.

(i) (Words of weight  $t < p$ ) Using Proposition B.5 we directly compute

$$\begin{aligned} m_n^{-p} \sum_{\mathbf{a}, \mathbf{b}: \text{wt}(\mathbf{ab}) < p} \left| \mathbb{E}[r_{\mathbf{u}}^{\mathbf{a}} r_{\mathbf{v}}^{\mathbf{b}}] \mathbb{E}[\text{tr}(\sigma_{\mathbf{u}}^{\mathbf{a}}) \text{tr}(\sigma_{\mathbf{v}}^{\mathbf{b}})] \right| &= m_n^{-p} (W(m_n, p) - N(m_n, p)) \\ &\leq (4e)^p p! \frac{1}{m_n} \end{aligned} \quad (91)$$

(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,

$$\begin{aligned} \mathbb{E}[\text{tr}(\sigma_{\mathbf{u}}^{\mathbf{a}}) \text{tr}(\sigma_{\mathbf{v}}^{\mathbf{b}})] &= \mathbb{E}[\text{tr}(\sigma(1)_{\mathbf{u}}^{\mathbf{a}}) \text{tr}(\sigma(1)_{\mathbf{v}}^{\mathbf{b}}) \cdots \text{tr}(\sigma(n)_{\mathbf{u}}^{\mathbf{a}}) \text{tr}(\sigma(n)_{\mathbf{v}}^{\mathbf{b}})] \\ &= \mathbb{E}[\text{tr}(\sigma(1)_{\mathbf{u}}^{\mathbf{a}}) \text{tr}(\sigma(1)_{\mathbf{v}}^{\mathbf{b}})]^n, \end{aligned}$$

we may apply Proposition B.5 and Lemma B.6 to see that

$$m_n^{-p} \sum_{\substack{\mathbf{a}, \mathbf{b}: \text{wt}(\mathbf{ab})=p \\ \text{Case 1}}} \left| \mathbb{E}[r_{\mathbf{u}}^{\mathbf{a}} r_{\mathbf{v}}^{\mathbf{b}}] \mathbb{E}[\text{tr}(\sigma_{\mathbf{u}}^{\mathbf{a}}) \text{tr}(\sigma_{\mathbf{v}}^{\mathbf{b}})] \right| \leq \frac{(2p)!}{2^p p!} 2^{-2n} \leq (4e)^p p! 2^{-2n}, \quad (92)$$

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

$$\mathbb{E}[\text{tr}(\sigma_{\mathbf{u}}^{\mathbf{a}}) \text{tr}(\sigma_{\mathbf{v}}^{\mathbf{b}})] = \mathbb{E}[\text{tr}(\sigma(1)_{\mathbf{u}}^{\mathbf{a}})]^n \mathbb{E}[\text{tr}(\sigma(1)_{\mathbf{v}}^{\mathbf{b}})]^n = 1.$$

We then see that

$$\mathbb{E}[r_{\mathbf{u}}^{\mathbf{a}} r_{\mathbf{v}}^{\mathbf{b}}] \mathbb{E}[\text{tr}(\sigma_{\mathbf{u}}^{\mathbf{a}}) \text{tr}(\sigma_{\mathbf{v}}^{\mathbf{b}})] = \prod_{(a,b) \in \pi_{\mathbf{a}}} \mathbb{1}_{u_a = u_b} \prod_{(c,d) \in \pi_{\mathbf{b}}} \mathbb{1}_{v_c = v_d},$$

where  $\pi_{\mathbf{a}}$  is the non-crossing pair partition of the letters  $\{1, \dots, |\mathbf{u}|\}$  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  $\tau(X_{\mathbf{u}}) \tau(X_{\mathbf{v}})$ . Using the counting arguments as in the previous proposition, we have that

$$m_n^{-p} \sum_{\substack{\mathbf{a}, \mathbf{b}: \text{wt}(\mathbf{ab})=p \\ \text{Case 2}}} \mathbb{E}[r_{\mathbf{u}}^{\mathbf{a}} r_{\mathbf{v}}^{\mathbf{b}}] \mathbb{E}[\text{tr}(\sigma_{\mathbf{u}}^{\mathbf{a}}) \text{tr}(\sigma_{\mathbf{v}}^{\mathbf{b}})] = \tau(X_{\mathbf{u}}) \tau(X_{\mathbf{v}}) \frac{m_n(m_n - 1) \cdots (m_n - p + 1)}{m_n^p}.$$

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

$$\left| m_n^{-p} \sum_{\substack{\mathbf{a}, \mathbf{b}: \text{wt}(\mathbf{ab})=p \\ \text{Case 2}}} \mathbb{E} \left[ r_{\mathbf{u}}^{\mathbf{a}} r_{\mathbf{v}}^{\mathbf{b}} \right] \mathbb{E} \left[ \text{tr}(\sigma_{\mathbf{u}}^{\mathbf{a}}) \text{tr}(\sigma_{\mathbf{v}}^{\mathbf{b}}) \right] - \tau(X_{\mathbf{u}}) \tau(X_{\mathbf{v}}) \right| \leq (4e)^p p! \frac{1}{m_n}. \quad (93)$$

Combining (91)-(92)-(93) gives

$$\left| m_n^{-p} \sum_{\mathbf{a}, \mathbf{b}} \mathbb{E} \left[ r_{\mathbf{u}}^{\mathbf{a}} r_{\mathbf{v}}^{\mathbf{b}} \right] \mathbb{E} \left[ \text{tr}(\sigma_{\mathbf{u}}^{\mathbf{a}}) \text{tr}(\sigma_{\mathbf{v}}^{\mathbf{b}}) \right] - \tau(X_{\mathbf{u}}) \tau(X_{\mathbf{v}}) \right| \leq (4e)^p p! \left( 2m_n^{-1} + 2^{-2n} \right).$$

□

As a corollary of all the results, we can now prove [Theorem 4.8](#). We state it again for convenience.

**Theorem B.16** ([Theorem 4.8](#)). *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  $U_\gamma^n$  be the solution to*

$$dU_\gamma^n(s, t) = i \sum_{j=1}^d U_\gamma^n(s, t) A_j^n d\gamma_t^j, \quad U_\gamma^n(s, s) = I,$$

and let

$$\langle \gamma_{s,t} \rangle_{\mu_V^\infty} = \sum_{\mathbf{w} \in \mathcal{W}_d} i^{|\mathbf{w}|} \tau(X_{\mathbf{w}}) \mathcal{S}_{s,t}^{\mathbf{w}}(\gamma).$$

Then there exists  $C > 0$  depending only on the length of  $\gamma$  and the dimension of the underlying space  $V$  for which

$$\mathbb{E} \left[ \left( \text{tr}(U_\gamma^n(s, t)) - \langle \gamma_{s,t} \rangle_{\mu_V^\infty} \right)^2 \right] < C(m_n^{-1} + 2^{-2n}). \quad (94)$$

*Proof.* We start by expanding

$$\begin{aligned} \left( \text{tr}(U_\gamma^n(s, t)) - \langle \gamma_{s,t} \rangle_{\mu_V^\infty} \right)^2 &= \sum_{\mathbf{u}, \mathbf{v} \in \mathcal{W}_d} i^{|\mathbf{u}|+|\mathbf{v}|} \left( \text{tr}(A_{\mathbf{u}}) \text{tr}(A_{\mathbf{v}}) - \tau(X_{\mathbf{u}}) \tau(X_{\mathbf{v}}) \right. \\ &\quad \left. + \tau(X_{\mathbf{u}}) \tau(X_{\mathbf{v}}) - \text{tr}(A_{\mathbf{u}}) \tau(X_{\mathbf{v}}) \right. \\ &\quad \left. - \tau(X_{\mathbf{u}}) \text{tr}(A_{\mathbf{v}}) + \tau(X_{\mathbf{u}}) \tau(X_{\mathbf{v}}) \right) \mathcal{S}_{s,t}^{\mathbf{u}}(\gamma) \mathcal{S}_{s,t}^{\mathbf{v}}(\gamma). \end{aligned}$$

We start by dealing with the first term.

$$\begin{aligned}
& \sum_{\mathbf{u}, \mathbf{v} \in \mathcal{W}_d} \left| \mathbb{E}[\text{tr}(A_{\mathbf{u}}) \text{tr}(A_{\mathbf{v}})] - \tau(X_{\mathbf{u}}) \tau(X_{\mathbf{v}}) \right| \\
& \leq \sum_{p=0}^{\infty} \sum_{|\mathbf{u}|=0}^{2p} \sum_{|\mathbf{v}|=2p-|\mathbf{u}|} \left| \mathbb{E}[\text{tr}(A_{\mathbf{u}}) \text{tr}(A_{\mathbf{v}})] - \tau(X_{\mathbf{u}}) \tau(X_{\mathbf{v}}) \right| \left| \mathcal{S}_{s,t}^{\mathbf{u}}(\gamma) \mathcal{S}_{s,t}^{\mathbf{v}}(\gamma) \right| \\
& \leq \left( 2m_n^{-1} + 2^{-2n} \right) \sum_{p=0}^{\infty} \sum_{k=0}^{2p} \sum_{\substack{|\mathbf{u}|=k \\ |\mathbf{v}|=2p-k}} (4e)^p p! \cdot \left| \mathcal{S}_{s,t}^{\mathbf{u}}(\gamma) \mathcal{S}_{s,t}^{\mathbf{v}}(\gamma) \right| \\
& \leq \left( 2m_n^{-1} + 2^{-2n} \right) \sum_{p=0}^{\infty} \sum_{k=0}^{2p} \frac{(4e)^p p! \cdot d^{2p} \|\gamma\|_{\text{I-var}}^{2p}}{k! (2p-k)!} \\
& \leq \left( 2m_n^{-1} + 2^{-2n} \right) \sum_{p=0}^{\infty} \frac{(2p+1)(4e)^p \cdot d^{2p} \|\gamma\|_{\text{I-var}}^{2p}}{p!} \\
& \leq C_1 (\|\gamma\|_{\text{I-var}}, d) \left( 2m_n^{-1} + 2^{-2n} \right).
\end{aligned}$$

The remaining two terms may each be bounded by

$$\begin{aligned}
& \sum_{\mathbf{u}, \mathbf{v} \in \mathcal{W}_d} \left| \mathbb{E}[\text{tr}(A_{\mathbf{u}})] - \tau(X_{\mathbf{u}}) \right| \tau(X_{\mathbf{v}}) \left| \mathcal{S}_{s,t}^{\mathbf{u}}(\gamma) \mathcal{S}_{s,t}^{\mathbf{v}}(\gamma) \right| \\
& \leq \sum_{\mathbf{u} \in \mathcal{W}_d} \left| \mathbb{E}[\text{tr}(A_{\mathbf{u}})] - \tau(X_{\mathbf{u}}) \right| \left| \mathcal{S}_{s,t}^{\mathbf{u}}(\gamma) \right| \sum_{\mathbf{v} \in \mathcal{W}_d} \tau(X_{\mathbf{v}}) \left| \mathcal{S}_{s,t}^{\mathbf{v}}(\gamma) \right| \\
& \leq \left( 2m_n^{-1} + 2^{-2n} \right) \sum_{p=0}^{\infty} \frac{(4e)^p d^{2p} \|\gamma\|_{\text{I-var}}^{2p}}{p!} \sum_{q=0}^{\infty} \frac{2^q d^{2q} \|\gamma\|_{\text{I-var}}^{2q}}{(2q)!} \\
& \leq C_2 (\|\gamma\|_{\text{I-var}}, d) \left( 2m_n^{-1} + 2^{-2n} \right).
\end{aligned}$$

Taking  $C = 3 \max\{C_1, C_2\}$  yields the result.  $\square$

## C Proofs of Theorem 4.11 and Theorem 4.12

We now prove the correctness and provide the complexity analysis of the quantum and classical algorithms.

### C.1 Quantum algorithm: Proof of Theorem 4.11

We state Theorem 4.11 again for convenience.

**Theorem C.1 (Theorem 4.11).** *The positive integer parameters  $m, n, K, M$  may be chosen such that the output of  $\mathcal{A}(\{\Delta_l^\nu\}, \epsilon, \delta)$  estimates the unitary path development  $\langle \gamma \rangle$  to within additive error  $\epsilon$  with probability  $1 - \delta$ :*

$$\mathbb{P}(|Q_{\Delta}(M, m, n, K) - \langle \gamma \rangle| \geq \epsilon) < \delta. \quad (95)$$

*The circuit requires  $\log(1/\epsilon, 1/\delta)$  qubits and  $\text{poly}(1/\epsilon, 1/\delta)$  gates.*

*Proof.* We use the triangle inequality and union bound to find

$$\mathbb{P}(|Q_{\Delta}(M, m, n, K) - \langle \gamma \rangle| > \epsilon) \leq \mathbb{P}\left(\left|Q_{\Delta}(M, m, n, K) - \mathbb{E}[\text{tr}(U_{\gamma}(\alpha, n, K))]\right| > \frac{\epsilon}{3}\right), \quad (96)$$

provided

$$|\mathbb{E}[U_{\gamma}(\alpha, n, K)] - \mathbb{E}[U_{\gamma}(\alpha, n)]| < \frac{\epsilon}{3}, \quad |\mathbb{E}[U_{\gamma}(\alpha, n)] - \langle \gamma \rangle| < \frac{\epsilon}{3}. \quad (97)$$

The first of these inequalities is ensured by Lemma 4.3 provided  $K > 3\Delta_\gamma^2 m/\epsilon$  and the second is ensured by Theorem 4.8 with  $m = n$  provided  $n > \max(6C/\epsilon, \log(6C/\epsilon))$ . Finally, we ensure that the right-hand side of (96) is bounded by  $\delta$  using Hoeffding's inequality (noting  $|\text{tr}(U)| < 1$ ). We thus take  $M > \frac{2}{\epsilon^2} \log(2/\delta)$ .  $\square$

## C.2 Classical Monte Carlo algorithm: Proof of Theorem 4.12

We now prove Theorem 4.12. We restate the classical algorithm and the theorem statement for convenience.

---

### Algorithm 3 Classical Sampling $\mathcal{A}^c$

---

**Input:** Data: path increments  $\{\Delta_l^1, \dots, \Delta_l^d\}_{l=1}^L$ . Parameters: Integers  $N > 0$ ,  $M > 0$  and  $K > 0$ .

Draw  $d$  independent  $N \times N$  GUE matrices  $\{A_\nu\}_{\nu=1}^d$

**for**  $m = 1$  to  $M$  **do**

**for**  $l = 1$  to  $L$  **do**

        Compute  $A_l := \sum_{\nu=1}^d \Delta_l^\nu A_\nu$

        Compute  $U_l^K := (1 + A_l/K)^K$  using a truncated approximation dependent on  $K > 0$

    Compute  $Q_m := \text{Tr}(U_1^K U_2^K \dots U_L^K)$

**return**  $Q_\Delta(M, N, K) := \frac{1}{M} \sum_{m=1}^M Q_m$

---

**Theorem C.2.** *The output of  $\mathcal{A}^c(\{\Delta_i^\nu\})$  satisfies*

$$\mathbb{P}(|Q_\Delta(N, M, K) - \langle \gamma \rangle| > \epsilon) < \delta, \quad (98)$$

for small  $\epsilon > 0$  provided that  $M > 2/\epsilon^2 \log(2/\delta)$  samples are taken, the dimension  $N$  of matrices satisfies  $N > e^{2\Delta_\gamma}/\epsilon^2$  and the matrix exponential is truncated to  $K > \Delta_\gamma e^{2\Delta_\gamma}/\epsilon$ .

We begin with a series of lemmas.

**Lemma C.3.** *The estimator  $Q_\Delta(N, M, K)$  of  $\mathbb{E}[\text{tr}(U(N, K))]$  satisfies*

$$\mathbb{P}(|Q_\Delta(M, N, K) - \mathbb{E}[\text{tr}(U(N, K))]| > \epsilon) < \delta, \quad (99)$$

provided  $M > \frac{2}{\epsilon^2} \log(2/\delta)$  by Hoeffding's inequality.

**Lemma C.4.** *The truncated matrix exponential  $U(N, K) := U_1^K \dots U_L^K$  satisfies*

$$\mathbb{E}[\text{tr}(U(N, K)) - \text{tr}(U(N))] < \epsilon, \quad (100)$$

provided  $K > \frac{\Delta_l e^{2\Delta_l}}{2/\epsilon}$ .

*Proof.* We write  $A_l := \sum_{\nu=1}^d \Delta_l^\nu A_\nu$ . It is then a standard GUE result that

$$\|A_l\| \leq \max_\nu(\Delta_l^\nu) \left(1 + \frac{\mu}{N^{2/3}} + o(N^{-2/3})\right), \quad (101)$$

where  $\Delta_l := \max_\nu(\Delta_l^\nu)$  and  $\mu$  is the mean of the Tracy-Widom distribution. Now we write  $f_K(A) = (1 + A/K)^K$ . It is straightforward to obtain the bound

$$\mathbb{E}[\|\exp(A_l) - f_K(A_l)\|] \leq \frac{\Delta_l}{K} e^{2\Delta_l} \left(1 + \frac{\mu}{N^{2/3}} + o(N^{-2/3})\right), \quad (102)$$

for each  $l = 1, \dots, L$  this quantity is thus bounded by  $\epsilon$  provided  $K > \Delta_\gamma e^{2\Delta_\gamma}/K$ . Finally, we note

$$\mathbb{E}[\text{tr}(U_1 \dots U_L) - \text{tr}(f_K(A_1) \dots f_K(A_L))] < \epsilon. \quad (103)$$

$\square$

**Lemma C.5.** *The convergence of the finite  $N$  development to the unitary development satisfies*

$$\mathbb{E} \left[ \left| \text{tr}(U(N)) - \langle \gamma \rangle \right| \right] < \epsilon, \quad (104)$$

provided  $N > e^{2\Delta_\gamma} / \epsilon^2$ .

We now turn to the proof of [Theorem 4.12](#).

*Proof.* We first note that

$$\left| Q_\Delta - \langle \gamma \rangle \right| \leq \left| Q_\Delta - \mathbb{E}[\text{tr}(U(N, K))] \right| + \left| \mathbb{E}[\text{tr}(U(N))] - \mathbb{E}[\text{tr}(U(N, K))] \right| + \left| \mathbb{E}[\text{tr}(U(N))] - \langle \gamma \rangle \right|. \quad (105)$$

Thus

$$\mathbb{P}(|Q_\Delta - \langle \gamma \rangle| > \epsilon) \leq \mathbb{P}(|Q_\Delta - \mathbb{E}[\text{tr}(U(N, K))]| > \epsilon) \quad (106)$$

provided, from [Lemma C.5](#) and [Lemma C.4](#),  $N > e^{2\Delta_\gamma} / \epsilon^2$  and  $K > \frac{\Delta_\gamma e^{2\Delta_\gamma}}{2/\epsilon}$ . Finally, we apply Hoeffding's inequality and [Lemma C.3](#), to obtain the result.  $\square$