Non Commutative Fourier Transforms

Application to path spaces

Leon CHATAIGNAULT

October 8, 2024

Abstract

Fourier Analysis is a powerful tool to capture symmetries and encode global stationary information. Harmonic Analysis extends the study to other algebraic structures than the real line, including groups. The theory of representation enables the analysis on non-commutative groups such as matrix or paths groups. As the topologies on paths spaces are beyond the classical scope, tailored embedding methods have been developed : detailing each with their properties is the core objective of this thesis.

Acknowledgements

I am warmly thankful to Pr. Cristopher Salvi for providing this interesting subject and for his insightful supervision throughout the year. I would also like to thank Pr. Laurent Moonens for giving me a taste for real harmonic analysis and related advised readings.

Contents

0 Introduction

0.1 Motivation

The push to develop a Non-Commutative Fourier Transform (NCFT) on paths partly takes its roots in dual problems of quantum mechanics, inasmuch as the Lagrangian approach uses path integrals whereas the Hamiltonian approach uses non-commutative variables.

Besides, the study of the Heisenberg group is fundamental to understanding the algebra of quantum observables. Furthermore, it relates to the Wilson-loop functional in Yang-Mills theory which aims at describing nuclear bindings. Fourier series on non-commutative spaces is still an active field of research, for example to solve sublaplacian equations on abstract algebraic groups [Bah+23].

From a practical point of view, having a Fourier transform and a dual space on spaces of paths could find inference applications in characterising measures or functions on paths. It has already been successfully applied in distribution regression [Lem+21] or as a fixed-size feature extraction layer [LLN22].

The objective of the project is to describe possible analogs of the Fourier transform for functions on the space of paths in $\mathbb{R}^n$. Existing methods have gained interest in machine learning [LLN22] [LLN24] and statistics [CO18]. Besides, bridging the gap between rough paths and non-commutative harmonic analysis has previously been done in various ways [Faw03] [Bau04]. One underlying objective is, given different such approaches, to compare their analytic and algebraic properties, and to a lesser extent their applicability and prospects on computational efficiency.

For instance, in the limelight of Chen's relation with the iterated integrals [Che54] - that the signature is a group homomorphism, algebraic properties are sought to carry out calculus on some dual space.
Expected applications include :

- the ability to approximate functions on paths with a set of statistics (universality),
- characterise convergence of measures on paths (characteristicness),
- carry out calculus on a dual space, possibly isometric,
- leverage linear symmetries, global properties, invariances, ...

Setup It is important to detail what is meant by the space of paths in $\mathbb{R}^n$ as well as what class of functions or measures that is to study.
The general case will be referred to as follows :

Let V be a Banach space and $\mathcal{C}$ the space of unparameterized, tree-reduced paths on V

For instance :

$$\mathcal{C} \xrightarrow{\quad f \quad} \mathbb{C}$$

Where f is continuous or lies in some L^p, including the case of probability measures.
Note that continuity and the definition of a measure on $\mathcal{C}$ depends on the choice of topology [CT24b].

$$(\mathcal{C}, \tau) \xrightarrow{\quad \mu \quad} \mathbb{C}$$

Given the known correspondence between paths and their signature with the product topology, the harmonic decomposition of f is expected to be some graded, formal series of non-commuting operators - or a transformation of these.

0.2 Intuition walk-through

The main results of Fourier analysis on the real line or the circle group is that the dual space is isometric and that the Fourier series are dense in the original L^2 space [Wil95]. In this sense a function with the right properties can be decomposed as a sum of sines and cosines. The spectral interpretation of the dual group yields a decomposition that leverages the periodic symmetries of a signal or function.

This can be extended canonically to Euclidean spaces $\mathbb{R}^n$, with :

$$\forall u \in \mathbb{R}^n,\ \mathcal{F}(f)(u) = \iiint_{s\in\mathbb{R}^n} f(s)e^{i\langle u,s\rangle}ds$$

An identical idea for the inverse works as long as values are well defined (integrability). The ideal case takes place in the Scharz space, stable under Fourier transform, which is dense in L^2 and can be orthogonalised, paving the way for functional analysis tricks.

Another application of Harmonic Analysis is to study PDEs - like solving the Laplace equation on a disk with singularities - or solve integral equations [AB12].

Some of the main applications have been found in signal processing for filtering and denoising, image processing for pattern recognition, signal encoding for dimensionality reduction, solving linear differential equations. To give one precise example, in bio-imaging the spectrogram (spectral decomposition of a signal in time) underpins brain and heart diagnosis. In many cases, a spectral decomposition is only a building block for complex models (in the case of harmful brain activity detection the spectrogram can be processed and seen as an image, with which one can leverage computer vision architectures like CNNs) - the spectogram can be seen as a linear lift in frequency domain.

Ramping up the Fourier toolbox to path spaces Real Fourier analysis is underpinned by the exponential function, that seen as a basis with different frequencies, can separate functions. These support functions will be called *characters* in the commutative case. The idea being to have some frequency-specific basis function defined globally - in some L^p sense, as opposed to locally for wavelets (as another approach at the cross-section of harmonic analysis and non-parametric statistics).

The foundational attractiveness of Fourier analysis could be ideally leveraged spatially on path spaces in a similar way.

Why is the loss of commutativity a big deal ? Non-commutativity of the underlying algebra is problematic in the usual sense of Fourier transform, starting from the definition of exp as group homomorphism :

$$e : (\mathbb{R}, +) \longrightarrow \left(\mathbb{R}_+^*, \times\right)$$

or :

$$e : (\mathbb{C}, +) \longrightarrow (\mathbb{C}^*, \times)$$

Indeed, defining exp as the exponential formal series, and taking two non-commutating elements S and T :

$$\exp(T+S) = \sum \frac{(T+S)^k}{k!} \neq \exp(T)\exp(S) \neq \exp(S)\exp(T)$$

So the Fourier transform can't be defined as usual, not to mention the inverse transform. To avoid confusion, it is convenient to view $\exp(S)$ and $\exp(T)$ as "free" words, which can be combined formally into sequences.

From then on, the Fourier series can't rely on a decomposition of sines and cosines. Instead, the theory of representation decomposes functions into matrix coefficients of irreducible representations [Dei05], effectively acting as non-commuting "characters".

Define a path-wise exponential instead Since $\exp(T)$ is formally always well defined (as a formal series), and that any path in $\mathbb{R}^n$ can be approximated in the limit by piece-wise constant paths, lifting the exponential of $S+T$ in the non-commutative domain can be thought path-wise in the limit [Kap09]:

$$E_\gamma(S,T) = \lim_N (X_N, Y_N)^{\gamma_N} \in \mathbb{C}\langle\langle S,T\rangle\rangle$$

Where $X_N = E(S/N)$ and $Y_N = E(T/N)$ which are well defined, γ_N is an approximation of γ which is piece-wise constant (N steps in each direction) and $\mathbb{C}\langle\langle S,T\rangle\rangle$ the space of formal series in two non-commuting variables, which is isomorphic to the tensor algebra $T((\mathbb{C}^2))$.

This limiting approach in [Kap09] leads to the definition of a holonomy, which at this point restricts the study to piecewise-continuous paths. We'd like however to at least include Stochastic Differential Equations driven by brownian paths in the study.

A better and more direct approach is to consider that the signature of a path is solution to an exponential-like equation in the tensor algebra :

Property 0.1 (Signature as exponential).
Let $x \in \mathcal{C}$. Given any parameterization of x, $t \mapsto S(x_t)$ is well-defined and solves the following Controlled Differential Equation (CDE) :

$$\begin{cases} dS_t = S_t \otimes dx_t \\ S_0 = 1 \in T((V)) \end{cases}$$

The signature is not the only exponential-like lift of the space of paths.

Property 0.2 (Path development as exponential).
Let $x \in \mathcal{C}$. Given any parameterization of x, $t \mapsto S(x_t)$ is well-defined and solves the following Controlled Differential Equation (CDE) :

$$\begin{cases} dZ_t = Z_t M(dx_t) \\ Z_0 = I \in G \end{cases}$$

0.3 Findings and prospects in brief

Notation	Abelian	Name	Topological group
$b \quad : \quad \mathbb{Z}^n \longrightarrow \mathbb{C}$	Yes	Linear paths	discrete
$b \quad : \quad \mathbb{F}_n \longrightarrow \mathbb{C}$	No	Piecewise-constant	discrete
$b \quad : \quad \Pi_n \longrightarrow \mathbb{C}$	No	$\mathbb{R}^n$-valued continuous tree-reduced	non locally compact

Figure 1: Table of path-like groups

Discrete topological paths-like groups have a canonical harmonic decomposition, which will be shown for linear paths and piecewise-constant paths (table 1).

The most interesting case is that of $\mathbb{R}^n$-valued, unparameterized, continuous tree-reduced paths, denoted Π_n. It's topological structure is a lot more complex and lies at the fringe of well-studied harmonic analysis on groups.
The main properties that will come up in the following sections are the following :

Paths Fourier Transforms It can be defined whether as the expected signature in the tensor algebra, or as a collection of irreducible representations on matrix Lie groups.

Closed-form formula A known result is that the expected signature (in the Stratonovich sense) of standard Brownian motion is the tensor equivalent of the Wiener measure :

$$\mathbb{E}\left[S(x)\right] = \exp\left(-\sum_{j=1}^{n} e_j^{\otimes 2}\right)$$

Similar closed-form formula for path-developments are not known and would be parameter-dependant unlike the signature.

Algebraic properties The transform of the convolution of measures is the product of the transform of each measure :

$$\mathcal{F}(\mu * \nu) = \mathcal{F}(\mu)\mathcal{F}(\nu)$$

which holds for the various definitions on Fourier transform that will be presented.

Weak convergence of measures The expected signature characterises functions in the limit of compact support, but can be extended with a scaling trick [CO18]. The characteristic function in [CL16] is characteristic and universal beyond the case of compact support.

Log-signature characteristicness Taking as example a simple class of diffusion processes, classic harmonic analysis method can be applied to characterise their distribution - but it is not the case in general.

0.4 Outline

To work the way up to the findings, basis of harmonic analysis will be explored [1]

(1.) Firstly, with commutative topological groups [Dei05], whose Fourier theory is, except technical considerations, naturally akin to real Fourier analysis.

Following is the study of non-commutative harmonic analysis on compact and locally compact groups [DE14], which relies on representation theory. Exemplification will be detailed with matrix groups. A focus on the Heisenberg group is taken, arguably the simplest example of locally compact matrix group, as it will appear in section 3.2 in analysing nilpotent diffusion processes.

(2.) Taking intuition on discrete path spaces to picture the added difficulties, the different topologies for continuous paths are reviewed with their geometric properties. It will pave the way for ramping up the study of spaces of paths in Π_n. Stochastic differential equation notations are introduced and motivated.

(3.) Next, Fourier analysis based on the signature is detailed with it's limitations. It can be thought of as a the moment generating function in the probabilistic case [CO18]. Besides, the log-signature at level two fully characterises a class of diffusion processes on nilpotent Lie groups that includes the Heisenberg group. Thus the harmonic dual from (1.) is applied to the process distribution.

(4.) Eventually path-development [LLN22] and related characteristic function on matrix Lie groups [CL16] will be explored, comparing approaches which are different sides of the same coin. An numerical simulation is detailed, in the context of characterising distributions, using theory from (1.).

A source of originality of this thesis is to apply harmonic analysis tools on compact matrix Lie groups being the image of the path development embedding, and to exemplify the Heisenberg group decomposition on 2-step nilpotent diffusion processes [2].

[1] TL;DR readers can focus on sections 3 and 4., indeed, the review on commutative harmonic analysis is a reminder and details of path topologies are only used to ensure continuity - other topologies are not compared.

[2] Some detours have been taken as an exercise and to show understanding. Expanding more on each would have provided more interpretation to the thesis at the cost of it's length and readability

1 Review of Harmonic analysis on groups

Extending the Fourier analysis on the real line has been widely studied and is the core of abstract harmonic analysis. The following review is largely taken from [Dei05] and [DE14]. Insights on the Heisenberg group are taken from [Pan82].

Definition 1.1. *A topological group is a group G with a topology, such that the group operations of multiplication and inversion are continuous.*

Definition 1.2. *A topological group is a locally compact group (LC) if the underlying topological space is locally compact and Hausdorff. A topological group is abelian if the underlying group is abelian (LCA).*

Examples of locally compact abelian groups include finite abelian groups, the integers (both for the discrete topology, which is also induced by the usual metric), the real numbers, the circle group $\mathbb{T}$ (both with their usual metric topology), and also the p-adic numbers (with their usual p-adic topology).

1.1 LCA case

Let G be a locally-compact abelian group bestowed with it's Haar measure with unit total mass [Dei05] (see A.2.1 on page 44 for the definition of the Haar measure).

1.1.1 Pontryagin duality

The dual of G is defined as the group of continuous group homomorphism between itself and the circle group :

$$\hat{G} = \mathrm{Hom}(G, \mathbb{T})$$

or more precisely :

$$\hat{G} = (\{\mathrm{hom} : G \to \mathbb{T},\ \text{continuous group homomorphism wrt } \tau\}, .)$$

The set of all characters $\hat{G}$, with pointwise product of functions and the topology given by uniform convergence on compact sets, is also an LCA group [Dei05] and is called the *Pontryagin dual* of G.

Theorem 1.3 (Pontryagin duality)**.** *Any LCA group G is naturally isomorphic to its bidual : $G \cong \hat{\hat{G}}$*

The Fourier inversion formula is a special case of this theorem.
Elements of the bidual of G are represented as evaluation functions :

$$\begin{aligned} \delta_a : \hat{G} &\longrightarrow \mathbb{T} \\ \chi &\longmapsto \chi(a) \end{aligned}$$

and the isomorphism can be defined as : $a \in G \mapsto \delta_a$.

In the following examples, the harmonic analysis results correspond to the Fourier theory on real functions. The Borel topology is a natural choice as being a branch of real harmonic analysis. The Pontryagin duality theorem is already known is those cases.

Example 1.4 ($G = (\mathbb{R}/\mathbb{Z}, +)$)**.**

- *The Haar measure is chosen as $\lambda/2\pi$ with λ the Lebesgue measure.*
- *The dual space is composed of Fourier coefficients :*

$$\hat{G} = Hom(G, \mathbb{T}) = \left\{x \mapsto e^{2\pi i x k} / k \in \mathbb{Z}\right\} \cong \mathbb{Z}$$

- *The harmonic transform corresponds to the Fourier coefficients of the periodic function f :*

$$c_k(f) = \frac{1}{2\pi} \int_{\mathbb{R}} f(x) e^{-2i\pi k x} dx = \hat{f}(\chi_k)$$

Example 1.5 ($G = (\mathbb{R}, +)$)**.**

- *The Lesbesgue measure is taken as the Haar measure.*
- *The dual space is composed of Fourier coefficients :*

$$\hat{G} = Hom(G, \mathbb{T}) = \left\{x \mapsto e^{2\pi i x y} / y \in \mathbb{R}\right\} \cong \mathbb{R}$$

- *Pontryagin dual elements relate to the Fourier transform of f via :*

$$\mathcal{F}f(u) = \int_{\mathbb{R}} f(x) e^{-2i\pi u x} dx = \hat{f}(\chi_u)$$

A main result in Fourier analysis is to have a decomposition of the original space in terms of elements of the dual. Similarly, the completeness of Fourier series in LCA groups is the consequence of the following theorems, akin to the usual Fourier analysis on the real line.

Completeness of Fourier Series Every function is the limit of it's inverse Fourier transform in L^2.

Lemma 1.6 (Bessel, LCA groups)**.** *Let G be an LCA group. Let K_n be an absorbing exhaustion of $\hat{G}$ and let $f \in L^2(G)$. In the same context, a Fourier series for $f \in L^2(G)$ can be defined.*

$$S_n(f) = \int_{K_n} \chi(f) \delta_{.}(\chi) d\chi$$

it belongs in $L^2(G)$ and :

$$||f - S_n(f)||_2^2 = ||f||_2^2 - ||S_n(f)||_2^2$$

Proof. Given $L^2(G)$ still is a Hilbert space, one has :

$$||f - S_n(f)||_2^2 = ||f||_2^2 - 2\langle f, S_n(f)\rangle + ||S_n(f)||_2^2$$

Then :

$$\begin{aligned}
\langle f, S_n(f)\rangle &= \int_G f(x) \overline{\int_{K_n} \chi(f) \delta_x(\chi) d\chi} dx \\
&= \int_{K_n} \overline{\chi(f)} \int_G f(x) \overline{\chi(x)} dx d\chi \\
&= \int_{K_n} \overline{\chi(f)} \chi(f) d\chi \\
&= \int_{K_n} |\chi(f)|^2 d\chi \\
&= ||S_n(f)||^2_{L^2(G)}
\end{aligned}$$

Where Fubini is assumed to hold to obtain the second equation. □

Therefore, when it exists, the inverse Fourier transform is given by :

$$S(f) : x \mapsto \int_{\chi \in \hat{G}} \chi(f) \delta_x(\chi) d\chi$$

Eventually, the Plancherel theorem can be extended to LCA groups.

Theorem 1.7 (Plancherel)**.** *There exist a unique Haar measure such that for any $f \in L^1_{bc}(G)$:*

$$||f||_{L^2(G)} = ||\hat{f}||_{L^2(\hat{G})}$$

That is the Fourier transform of f exists in L^2. It extends to a Hilbert space isomorphism

$$L^2(G) \to L^2(\hat{G})$$

and :

$$\forall \eta, \chi \in \hat{G}, \int_G \chi(x) \overline{\eta(x)} dx = \begin{cases} 1 \ \mathit{if} \quad \chi = \eta \\ 0 \ \mathit{otherwise} \end{cases}$$

Completeness of the Fourier series is provided by Bessel's lemma.

As the locally compact abelian case encompasses real Fourier analysis, it provides all the nice tools and properties that put the theory in the limelight of analysis. Loosing commutativity requires a different approach : using group representations.
Non-abelian topological groups include matrices, operators and path groups.

Having a compact group instead of locally compact gives rise to a thorough decomposition onto discrete "matrix" coefficients and is the main result of the following sub-section.

1.2 Compact case

Definition 1.8 (Representation). *Let G be a topological group. For a Hilbert space V, a representation of the group G on V is a continuous group homomorphism :*

$$\pi : G \to GL(V)$$

Such that the action :

$$\begin{aligned} G \times V &\longrightarrow V \\ (g, x) &\longmapsto \pi(g)x \end{aligned}$$

is continuous, where $GL(V)$ is the general linear group.

Definition 1.9 (Invariant subspace). *An invariant subspace $W \subset V$ is such that $\tau(x)W \subset W, \forall x \in G$. It is proper if it is neither empty nor the full space V.*

Given a topological group G and a representation (τ, V_τ):

- τ is *unitary* if $\langle \tau(x)v, \tau(x)w \rangle = \langle v, w \rangle, \quad \forall x \in G, \ v, w \in V$
- (τ, V_τ) *irreducible* if there are no proper closed invariant subspaces

Example 1.10 (Unitary representation on $\mathbb{C}^n$). *$U(n) < GL_n(\mathbb{C})$ as a closed matrix group and $\mathbb{C}^n$ is a n-dimensional complex Hilbert space. Matrix multiplication is continuous, therefore the identity mapping $id_{GL_n(\mathbb{C}}$ is a representation of the unitary group on $\mathbb{C}^n$.*

Let K be a compact group and $\hat{K}$ be the set of equivalence classes of irreducible unitary representations of K. The gist of the following part is the Peter-Weyl theorem for compact groups, and it generalises this duality property on abelian groups :

Property 1.11 (Compact dual). *If G is an abelian discrete group, it's dual group is compact and the converse is true : the dual group of a compact abelian group is discrete.*

1.2.1 Peter-Weyl decomposition

Property 1.12. *The dual space of a compact group reduces to finite-dimensional irreducible unitary representations :*

$$\hat{K} = \hat{K}_{finite}$$

Matrix coefficients For every class in $\hat{K}$, choose one representative (τ, V_τ). Choose also an orthonormal basis $(e_i)_{i=1...n}$ of V_τ.
A *matrix coefficient* for the representation τ is defined as :

$$\tau_{ij}(k) = \langle \tau(k)e_i, e_j \rangle$$

As $\tau : K \to V_\tau$ is continuous and the inner product on V_τ is bilinear, matrix coefficients are continuous and lie in $L^2(K)$.
They define an element in $GL(V_\tau) \cong GL_n(\mathbb{R})$ hence their name.

The following result defines the isomorphism between the space of functions on K and it's dual in terms of matrix coefficients.

Theorem 1.13 (Peter-Weyl).
The harmonic dual of $L^2(K)$ is :

$$L^2(K) \cong \widehat{\bigoplus}_{\tau \in \hat{K}_{fin}} End(V_\tau)$$

Where non-equivalent representations are orthogonal. The isomorphism is explicitly defined as :

$$L^2(K) \longrightarrow \widehat{\bigoplus}_{\tau \in \hat{K}_{fin}} End(V_\tau)$$
$$f \longmapsto \sum \tau(f)$$

Where :

$$\tau(f) = \int_K f(x)\tau(x)dx$$

as a Bochner integral in $\mathcal{B}(V_\tau)$

When applied to the unit circle, one finds the usual discrete Fourier decomposition. Indeed as $\mathbb{T}$ is abelian, all irreducible representations are one-dimensional.

Unitary representations generalise the Pontryagin dual

Theorem 1.14 (Schur). *Let (π, V_π) be a unitary representation on the locally compact group G. The following statements are equivalent :*

- *π is irreducible*
- *If T is a bounded operator on V_π such that $\forall g \in G, T\pi(g) = \pi(g)T$, then $T \in \mathbb{C}Id$*

Property 1.15. *An irreducible unitary representation on a LCA group G is one dimensional.*

Proof. [DE14] Let G be an LCA group. For any unitary and irreducible representation π of G on some Hilbert space V_π, one has :

$$\forall g, h \in G,\ \pi(g)\pi(h) = \pi(gh) = \pi(hg) = \pi(h)\pi(g)$$

Fix $h \in G$. $\pi(h)$ is a bounded operator on V_π and commutes with all $\pi(g), g \in G$.
By Schur's Lemma (1.14), $\pi(h) \in \mathbb{C}Id$ for all $h \in G$. In this case, every closed subspace of dimension greater than one is invariant.
Given the representation is irreducible, the dimension of V_π must be one, so V_π coincides with $\mathbb{C}$ and $\pi(h) \in \mathrm{GL}(\mathbb{C}) = \mathbb{C}^\times$.
As the representation is unitary :

$$\pi(h)x\overline{\pi(h)y} = \langle \pi(h)x, \pi(h)y\rangle = \langle x, y\rangle = x\bar{y} \Rightarrow \pi(h)\overline{\pi(h)} = 1,\ \forall h \in G$$

So it coincides with a character (it lies in $\mathbb{T}$ and is a topological homomorphism). □

Eventually the dual of G as defined with representation theory coincides with the Pontryagin dual of G as an LCA group.

1.3 The Heisenberg group

Definition 1.16. *A topological group is called locally compact if it is Hausdorff and locally compact in the sense that every point has a compact neighbourhood.*

The most general form of Plancherel theorem applies in the locally compact case. As characters are generalised to the non-abelian domain with irreducible unitary representations, the center of a locally compact group G is treated differently, hence it's importance.

The Heisenberg group will be a case study to exemplify this section. As it resembles a lot to $\mathbb{R}^3$ [Pan82], the structure of it's dual is characterised in a familiar way. It is relevant in quantum mechanics in Weyl's parameterisation to carry out calculus on operators. Other notorious locally compact matrix groups include $GL_n(\mathbb{C})$ or the group of upper-triangular matrices.

1.3.1 Description of the group structure

$$\mathcal{H} = \left\{ \begin{pmatrix} 1 & a & c \\ 0 & 1 & b \\ 0 & 0 & 1 \end{pmatrix}, \quad (a,b,c) \in \mathbb{R}^3 \right\} \leq \mathrm{Sl}_3(\mathbb{R})$$

Denote :

$$H(a,b,c) = \begin{pmatrix} 1 & a & c \\ 0 & 1 & b \\ 0 & 0 & 1 \end{pmatrix}$$

The space is isomorphic to $\mathbb{R}^3$ equipped with the group structure defined as :

$$\begin{cases} H(a,b,c)H(\alpha,\beta,\gamma) = H(a+\alpha, b+\beta, c+\gamma+a\beta) \\ H(a,b,c)^{-1} = H(-a,-b,-c+ab) \\ e = H(0,0,0) \end{cases}$$

H_3 is closed therefore it is a matrix Lie group.
It is a locally-compact non-abelian topological group (with the $\mathbb{R}^9$ product topology on matrices).

Property 1.17 (Haar measure on the Heisenberg group)**.**

$$\int_{h(a,b,c)\in\mathcal{H}_3} f(h(a,b,c))dadbdc$$

defines a Haar measure on the Heisenberg group.

Proof. Let $g \in \mathcal{H}_3$, and $f \in C_c(\mathcal{H}_3)$, then :

$$\begin{aligned} \int_{h\in\mathcal{H}_3} f(gh)dh &= \int_{a,b,c\in\mathbb{R}^3} f(a+\alpha, b+\beta, c+\gamma+\beta a)dadbdc \\ &= \int_{a,b,c\in\mathbb{R}^3} f(a,b,c)dadbdc \\ &= \int_{h\in\mathcal{H}_3} f(h)dh \end{aligned}$$

Where a linear change of variable is applied in the integral.
The outer-Radon property holds given it is expressed in term of Lebesgue measure. □

Besides, the Heisenberg group has generators :

$$\begin{cases} L_1 = I_3 + E_{1,2} \\ L_2 = I_3 + E_{2,3} \\ L_3 = I_3 + E_{1,3} \end{cases}$$

Each of which forms a 1-dimensional closed subgroup of $\mathrm{GL}_3(\mathbb{R})$. Let $\mathfrak{h} = T_{\mathcal{H}}(I_3)$ be it's associated matrix Lie algebra.
Let $V_1 = \mathrm{span}(E_{1,2}, E_{2,3})$ and $V_2 = \mathrm{span}(E_{3,3})$. One has : $\mathfrak{h} = V_1 \oplus V_2$ and $[V_1, V_1] = V_2$.

Definition 1.18 (Carnot group)**.** *A simply connected nilpotent Lie group is called a Carnot group if the subspace with eigenvalue* 1 *generates the Lie algebra.*

$\mathfrak{h}$ endowed with the commutator brackets is nilpotent of order 2. In this sense it is a Carnot group of order 2.

Engel groups are another example of Carnot groups and their properties are still under study [Bah+23]. The center of $\mathcal{H}$ is $Z(\mathcal{H}) = \{H(0,0,c)/c \in \mathbb{R}\}$, and $\mathcal{H}/Z(\mathcal{H}) \cong \mathbb{R}^2$.
It is a closed subgroup of $(GL_3(\mathbb{R}), \times)$ with embedded matrix norm. It is non-commutative and locally compact, but not compact.

The following paragraphs aim at understanding the world in this non-commutative space, in terms of distances and to understand it's symmetries which will give us intuition as to what kind of properties it is better suited to encode.

Geometry of the Heisenberg group Let γ be a smooth path in $\mathcal{H}$ parameterized on $[0,1]$. Then $\gamma' \in \mathfrak{h}$ and one can define the euclidean length of that path as :

$$L(\gamma) = \int_{[0,1]} \sqrt{|(\pi \circ \gamma)'(t)|}dt$$

The Carnot-Caratheodory distance between two points $h, h_0 \in \mathcal{H}$ is then :

$$d_c(h_0, h) = \inf \{L(\gamma), \gamma \in \Gamma(h_0, h)\}$$

Where $\Gamma(h_0, h)$ is the set of parameterized paths on $[0,1]$ joining h_0 and h.

Definition 1.19 (Carnot path)**.** *A path γ is called a Carnot path if it's speed is contained in a plane.*

Writing : $\gamma(t) = \begin{pmatrix} 1 & x(t) & z(t) \\ & 1 & y(t) \\ & & 1 \end{pmatrix}$, one has :

$$ln(\gamma(t))' = \gamma(t)^{-1}\gamma'(t) = \begin{pmatrix} 0 & x'(t) & z'(t) - x(t)y'(t) \\ & 0 & y'(t) \\ & & 0 \end{pmatrix} \in \mathfrak{h}$$

So it is a Carnot path if $z'(t) = x(t)y'(t)$, or $z(t) = z_0 + \int xdy$.

Property 1.20. *$\mathcal{H}$ is accessible with Carnot paths.*

Proof. It relies on the fact that a Carnot path can access the coordinates x and y, both belonging to the speed plane. Then the third coordinate is determined by the area spanned by the two first coordinates, which can be adjusted algebraically to match the third coordinate - taking an arc is sufficient [Pan82]. $\square$

Property 1.21. *There exist positive constants c, C such that for any $h = \begin{pmatrix} 1 & x & z \\ & 1 & y \\ & & 1 \end{pmatrix}$:*

$$c \max \left\{|x|, |y|, \sqrt{|z|}\right\} \leq d_c(I_3, h) \leq C \max \left\{|x|, |y|, \sqrt{|z|}\right\}$$

The unit sphere in $\mathcal{H}$ is flattened in the z direction, in the shape of an apple. Geometrically, it is a lot more costly to travel upward than it is in the (x, y)-plane. This property will help us understand how path embeddings compare in the following section.

The scaling automorphism $\delta_\epsilon : (a, bc) \mapsto (\epsilon a, \epsilon b, \epsilon^2 c)$ acts on the unit sphere in opposite ways in small and large scales : small balls are flattened and conversely large balls are tall in the z direction.

Automorphisms and symmetries In $\mathrm{Sl}_3(\mathbb{R})$, the centralizer of $\mathcal{H}$ - elements g such that $g^{-1}\mathcal{H}g = \mathcal{H}$ is the subgroup of upper triangular matrices, which contains the subgroup of diagonal matrices which is 2-dimensional.

A 1-dimensional subgroup of diagonal matrices that commutes with $\mathcal{H}$ is the group of dilations :

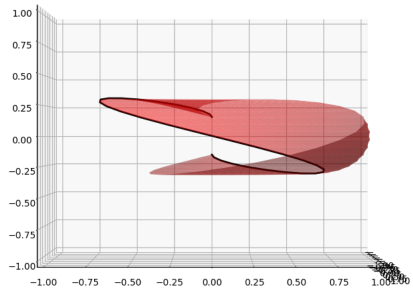
$$\left\{ \delta_\epsilon = \begin{pmatrix} \epsilon & & \\ & 1 & \\ & & \epsilon^{-1} \end{pmatrix} / \epsilon > 0 \right\}$$

Whose action as conjugate on $\mathcal{H}$: $\delta_\epsilon(\gamma)(t) = \delta_\epsilon \gamma(t) \delta_\epsilon^{-1}$, is homogenous in the (x, y)-plane projection. It is also homogenous with the length of a path :

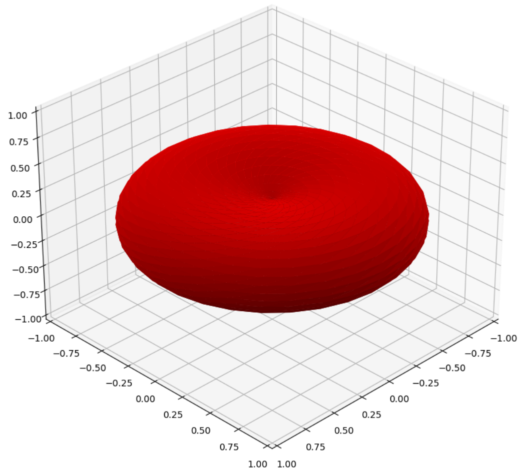
$$L\left(\delta_\epsilon(\gamma)\right) = \epsilon L(\gamma)$$

Hence it is also homogenous in term of Carnot-Caratheodory distance.
The action of general diagonal groups is given by :

$$\Delta_{a,b} \begin{pmatrix} 1 & x & z \\ & 1 & y \\ & & 1 \end{pmatrix} = \begin{pmatrix} 1 & \frac{x}{a^2 b} & \frac{b}{a} z \\ & 1 & ab^2 y \\ & & 1 \end{pmatrix}$$



(a) Horizontal view of a parameterized curve and a slice of the unit sphere



(b) Full unit sphere

Figure 2: Unit sphere of the Heisenberg group in the Carnot-Caratheodory distance

With : $\Delta_{a,b} = \begin{pmatrix} a & & \\ & \frac{1}{ab} & \\ & & b \end{pmatrix}$.
Other automorphisms can be found as a linear bijection that can pass through the Lie bracket :

$$l([x, y]) = [l(x), l(y)]$$

defines an automorphism on the Lie group through the exponential :

$$\exists L,\ L(\exp(x)) = \exp(l(x))$$

Such a linear bijection necessarily preserves the z-axis. Explicitly writing l gives necessary conditions as follows :

$$l \begin{pmatrix} 1 & u & w \\ & 1 & v \\ & & 1 \end{pmatrix} \approx l(u, v, w) = \begin{pmatrix} a & b & g \\ c & d & h \\ e & f & i \end{pmatrix} \begin{pmatrix} u \\ v \\ w \end{pmatrix} = \begin{pmatrix} au + bv + gw \\ cu + dv + hw \\ eu + fv + iw \end{pmatrix}$$

$$l(0, 0, w) \in (0, 0, \mathbb{R}) \Leftrightarrow (g, h) = (0, 0)$$

$$l \in GL_3(\mathbb{R}) \Leftrightarrow ad - bc \neq 0,\ i \neq 0$$

In order to preserve the bracket :

$$l([M, M']) = (0, 0, i(uv' - u'v)) = [l(M), l(M')] = (0, 0, (ad - bc)(uv' - u'v)$$

Hence $i = ad - bc$. Eventually this group of automorphisms has 6 degrees of liberty :

$$l = \begin{pmatrix} a & b & 0 \\ c & d & 0 \\ e & f & ad - bc \end{pmatrix}, ad - bc \neq 0 \Rightarrow L \text{ automorphism}$$

The Heisenberg group has almost as many automorphisms as the commutative group $\mathbb{R}^3$.

To understand what equivalent clusters look like in the Heisenberg geometry, one can be interested to know which of these automorphisms are isometries in the Carnot-Caratheodory distance.

To start with, the left-translation map is such an isometry. The action of diagonal matrices are not isometries in general. Some automorphisms defined from a linear bijection in the Lie algebra are isometries :

Property 1.22. *If l linear bijection on $\mathfrak{h}$ defines an isometry of the Carnot-Caratheodory distance on $\mathcal{H}$ then $l(V_1) = V_1$*

These maps are rotations of axis z and symmetries along an axis of the (x, y)-plane. For $h \in \mathcal{H}$, define hV_1 the left-translation of V_1 by h. The map $K : h \mapsto hV_1$ is called the field of horizontal planes.

Theorem 1.23. *Carnot-Caratheodory isometries are characterised by composing left-translations and isometric isomorphisms.*

1.3.2 The unitary dual of $\mathcal{H}$

Let $\hat{\mathcal{H}}$ be the set of irreducible unitary representations on $\mathcal{H}$. Define :

$$\hat{\mathcal{H}}_0 = \left\{ \pi \in \hat{\mathcal{H}} \ / \ \pi(h) = 1, \ \forall h \in Z(\mathcal{H}) \right\}$$

$$\hat{\mathcal{H}}_0 = \widehat{\mathcal{H}/Z(\mathcal{H})} \cong \widehat{\mathbb{R}^2} \cong \hat{\mathbb{R}}^2 \cong \mathbb{R}^2$$

Elements of $\mathcal{H}/Z(\mathcal{H})$ can be identified as $H(a, b, 0)$, and dual elements can be identified :

$$\begin{aligned} \chi_{a,b} : \mathcal{H} &\longrightarrow \mathbb{T} \\ (a, b, c) &\longmapsto e^{i(ax+by)} \end{aligned}$$

Schur's lemma (1.14 on page 10) then implies that all representations of $Z(\mathcal{H})$ are one-dimensional. It shows the importance of the behaviour of the center under a unitary representation.

Definition 1.24 (central character)**.** *Let π be a unitary representation of locally compact group G. For all $z \in Z(G)$, Schur's lemma ensures that $\pi(z)$ is a multiple of the identity :*

$$\forall \pi \in \hat{G}, \ \exists \chi_\pi : Z(G) \to \mathbb{T} : \pi(z) = \chi_\pi(z) Id_V, \forall z \in Z(G)$$

χ_π is called the central character of the representation π.

Then for a character $\chi \neq 1$ of $Z(\mathcal{H})$, one can construct a unitary representation of $\mathcal{H}$ that has χ for central character. Take as reference the character : $\pi_1(0, 0, c) = e^{2\pi i c}$.
The group of unitary operators on $L^2(\mathbb{R})$ generated by :

$$\phi(x) \mapsto \phi(x + a), \quad \phi(x) \mapsto e^{2\pi i b x}\phi(x), \quad \pi_1(0, 0, c)$$

Is isomorphic to $\mathcal{H}$.
On $L^2(\mathbb{R})$ define the operator $\pi_1(a, b, c)$:

$$\pi_1(a, b, c)\phi(x) = e^{2\pi i(bx+c)}\phi(x + a)$$

Property 1.25 (Representations on the Heisenberg group)**.** *π_1 defines an irreducible unitary representation on $\mathcal{H}$:*

$$\pi_1(a, b, c)\pi_1(\alpha, \beta, \gamma)\phi(x) = \pi_1((a, b, c).(\alpha, \beta, \gamma))\phi(x)$$

Proof. Let (a, b, c) and (α, β, γ) seen as elements in the Heisenberg group. Let $\phi \in L^2(\mathbb{R})$ and $x \in \mathbb{R}$.

$$\begin{aligned} \pi_1(a, b, c)\pi_1(\alpha, \beta, \gamma)\phi(x) &= \pi_1(a, b, c) \left[e^{2\pi i(\beta . + \gamma)}\phi(. + \alpha) \right](x) \\ &= e^{2\pi i((b+\beta)x + c + \gamma + \beta a)}\phi(x + a + \alpha) \\ &= \pi_1 \left((a + \alpha, b + \beta, c + \gamma + \beta a) \right) \phi(x) \\ &= \pi_1 \left((a, b, c)(\alpha, \beta, \gamma) \right) \phi(x) \end{aligned}$$

Hence π_1 is a group morphism from the Heisenberg group to the group of invertible linear operators in $L^2(\mathbb{R})$.
To conclude that it is a representation, one needs to show the group action is continuous.

Let $\epsilon > 0$, $(a, b, c) \in H$ and $(\delta a, \delta b, \delta c) \in H$ and a neighbourhood of the unit. Let also $\phi \in L^2(\mathbb{R})$ and $\delta\phi \in B_{L^2(\mathbb{R})}(0, \epsilon)$.

$$\begin{aligned}||\pi_1(a+\delta a, b+\delta b, c+\delta c)(\phi+\delta\phi) - \pi_1(a,b,c)\phi||_{L^2} &\leq || \left(\pi_1(a+\delta a, b+\delta b, c+\delta c) - \pi_1(a,b,c)\right)\phi|| \\ &\quad + ||\pi_1(a+\delta a, b+\delta b, c+\delta c)\delta\phi|| \\ &\leq || \left(\pi_1(a+\delta a, b+\delta b, c+\delta c) - \pi_1(a,b,c)\right)\phi|| \\ &\quad + ||\delta\phi||\end{aligned}$$

Given π_1 is an isometry on $L^2(\mathbb{R})$. As $C_c(\mathbb{R})$ is dense in $L^2(\mathbb{R})$, let $\phi_c \in L^2(\mathbb{R})$ such that $||\phi - \phi_c||_{L^2} \leq \epsilon$.

$$\begin{aligned}||\pi_1(a+\delta a, b+\delta b, c+\delta c)\phi - \pi_1(a,b,c)\phi|| &= ||\pi_1(\delta a, \delta b, \delta c + a\delta b)\phi - \phi|| \quad \text{(direct calculation)} \\ &\leq \; ||\phi(x+\delta a) - \phi_c(x+\delta a)|| \\ &\quad + ||e^{2\pi i(\delta bx + \delta c + a\delta b)}\phi_c(x+a) - \phi_c(x+a)|| \\ &\quad + ||\phi_c(x+a) - \phi_c(x)|| \\ &\quad + ||\phi_c(x) - \phi(x)||\end{aligned}$$

The first and last term are smaller than ϵ by assumption. Then, as ϕ_c is continuous on it's compact support K, it is uniformly continuous.

$$\exists \delta > 0, \delta a < \delta \Rightarrow ||\phi_c(. + \delta a) - \phi_c||_\infty < \frac{\epsilon}{2}$$

Taking at the same time for δa small enough :

$$\int_{\mathbb{R}-K} |\phi_c(x+a) - \phi_c(x)| dx < \frac{\epsilon}{2}$$

One controls the term :

$$||\phi_c(x+a) - \phi_c(x)|| < \epsilon$$

And for the last term :

$$\begin{aligned}||e^{2\pi i(\delta bx+\delta c+a\delta b)}\phi_c(x+a) - \phi_c(x+a)||_{L^2(\mathbb{R})} &= ||e^{2\pi i(\delta bx+\delta c+a\delta b)}\phi_c(x+a) - \phi_c(x+a)||_{L^2(K)} \\ &\leq ||e^{2\pi i(\delta bx+\delta c+a\delta b)} - 1||_{L^2(K)}.||\phi_c||_{L^2(K)} \\ &\quad \text{(Schwarz inequality)}\end{aligned}$$

With the first term converging uniformly to 0 on the compact set K so is also controlled by ϵ.
Gathering all inequalities :

$$||\pi_1(a+\delta a, b+\delta b, c+\delta c)(\phi+\delta\phi) - \pi_1(a,b,c)\phi||_{L^2(\mathbb{R})} \leq \epsilon$$

Eventually let $(a, b, c) \in H$ and $\phi, \psi \in L^2(\mathbb{R})$.

$$\begin{aligned}\langle \pi_1(a,b,c)\phi, \pi_1(a,b,c)\psi\rangle_{L^2(\mathbb{R})} &= \int_{\mathbb{R}} e^{2\pi i(bx+c)}\phi(x+a)\overline{e^{2\pi i(bx+c)}}\,\overline{\psi(x+a)}dx \\ &= \int_{\mathbb{R}} \phi(x+a)\overline{\psi(x+a)}dx \\ &= \langle \phi, \psi\rangle\end{aligned}$$

Therefore it is unitary. □

Define the shift map θ_t as $\theta_t(a, b, c) = (a, tb, tc)$. Then

$$\pi_t = \pi_1 \circ \theta_t$$

is also an irreducible representation, with $\chi_t = \chi_1 \circ \theta_t$ as central character. $\chi_t(0, 0, c) = \chi_1(0, 0, ct) = e^{2\pi ict}$

In summary π_t is unique up to isomorphism, and is a unitary representation of $\mathcal{H}$ with central character χ_t [Tay86].
It it expressed as :

$$\pi_t(a,b,c)\phi(x) = e^{2\pi i(bx+c)t}\phi(x+a)$$

Theorem 1.26 (Stone-Von Neuman)**.** *The unitary dual of $\mathcal{H}$ is :*

$$\hat{\mathcal{H}} = \hat{\mathbb{R}}^2 \cup \left\{\pi_t / t \in \mathbb{R}^\times\right\}$$

Let G be a locally compact group and $f \in \mathcal{C}_c(G)$. Fix the Haar measure on G. For a unitary representation (π, V_π) of G, define formally the operator :

$$\pi(f) = \int_G f(x)\pi(x)dx$$

Then $\forall v, w \in G$,

$$\langle \pi(f)v, w\rangle = \int_G \langle \pi(x)v, w\rangle dx$$

Property 1.27.

$$||\pi(f)|| \leq ||f||_1 = \int_G |f(x)|dx$$

In the case of the Heisenberg group $\mathcal{H} \cong \mathbb{R}^3$, the Schwarz space $\mathcal{S}(\mathbb{R}^3)$ is adapted for test functions. More explicitly, the Fourier transform of functions on the Heisenberg group reads :

$$\langle \pi_t(f)\phi, \psi\rangle_{L^2(\mathbb{R})} = \iiint_{\mathbb{R}^3} f(a,b,c) \int_{\mathbb{R}} e^{2\pi i(bx+c)t}\phi(x+a)\overline{\psi(x)}dx \; dadbdc$$

Any basis of $L^2(\mathbb{R})$ can be used to "realise" the action of the operator. An explicit example will be taken in section 3.

Theorem 1.28 (Plancherel)**.** *The unique Haar measure on $\mathcal{H}$ called Plancherel measure, is such that for $f \in \mathcal{S}(\mathcal{H}), \forall t \in \mathbb{R}^\times, \pi_t(f)$ is Hilbert-Schmidt and :*

$$\int_{\mathbb{R}^\times} ||\pi_t(f)||^2_{HS}|t|dt = \int_{\mathcal{H}} ||f(h)||^2 dh$$

Property 1.29 (Orthonormality of characters)**.** *If $\chi \neq \eta$, $\langle \chi, \eta\rangle = 0$*

Proof. From [Dei05] :
Let χ, η be two characters.
It is immediate that if $\chi = \eta$, then $\langle \chi, \eta\rangle = 1$ provided the Haar measure is chosen such that $\int_G 1dx = 1$.
Then assume $\eta \neq \chi$. $\exists a \in G$, $\eta(a) \neq \chi(a)$.
Let $\alpha = \chi\overline{\eta} = \chi\eta^{-1}$; $\alpha(a) \neq 1$.

$$\begin{aligned}\alpha(a)\int_G \alpha(x)dx &= \int_G \alpha(a)\alpha(x)dx \\ &= \int_G \alpha(x)dx\end{aligned}$$

Due to the translation invariance of the Haar measure. Then

$$(1-\alpha(a))\int_G \alpha(x)dx = 0$$

Implies that :

$$\int_G \chi(x)\overline{\eta(x)}dx = 0$$

□

2 Path spaces

One interest to study paths as a whole is to unify the continuous-time theory with discrete-time sequences by linearly interpolating data points. However their topology is not trivial as depending on the choice it might not be locally compact and thus extra care is needed to ensure functions are continuous.

An important class of paths is that of solutions to SDEs driven by brownian motion. The power of the following sections is that they extend well to more general classes of paths like fractional Brownian motions or geometric rough paths.

2.1 Discrete groups of paths : $\mathbb{Z}^n, \mathbb{F}^n$

This section exemplifies some difficulties of going from the commutative discrete group $\mathbb{Z}^n$ to non-commutative discrete paths $\mathbb{F}^n$ as detailed in introduction of [Kap09], see table 1.
$\mathbb{F}^n$ is defined as paths starting at zero, which can be decomposed into unit length paths in the direction of one coordinate of the canonical basis of $\mathbb{R}^n$.

Example 2.1 ($\mathbb{Z}^n$). *$a : \mathbb{Z}^n \to \mathbb{C}$, then :*

$$\hat{a} : \mathbb{T}^n \to \mathbb{C} : \chi \mapsto \hat{a}(\chi_t) = \int_{\mathbb{Z}^n} a(x)\overline{\chi_t(x)} d\chi_t = \sum_{x \in \mathbb{Z}^n} a(x) e^{2\pi i t x}$$

With :

$$\chi_t : x \mapsto e^{2\pi i x t}$$

Therefore the identification to the Fourier transform is canonical in the space of Laurent series $\mathbb{R}[[X]]$ with :

$$\hat{a} = \sum_{x \in \mathbb{Z}^n} a_x X^x$$

The inverse Fourier transform is given by :

$$a(x) = \int_{\chi \in \widehat{\mathbb{Z}^n}} \hat{a}(\chi)\chi(x) d\chi$$

Indeed :

$$\begin{aligned}
\int_{\chi \in \widehat{\mathbb{Z}^n}} \hat{a}(\chi)\chi(x) d\chi &= \int_{t \in \mathbb{T}^n} \hat{a}(\chi_t)\chi_t(x) dt \\
&= \int_{t \in \mathbb{T}^n} \int_{y \in \mathbb{Z}^n} a(y)\overline{\chi_t(y)} dy \chi_t(x) dt \\
&= \int_{t \in \mathbb{T}^n} \sum_{y \in \mathbb{Z}^n} a(y)\overline{\chi_t(y)}\chi_t(x) dt \\
&= \sum_{y \in \mathbb{Z}^n} a(y) \int_{t \in \mathbb{T}^n} \overline{\chi_t(y)}\chi_t(x) dt \\
&= \sum_{y \in \mathbb{Z}^n} a(y) \int_{t \in \mathbb{T}^n} e^{2\pi i (x-y)} dt \\
&= a(x)
\end{aligned}$$

Example 2.2 ($\mathbb{F}^n$). *Take $b : \mathbb{F}_n \to \mathbb{C}$.*
$\mathbb{F}_n$ is a discrete group so the Fourier transform can still formally be expressed foramlly as :

$$\hat{b} = \sum_{\gamma \in \mathbb{F}_n} b_x X^\gamma$$

where $X = (X_0, \ldots, X_n)$ are non-commutative variables. It ceases to be a tautology.
Indeed the inversion formula in this case is given by [Voi91] :

$$b_\gamma = \lim_{N \to \infty} tr \int_{X_1, \ldots, X_N \in U(N)} b(X_1, \ldots, X_n) X^{-\gamma} d^* X_1 \ldots d^* X_n$$

Eventually, when the group is continuous, the choice of topology is not straightforward.

2.2 Paths in $\mathbb{R}^n$

Let C be the group of finite variation paths in $\mathbb{R}^n$. The p-variation norm of a path γ is :

$$||\gamma||_p = \left(\sup_{\sigma \subset [0,1]} \sum ||\gamma_{t_{i+1}} - \gamma_{t_i}||^p \right)^{\frac{1}{p}}$$

Then, the tree-like equivalence relation $\sim_\tau$ [CT24b] defines the space of unparameterised, tree-reduced paths $\mathcal{C}$, and makes the signature map $\mathcal{C} \to T((V))$ injective. To complete the group structure, for a given path, the inverse is taken as the reversed path, and the group operation is that of concatenation. A definition of this group as equivalence classes on the space of finite 1-variation paths can be found in these lecture notes [CS24].

Choice of topology General definitions of topological spaces are reminded in Appendix (see A.2 on page 44). As detailed in [CT24b], different topologies can be defined on the space of finite 1-variation unparameterized tree-reduced paths $\mathcal{C}$. The weaker topology that makes each graded canonical projection of the signature continuous (henceforth linear functionals are continuous too) is the product topology on the tensor algebra, where the signature transform is seen as a topological embedding $S : \mathcal{C} \to T((V))$.

For a detailed introduction on the signature transform we refer again to [CS24].

Definition 2.3 (Signature transform)**.** *Given a parameterized path x with values in $\mathbb{R}^n$, the signature of x is the unique solution to the following controlled differential equation in the tensor algebra $T((\mathbb{R}^n))$:*

$$\begin{cases} dS_t = S_t \otimes dx_t \\ S_0 = \mathbf{1}_{T((\mathbb{R}^n))} \end{cases}$$

It is parameterization-independent, with factorial decay of it's graded coefficients and satisfies Chen's relation. Eventually, it is a topological group morphism from the group of unparameterized paths Π_n to the group of group-like elements existing in the Banach algebra $T((\mathbb{R}^n))$.

In the case of smooth paths it can be defined as as holonomy [Kap09] : let $\gamma \in \Pi_n$,

$$S(\gamma) = P \exp \int_\gamma \Omega \in T((\mathbb{R}^n))$$

which in quantum field terminology is called the Wilson loop functional.

The universal approximation theorem states that, for any compact subset of Π_n, it is possible to find a linear functional F to approximate f arbitrarily close.

$$\begin{tikzcd} \Pi_n \arrow[r, "f"] \arrow[d, dashed, "S"'] & \mathbb{C} \\ T((\mathbb{R}^n)) \arrow[ur, "F"'] & \end{tikzcd}$$

In our case f can be a measure on paths instead.

This topology is separable and Hausdorff, metrisable and σ-compact, but not Baire and therefore neither Polish or locally compact [CT24b].

Property 2.4. *$\forall p \in \mathbb{N}^*, (\mathcal{C}_p, \tau_p)$ is not Baire, therefore it is not Polish.*

The other topologies suggested are stronger therefore neither locally compact, which makes the previous theory of locally compact groups unapplicable. Noticeably, completions are sought to alleviate some issues. For example in the product topology it is defined as :

$$\bar{S} = \left\{ x \in T((V)) / \ \pi_n(x) \in S^{(n)}(\mathcal{C}), \ \forall n \in \mathbb{N} \right\}$$

Property 2.5. *$(\mathcal{C}_1, \tau_{pr})$ is Lusin.*

It entails that the weak topology on $\mathcal{P}(X)$ is metrisable.

Definition 2.6 (Weak topology on $\mathcal{P}(X)$)**.** *Let $\mu \in \mathcal{P}(X)$. An open ball with radius ϵ around μ in the weak topology is defined as :*

$$\{\nu \in \mathcal{P}(X)/\ |\mu(f) - \nu(f)| < \epsilon, \forall f \in \mathcal{C}_b(X)\}$$

It is a first step towards characterising the convergence of measure along with the following theorem.

Theorem 2.7 (Prohorov)**.** *Let X be a Lusin space and $H \subset \mathcal{P}(X)$. H is relatively compact in the weak topology if H is tight :*

$$\forall \epsilon, \exists K \subset X \ \textit{compact}\ , \forall \mu \in H, \mu(K^c) < \epsilon$$

If the space was Polish the condition would also be necessary.

Definition 2.8 (Borel sets of τ_{pr})**.** *Let p_d be the projection map from $T((V))$ to $V^{\otimes d}$. Let $U \subset S^d$ be open. $p_d^{-1}(U)$ is called an open cylinder. The set of such open cylinders form a basis of the projective limit (initial) topology on $\mathcal{S}$. Let $\mathfrak{S}$ be the σ- algebra generated by the open cylindrical sets. Elements of this σ-algebra are the Borel subsets of $\mathcal{S}$.*
Their signature pre-image defines Borel sets in the space of paths.

The product topology also ensures that the group operations are continuous, so that the group of paths is a topological group.

Just as inverting the signature of a path is a hard problem [LX18] [LX17], describing geometrically the unit sphere in the product topology hardly relates the the unit sphere in term of bounded path-length. Indeed, as shown with an example in [CS24], any neighbourhood of the null path is unbounded in length.

Still, the signature embedding lives in a very structured space, which will enable different approaches to harmonic analysis.

Property 2.9. *Let V be a Hilbert space, then $T_H((V)) = \left(\{x \in T((V)),\ ||x|| < \infty\}, ||.||_{T((V))}\right)$ is a Hilbert space and $(T_H((V)), +, \otimes)$ is a Banach algebra.*

Notations between the free tensor algebra $T((V))$ and the Hilbert space $T_H((V))$ can be recovered from context so the distinction might be omitted. The Gelfand theory provides a different appproach than group harmonic analysis. Using the Banach algebra property of the tensor embedding has been less frugal but basic theory and some attempts are detailed in appendix B on page 47.

2.3 SDE preliminaries

An important class of paths is that of solutions of stochastic differential equations (SDE) driven by Brownian motion. A classic 1-dimensional time-homogenous SDE might read :

$$dX_t = b(X_t)dt + \sigma(X_t) \circ dB_t = b(X_t)dt + \sigma(X_t)dB_t + \frac{1}{2}\frac{d}{dx}\sigma(X_t)dt$$

Stratonovich integral matches the Ito integral SDE if σ does not depend on x.

Assuming we are allowed to use Girsanov theorm incorporate the drift in the measure, let us re-write the SDE in integral form with $X_t \in \mathbb{R}^n$ and B a d-dimensional brownian motion, as :

$$X_t^{x_0} = x_0 + \sum_{i=1}^{d} \int_0^t V^i(X_s^{x_0}) \circ dB_s^i$$

where the V^i's are smooth vector fields on $\mathbb{R}^n$.

The advantage of using Stratonovich integration is that Ito's formula expands like normal calculus. Let $f \in C_{bc}^{\infty}(\mathbb{R}^n)$:

$$f(X_t) = f(x_0) + \int_0^t \sum_{i=1}^{d} V^i(f)(X_s) \circ dB_s^i$$

Definition 2.10. *The flow of a stochastic process is the map* Φ^* *:*

$$(\Phi_t^* f)(x_0) = f(X_t^{x_0})$$

The approach in [Bau04] to the signature of a path is to iterate the Stratonovich expansion above :

$$f(X_t) = f(x_0) + \iint_0^t \sum_{1 \le i,j \le d} V^i V^j (f)(X_{s_1}^{x_0}, X_{s_2}^{x_0}) \circ dB_{s_1}^i \circ dB_{s_2}^j$$

Here, the vector fields V^i are seen as derivation operators acting on smooth functions defined in the following way [Bau04] :

$$\forall x \in \mathrm{dom}(V),\ V(x) = \sum_{i=1}^d v_i(x) \frac{\partial}{\partial x_i}$$

If the V's vanish after some finite iteration, the integration has a finite number of terms.

Diffusion on nilpotent groups Only one case-study will be defined as it will be interpreted later when describing the Heisenberg group's link with the signature. Nilpotent Lie group have the advantage that their Lie algebra has finite support.

Definition 2.11 (Free 2-step Carnot group on $\mathbb{R}^d$)**.**

$$\mathbb{G}_2(\mathbb{R}^d) = \left(\mathbb{R}^d \times \mathcal{AS}_d, \circledast\right)$$

Where for $x, y \in \mathbb{R}^d$ *and* $A, B \in \mathcal{AS}_d$*, the group operation is defined as:*

$$(x, A) \circledast (y, B) = \left(x + y, A + B + \frac{1}{2}\left(x_i y_j - x_j y_i\right)_{1 \le i,j \le d}\right)$$

And the inverse is :

$$(x, A)^{-1} = (-x, -A)$$

Definition 2.12 (Left Brownian motion)**.** *A process* $(X_t)_{t \ge 0}$ *in a filtered probability space* $(\Omega, \mathcal{F}, \mathbb{P}, \mathcal{F}_t)$ *with values in a topological group* G *is called a left brownian motion if :*

- *It has continuous paths,*
- *For* $0 \le s < \infty, t,\ X_s^{-1} X_{t+s}$ *is independent of* $\mathcal{F}_s$*,*
- *Increments are stationary : for* $0 \le s < \infty, t,\ X_s^{-1} X_{t+s} \sim \mathcal{N}(0_G, \Sigma)$*.*

Example 2.13. *Let* $(B_t)_{t \ge 0}$ *be a Brownian motion in* $\mathbb{R}^2$*. The lift of the Brownian motion in* $\mathbb{G}_2(\mathbb{R}^2)$ *is given by :*

$$B_t^* = \left(B_t^1, B_t^2, \frac{1}{2} \int_0^t B_s^1 dB_s^2 - B_s^2 dB_s^1\right)$$

Property 2.14. B_t^* *is a left Brownian motion.*

Proof. The continuity of the process flows from the continuity of brownian motion paths and that the integral operator is continuous.
Then :

$$\begin{aligned}
(B_s^*)^{-1} B_{t+s}^* &= \left(B_{t+s}^1 - B_s^1, B_{t+s}^2 - B_s^2, \frac{1}{2} \int_s^{t+s} B_u^1 dB_u^2 - B_u^2 dB_u^1 + \frac{1}{2}(B_{t+s}^1 B_s^2 - B_{t+s}^2 B_s^1)\right) \\
&= \left(B_{t+s}^1 - B_s^1, B_{t+s}^2 - B_s^2, \frac{1}{2} \int_s^{t+s} (B_u^1 - B_s^1) dB_u^2 - (B_u^2 - B_s^2) dB_u^1\right) \\
&\sim \left(B_t^1, B_t^2, \frac{1}{2} \int_0^t B_u^1 dB_u^2 - B_u^2 dB_u^1\right) = B_t^*
\end{aligned}$$

Which is independent of $\mathcal{F}_s$ and stationary (the law does not depend on s). ☐

From stochastics to harmonic analysis The usual interest in stochastic analysis is to approximate the flow and infer properties on the final distribution. Besides, the signature transform is a parameter-free method to vectorize a path (at some level of truncation in practice).

In the context of harmonic analysis, the interest is rather to characterize a given distribution. The idea is therefore to vectorize a measure on paths, to be able to compare different measures and hopefully infer weak properties, or independance of variables.

From the lift of a random path into a vector to the characteristic graded vector of a measure, a natural step is to take the expectation. It is exactly what is done in [Bau04], [Kap09], [CO18] and [CL16].

The aim of the next section is to explore the properties of the expected signature, to unify different approaches in the aforementioned articles and finally list some practical limitations before introducing a similar but more powerful embedding.

3 Path signature embedding

As a non-locally compact group, the theory of representation does not apply directly on the space of paths in $\mathbb{R}^n$, or $T((\mathbb{R}^n))$. However the topological embedding of the signature transform paves the way for a formal definition of a Fourier transform. Indeed, as monomials on a compact subset of $\mathbb{R}^n$ characterise continuous functions (Stone-Weierstrass) and moments of a bounded random variable characterise it's law [CO18], one could expect signature moments to characterise the law of measures on paths.

Two sides of the same coin : Π_n and S The approach in [Kap09] is based on algebra and differential geometry. It defines the dual element of F rather than f, being a function or a measure on the image of the signature.
The method is extended a lot in [CL16] which is based on functional analysis and topology.

Article	Space
[Kap09]	$\mathcal{S}$
[CL16]	$\mathcal{S}$
[CO18]	Π_n

Table 1: Summary of articles point of view on the expected signature

This difference is important in the case of a measure because when taking a measure on path spaces, one might not automatically assume to take the Borel sets from the tensor algebra, but rather a stronger topology on some parameterisation, with local properties - taking an arbitrary discretisation of the path if needed. In the case of Brownian motion :

$$\mu\left((t_1,\ldots,t_n),(U_1,\ldots,U_n)\right) = \int_{U_1}\cdots\int_{U_n} \frac{e^{-\frac{x_1^2}{2t_1}}}{\sqrt{2\pi t_1}} \prod_{i=1}^{n-1} \frac{1}{\sqrt{2\pi(t_{i+1}-t_i)}} e^{\left(-\frac{(x_{i+1}-x_i)^2}{2(t_{i+1}-t_i)}\right)} dx_n \ldots dx_1$$

Existence of random variables on the product topology
A measure on the product topology can be constructed incrementally from finite projections as follows : let $\mathcal{S}$ be the image of Π_n through the signature transform. $\mathcal{S} \subset G$ the set of group-like elements. It belongs in the full tensor algebra $T((\mathbb{R}^n))$. Let $\mathcal{S}^d$ be the projection of $\mathcal{S}$ on the truncated tensor algebra $T^d(\mathbb{R}^n)$ (denoted $G_{n,d}$ in [Kap09]).

For $d \geq d'$, let :

$$p_{dd'} : \mathcal{S}^d \to \mathcal{S}^{d'}$$

the canonical projection. Then $\mathcal{S}$ is the projective limit of the $\mathcal{S}^d$ as d tends to infinity.

Definition 3.1. *A pro-measure is defined as a compatible system of measures on the truncated image of the signature, that is $(\mu_d)_{d\in\mathbb{N}} = \mu_\bullet$ such that :*

$$\forall d' \geq d, p_{dd'} * \mu_d = \mu_{d'}$$

where $*$ denotes the push-forward defined as :

$$p_{dd'} * \mu_d(B) = \mu_d(p_{dd'}^{-1}(B))$$

For any cylinder Borel set on $\mathcal{S}$.

Then for any continuous function f on $\mathcal{S}$ the following holds :

$$\int_{\mathcal{S}^d} f(x) p_{dd'} * \mu_d(dx) = \int_{\mathcal{S}^{d'}} f(x)\mu_{d'}(dx)$$

A pro-distribution can be defined in the same way, where a distribution is defined as a functional on $\mathcal{C}^\infty$. Kolmogorov's extension theorem ensures the existence of random variables on $\mathcal{S}$ with the Borel topology, which is used in [CL16], [Kap09].

Property 3.2. *Let $\mu \in \mathcal{P}(\Pi_n, \tau)$, where τ is a topology that makes the signature continuous. The push-forward measure :*

$$S * \mu : B \in \tau_{pr} \mapsto \mu(S^{-1}(B))$$

defines a pro-measure.

Conversely, given a measure $\nu \in \mathcal{P}(\mathcal{S}, \tau_{pr})$, one can define the composed measure as :

$$\tilde{\nu} : A \mapsto \nu(S(A))$$

3.1 Expected signature

Definition 3.3 (Expected Signature)**.** *Let $\mu_\bullet$ be a pro-distribution on $\mathcal{S}$. The formal Fourier transform of $\mu_\bullet$ belongs to $\mathbb{R}\langle\langle Z_1, \dots, Z_n\rangle\rangle$ identified as $T((\mathbb{R}^n))$ and is defined as :*

$$\hat{\mathcal{F}}(\mu_\bullet) = \sum_{p=0}^{\infty} \sum_{i_1,\dots,i_p} \left(\int_{\psi \in \mathcal{S}^d} \psi_{i_1,\dots,i_p} d\mu_p \right) Z_{i_1} \dots Z_{i_p}$$

It is a generalisation of the discrete case $\mathbb{F}_n$ where the Fourier transform can be canonically expressed as $\sum_{\gamma \in \mathbb{F}_n} a_\gamma X^\gamma$.

It coincides with the definition of the expected signature of a measure μ on Π_n [CL16] by the following relation :

$$\begin{aligned}
\mathbb{E}_{x\sim\mu}[S(x)] &= \int_{x\in\Pi_n} S(x) d\mu(x) \\
&= \int_{\psi\in\mathcal{S}} \psi d\mu(S^{-1}(\psi)) \\
&= \int_{\psi\in\mathcal{S}} \psi d(S * \mu)(\psi)) \\
&= \hat{\mathcal{F}}(\tilde{\mu}_\bullet)
\end{aligned}$$

Where $\tilde{\mu}_\bullet = (S * \mu)$ is the pushdown of the measure μ to a pro-measure on $\mathcal{S}$ via the signature transform, which is a bijection from the space of unparameterised, tree-reduced paths to it's image in the tensor space $\mathcal{S}$.

The following two definitions are dear to inference tasks and oftentimes equivalent [SS18].

Definition 3.4. *A map ϕ is characteristic on $\mathcal{P}(\mathcal{X})$ if the application :*

$$\mu \in \mathcal{P}(\mathcal{X}) \mapsto \int_{x\in\mathcal{X}} \phi(x)\mu(dx)$$

is injective.

Definition 3.5. *A map ϕ is universal on $\mathcal{C}(\mathcal{X})$ if any element can be approximated uniformly and arbitrarily close by a linear functional on $\phi(\mathcal{X})$.*

The expected signature is not characteristic in general. However, characteristicness holds if the underlying space is compact [CS24] which is not convenient as Π_n is not locally compact. It can also hold if the tail of distribution verifies a decrease condition [CL16], but it is hard to verify in practice [CO18]. The alternative method of scaling the signature proposed in [CO18] grants characteristicness beyond compact support.

The same holds for universality and the normalisation also turns the signature map universal. Moreover, the scaling has robustness properties [CO18].

Property 3.6. *The expected signature of the Gaussian measure is :*

$$\mathbb{E}_{x\sim\mu}[S(x)] = \exp\left(-\frac{1}{2}\sum e_i^{\otimes 2}\right) \in T((\mathbb{R}^n))$$

The proof can be found in [Faw03] [Bau04].

Property 3.7 (Convolution of measures). *Let $\mu, \nu \in \mathcal{P}(\Pi_n)$, then :*

$$\mathbb{E}_{\mu*\nu}\left[S(x)\right] = \mathbb{E}_{\mu}\left[S(x)\right]\mathbb{E}_{\nu}\left[S(x)\right]$$

with equality in $T((\mathbb{R}^n))$.

Proof. Let $p \in \mathbb{N}$,

$$\begin{aligned}
\mathbb{E}_{\mu*\nu}\left[S(x)^p\right] &= \int_{\Pi_n} S(x)^p \mu * \nu(dx) \\
&= \int_{\Pi_n} S(x)^p \int_{y\in\Pi_n} \mu(dxy^{-1})\nu(dy) \\
&= \int\int_{\Pi_n^2} S(x)^p \mu(dxy^{-1})\nu(dy)
\end{aligned}$$

Translating the variable $x = x' * y$:

$$\begin{aligned}
\mathbb{E}_{\mu*\nu}\left[S(x)^p\right] &= \int\int_{\Pi_n^2} S(x*y)^p \mu(dx)\nu(dy) \\
&= \int\int_{\Pi_n^2} \sum_{k=1}^{p} S(x)^k S(y)^{p-k} \mu(dx)\nu(dy) \\
&= \sum_{k=1}^{p} \int_{\Pi_n} S(x)^k \mu(dx) \int_{\Pi_n} S(y)^{p-k}\nu(dy) \\
&= \sum_{k=1}^{p} \mathbb{E}_{\mu}\left[S(x)^k\right]\mathbb{E}_{\nu}\left[S(y)^{p-k}\right] \\
&= \left(\mathbb{E}_{\mu}\left[S(x)\right]\mathbb{E}_{\nu}\left[S(y)\right]\right)_p
\end{aligned}$$

□

It also works taking measures directly on the tensor algebra, which is the approach in [Kap09] and [CL16], see section A.1.1 on page 43 for the proof of the statement (straight application of the definition).

Being an infinite dimensional object, it is hard to use the full potential of the expected signature as is. The most popular inference methods with the signature involve Kernel Mean Embedding (KME) [CT24a] - be it random, truncated - and Maximum Mean Discrepancy metrics on distributions, which in the best case metrizes the weak topology [CS24].

The following sub-section describes selected cases where the expected signature can be enough, hopefully providing intuition on how to use it best.

The transform exposed the in the following section can be thought of as a data-driven approach to benefit from the full potential of the signature by selecting some transform on the signature of a path.

3.2 Log-signature and nilpotent diffusions

The objective and originality of this sub-section is to bridge the gap between the signature embedding and the harmonic analysis on the Heisenberg group from section 1.3.

In a nutshell, the log-signature at level two can be seen as a lift into the Heisenberg group. If a measure on path has a density, it's pushforward with the log-signature can be viewed as a function from the Heisenberg group to $\mathbb{R}$ which under integrability conditions can be decomposed with the Plancherel theorem on the Heisenberg group.
The focus is given on a class of paths that would be fully characterised by it's level-2 truncation.

One interested in the study of stochastic differential equations (SDE) can leverage Ito's calculus time-wise.

$$dX_t = \sum_{i=1}^{d} V_i(X_t) \circ dB_t^i \in \mathbb{R}^n$$

However in the path-wise study the Ito map $B \mapsto X$ is not ideal given it is not continuous. Taking the signature transform of the driving path of diffusion processes can be leveraged to approximate the flow of SDEs defined in section 2.3.

To some extent, it is possible to do so with the log-signature of a path :

Theorem 3.8 (Chen-Strichartz). *[Bau04] The logarithm of the signature in the sense of Stratonovich of Brownian motion is a Lie element given by the relation :*

$$S(\circ B)_t = \exp\left(\sum_{k\geq 1}\sum_{i_1,\dots,i_k=I}\Lambda_I(B)X_I\right)$$

With :

$$\Lambda_I(B)_t = \sum_{\sigma\in \mathfrak{t}} \frac{(-1)^{e(\sigma)}}{k^2\binom{k-1}{e(\sigma)}}\int_{\Delta^k(0,t)} \circ dB^{\sigma^{-1}.T}$$

Where $X_I = [X_{i_1},[\dots,[X_{i_{n-1}},X_{i_n}]\dots]]$ *and* $e(\sigma) = \#\left\{j\in[1,\dots,k-1], \sigma(j) > \sigma(j+1)\right\}$

This formula generalises to the signature of any submartingale. It relates to the Baker-Campbell-Hausdorff formula for Lie algebras [Reu03].

Example 3.9. *The first term of the log-signature is :*

$$ln(S(\circ B)_t)^{(1)} = \sum_{i=1}^d B_t^i X_i$$

And the second term reads :

$$ln(S(\circ B)_t)^{(2)} = \frac{1}{2}\int_0^t \sum_{1\leq i,j\leq d}(B_s^i\circ dB_s^j - B_s^j\circ dB_s^i)[X_i,X_j]$$

where Ito and Stratonovich integrals coincide in this case.

A simple case where the flow of an SDE is fully characterised by the log-signature is that of diffusion processes in n-step stratified nilpotent Lie groups (Carnot groups) [Bau04] [McK24].

If $n = 2$ and $d = 2$, the nilpotent Lie group it is isomorphic to the Heisenberg group described at section 1.3 on page 10.

Property 3.10. *The Heisenberg group is isomorphic to the free 2-step nilpotent Lie group defined in 2.11 on page 20 with the explicit isomorphism :*

$$\begin{aligned}\phi : \mathbb{G}_2(\mathbb{R}^2) &\longrightarrow \mathcal{H}_3\\ (x,\omega) &\longmapsto (x,\omega+\frac{1}{2}x^1x^2) = \begin{pmatrix}1 & x^1 & \omega+\frac{1}{2}x^1x^2\\ 0 & 1 & x^2\\ 0 & 0 & 1\end{pmatrix}\end{aligned}$$

Proof. Let $(\alpha^1,\alpha^2,\omega),(\beta^1,\beta^2,\omega')\in\mathbb{G}_2(\mathbb{R}^2)$.

$$\begin{aligned}\phi\left((\alpha^1,\alpha^2,\omega)\circledast(\beta^1,\beta^2,\omega')\right) =&\phi\left((\alpha^1+\beta^1,\alpha^2+\beta^2,\omega+\omega'+\frac{1}{2}(\alpha^1\beta^2-\beta^1\alpha^2))\right)\\ =&(\alpha^1+\beta^1,\alpha^2+\beta^2,\omega+\omega'+\frac{1}{2}(\alpha^1\beta^2-\beta^1\alpha^2)+\frac{1}{2}(\alpha^1+\beta^1)(\alpha^2+\beta^2)\\ =&(\alpha^1+\beta^1,\alpha^2+\beta^2,\omega+\omega'+\alpha^1\beta^2+\frac{1}{2}(\alpha^1\alpha^2+\beta^1\beta^2)\\ =&\phi(\alpha^1,\alpha^2,\omega)\phi(\beta^1,\beta^2,\omega')\end{aligned}$$

And :

$$\phi\left((\alpha^1,\alpha^2,\omega)^{-1}\right) = \phi\left((-\alpha^1,-\alpha^2,-\omega)\right) = \left((-\alpha^1,-\alpha^2,-\omega+\alpha^1\alpha^2)\right) = \phi\left((\alpha^1,\alpha^2,\omega)\right)^{-1}$$

Therefore ϕ is a group homomorphism. It's inverse is immediate. $\square$

Property 3.11. *Let V^i be the basis of a 2-step nilpotent Lie algebra of vector fields. The strong solution to the diffusion equation :*

$$X_t = x + \int_0^t \sum_{i=1}^d V^i(X_s) \circ dB_s^i$$

is :

$$\phi^{-1}\left(F(x, B_t^*)\right) = \phi^{-1}\left(\exp\left(\sum_{i=1}^d B_t^i V^i - \frac{1}{2}\sum_{1\le i<j\le d}[V_i, V_j] Levy_{i,j}(B)(t)\right)\right)$$

The left-brownian motion on the Heisenberg group is universal in solving linear 2-step nilpotent diffusion equations.

Proof. We proove the above explicitly for $d = 2$ and $n = 3$ with explicit matrix to have an example of what a nilpotent diffusion looks like. For the general case, it is a consequence of the Chen-Strichartz formula. Let $V_1 = ae_{1,2}$ and $V_2 = be_{2,3}$. $(V_1, V_2, [V_1, V_2])$ form a basis of $\mathfrak{h} \cong \mathfrak{g}_2(\mathbb{R}^2)$. The nilpotent diffusion can be rewritten :

$$X_t = x + \int_0^t \begin{pmatrix} aX_s^2 dB_s^1 \\ bX_s^3 dB_s^2 \\ 0 \end{pmatrix} = \begin{pmatrix} x_1 + ax_2B_t^1 + abx_3\int_0^t B_s^2 dB_s^1 \\ bx_3B_t^2 + x_2 \\ x_3 \end{pmatrix}$$

Besides :

$$\begin{aligned} F(x_0, B_t^*) =& \exp\begin{pmatrix} 0 & aB_t^1 & -\frac{ab}{4}\int_0^t B_s^1 dB_s^2 - B_s^2 dB_s^1 \\ 0 & 0 & bB_t^2 \\ 0 & 0 & 0 \end{pmatrix} x \\ =& \begin{pmatrix} 1 & aB_t^1 & -\frac{ab}{2}\int_0^t B_s^1 dB_s^2 - B_s^2 dB_s^1 + abB_t^1B_t^2 \\ 0 & 1 & bB_t^2 \\ 0 & 0 & 1 \end{pmatrix} x \end{aligned}$$

Then :

$$\begin{aligned} \phi^{-1}\left(F(x_0, B_t^*)\right) =& \begin{pmatrix} 1 & aB_t^1 & ab\int_0^t B_s^2 dB_s^1 \\ 0 & 1 & bB_t^2 \\ 0 & 0 & 1 \end{pmatrix} x \\ =& X_t \end{aligned}$$

For $n > 3$, the same technique can be used when the V^i's define diagonal blocks of 2-step nilpotent subspaces. For $d \geq 2, \mathfrak{g}_2(\mathbb{R}^d)$ has dimension $\frac{d(d+1)}{2}$ with basis $\left\{\left(V^i\right)_{1\le i\le d}, \left([V^i, V^j]\right)_{1\le i<j\le d}\right\}$. Having only zero as eigenvalue, a nilpotent matrix of size d can be written in Jordan form as :

$$A = P^{-1}\begin{pmatrix} 0 & 1 & & & & & \\ & 0 & & & & & \\ & & \ddots & & & & \\ & & & 0 & 0 & & \\ & & & & 0 & 1 & \\ & & & & & 0 & 1 \\ & & & & & & 0 \end{pmatrix} P \in \mathcal{M}_d(\mathbb{R})$$

with subspaces isomorphic to the Heisenberg Lie algebra $\mathfrak{h}$ □

The result holds for non-linear vector fields where the SDE might not have a closed-form solution. In practice, given a measure on Brownian paths, the log-signature embedding seen in the Heisenberg group can leverage well-defined harmonic analysis.

The embedding is reminiscent of a Carnot path, but instead of having $z' = xy'$ one has $z' = x'y$.

The interpretation of the geometry on the Heisenberg group is telling of what the log-signature takes as information at level two. Indeed it is more costly by an order of magnitude to join paths in the z-coordinate which corresponds to the Levy area of the path. Therefore more importance is given on path correlation compared to the endpoint of the path - the (x, y)-coordinates.

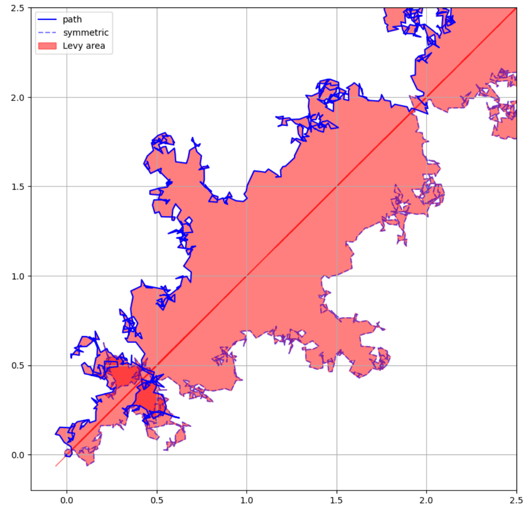


Figure 3: Geometric interpretation of the Levy area

Application of the Plancherel dual In section 1.3, the Fourier transform for the Heisenberg group was given as a representation on square integrable functions. Given $f : \mathcal{H} \to \mathbb{R}$:

$$\pi_t(f) = \iiint f(a,b,c)\pi_t(a,b,c)dadbdc \in \mathcal{B}\left(L^2(\mathbb{R})\right)$$

An explicit choice of basis can be telling of the action of these operators :

Basis function	Orthogonal
$x \mapsto 1_{[n/k,(n+1)/k[}$	no
$x \mapsto e^{-\frac{(x-\alpha)^2}{\beta}}$	no
Haar wavelets	yes

Table 2: Generating families of $L^2(\mathbb{R})$

Taking the example of $\phi_{\alpha,\beta} : x \mapsto e^{-\frac{(x-\alpha)^2}{\beta}}$ for $\alpha \in \mathbb{R},\ \beta > 0$, the Fourier transform can written more explicitly :

$$\langle \pi_t(f)\phi_{\alpha,\beta}, \phi_{\alpha',\beta'}\rangle_{L^2(\mathbb{R})} = \iiint_{\mathbb{R}^3} f(a,b,c) \int_{\mathbb{R}} e^{-A(x)}dx\ dadbdc$$

$$\begin{aligned}
A(x) =& \frac{1}{\beta\beta'}\left[(x+a+\alpha)^2 + (x+\alpha')^2 - 2\pi i(bx+c)\beta\beta' t\right] \\
=& \frac{2}{\beta\beta'}\left(x + \frac{1}{2}(a+\alpha+\alpha' - \pi i b\beta\beta' t)\right)^2 + \frac{1}{\beta\beta'}\left((a+\alpha)^2 + \alpha'^2 - \frac{1}{2}\left(a+\alpha+\alpha' - i\pi bt\beta\beta'\right)^2\right) - 2\pi i ct \\
=& \frac{2}{\beta\beta'}\left(x + \frac{1}{2}(a+\alpha+\alpha' - \pi i b\beta\beta' t)\right)^2 + B(a,b,c,\alpha,\alpha',\beta,\beta',t)
\end{aligned}$$

Then :

$$\langle \pi_t(f)\phi_{\alpha,\beta}, \phi_{\alpha',\beta'}\rangle_{L^2(\mathbb{R})} = \iiint_{\mathbb{R}^3} f(a,b,c)e^{-B(a,b,c)} \underbrace{\int_{\mathbb{R}} e^{-\frac{2}{\beta\beta'}\left(x+\frac{1}{2}(a+\alpha+\alpha'-\pi i b\beta\beta' t)\right)^2}dx}_{I}\ dadbdc$$

Let $\gamma = \gamma_r \cup \gamma_I \cup \gamma_1 \cup \gamma_2$. With a linear change of variable :

$$I = \int_{\mathbb{R}} e^{-\frac{2}{\beta\beta'}\left(x - \frac{1}{2}\pi i b\beta\beta' t\right)^2}dx = \lim_{B\to+\infty}\left(\int_{\gamma_I} e^{-\frac{2}{\beta\beta'}z^2}dz\right)$$

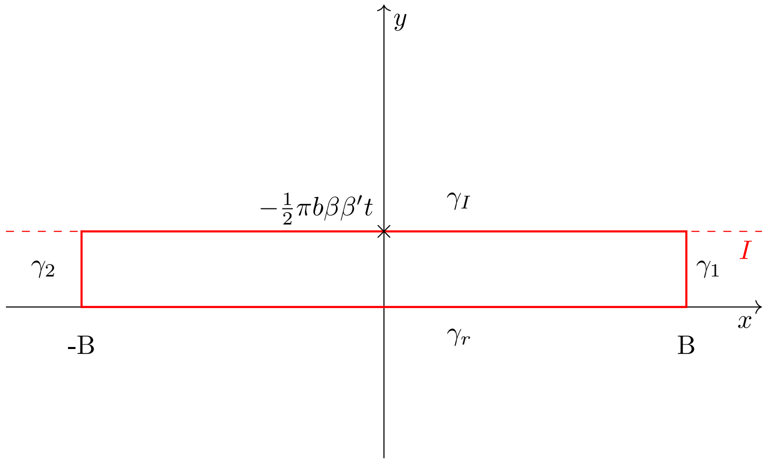


Figure 4: Contour integral of I

For any $c \in \mathbb{C}$, $z \mapsto e^{cz^2}$ is entire on $\mathbb{C}$. By Cauchy's integral theorem :

$$\oint_\gamma e^{-\frac{2}{\beta\beta'}z^2} dz = 0$$

The integral on each side tend to zero as B tends to infinity, eventually giving :

$$I = \sqrt{\frac{\beta\beta'\pi}{2}}$$

$$\langle \pi_t(f)\phi_{\alpha,\beta}, \phi_{\alpha',\beta'} \rangle_{L^2(\mathbb{R})} = \sqrt{\frac{\beta\beta'\pi}{2}} \iiint_{\mathbb{R}^3} f(a,b,c) e^{-B(a,b,c)} dadbdc$$

This example was taken for ease of computation rather than intuition and using a different basis would give a completely different result. Using an orthogonal countable basis of functions like Haar wavelets, whose computation and numerical implementation seem feasible, could be preferable to extract enough independant information on the function f to characterise it.

In the case of nilpotent diffusion, the function f can be taken as the pushforward of the measure on paths with the log-signature at level 2 :

$$f(a,b,c) = \log S^{(2)} * \mu = \mu(\log S^{-1})^{(2)}$$

when the measure has a density with respect to the Lebesgue measure.

A generalisation can be sought for nilpotent groups of order n greater than 2, but the harmonic decomposition and Plancherel theorem that have been proven for the Heisenberg group would have to be verified. Besides, the log-signature is not in general universal.

Conclusion on expected signature Topological structures on the space of paths don't fit in the compact or locally compact cases where the Peter-Weyl or Plancherel theorem hold. Instead of working directly in the space of paths, one can embed it, for instance with the signature transform into the tensor algebra. The limitation to compact subsets of paths has been tackled in [CO18] using an appropriate scaling, but controlling the truncation error remains challenging. Besides, the use of a Kernel Mean Embedding into an Reproducing Kernel Hilbert Space is a known trick against the curse of dimensionality as long as the Maximum Mean Discrepancy metrizes weak convergence [CS24].

As explained in [CL16], the expected signature can be thought of as a moment generating function for variables on the space of paths. Another natural approach to weak description of random variables is the characteristic function.

Towards a characteristic function The main finding in [CL16] is the definition of an equivalent characteristic function for measures on the image of the signature.

Besides, an exponential-like equation into matrix Lie groups coined *path development* [LLN22], with similar algebraic properties than the signature transform, is a representation of paths directly in those matrix Lie groups. Without going through the tensor algebra, this latter CDE solution has a canonical relation with the signature [LLN22] which will be reviewed.

The objective of the following section is to study the applicability of path developments to compact matrix Lie groups, combined with harmonic analysis tools to characterise functions on the space of paths. Expected benefits include avoiding the curse of dimensionality, still having interpretable and well structured dual matrix coefficients, and being able to adapt coefficients to specific situations.

4 Path Development embedding

Path development is an embedding from paths into a matrix Lie group such as $GL_n(\mathbb{C})$ or the symplectic group $Sp(n)$.

Statement of objective If the Lie group is compact, one can apply harmonic analysis to it, which is the objective thereof. Compact matrix Lie groups include $O(n), SO(n), U(n), SU(n)$ - they are closed and bounded. Thus our primary focus in this section is to apply compact group tools of harmonic analysis, namely the Peter-Weyl decomposition of L^2 functions.

4.1 Definition and properties

Embedding of paths in Lie groups has been pioneered in [McK24] for parameterised stochastic processes and is defined in [Kap09] with differential geometry tools :

Definition 4.1 (Path development as a holonomy on smooth paths)**.**
Let $G < GL_n(\mathbb{C})$ closed and $\mathfrak{g} \subset \mathcal{M}_n(\mathbb{C})$ it's Lie algebra. Let $A = \sum A_i(y)dy_i$ be a smooth, $\mathfrak{g}$-valued 1-form on $\mathbb{R}^n$. The path development of a smooth path γ is :

$$Hol_\gamma(A) = P \exp \int_\gamma A \in G$$

This definition is limited as piecewise-smooth paths have measure zero in the Wiener distribution. Replacing the differential equation by an integral equation in a regularised sense like Itô or Stratonovich allows to extend the definition to continuous paths :

$$U(t) = I + \int_0^t U(s) \sum A_i(\gamma_s) \circ d\gamma_s$$

for example in the Stratonovich sense.

Recently, it has been successfully applied in [LLN22] as a learnable embedding of paths onto a finite dimensional manifold. The manifold constraint noticeably provides stability properties during training, which is relevant in sequential data learning with RNNs to tackle vanishing and exploding gradient issues. It has further been used in [CT24a] to efficiently compute signature kernels. In this latter application, M is randomized to define a kernel mean embedding (KME) and related maximum mean discrepancy (MMD) metric on measures.

The path development can be expressed similarly to the path signature through a controlled differential equation :

Definition 4.2 (Path development)**.**
Let $G < GL_n(\mathbb{C})$ a matrix Lie group and $\mathfrak{g}$ it's matrix Lie algebra. Let $M : \mathbb{R}^n \to \mathfrak{g}$ be linear.
Let $\gamma \in \Pi_n$ and X_t be it's unit-speed class representative, starting at the origin.
The following controlled differential equation has a unique solution :

$$\begin{cases} dZ_t = Z_t M(dX_t) \\ Z_0 = I \end{cases}$$

It is denoted Z_t^M and is called path development of X.

It has similar algebraic properties than the signature, including Chen's relation, which enables efficient computation of paths developments by linear approximation of the path and matrix multiplication.

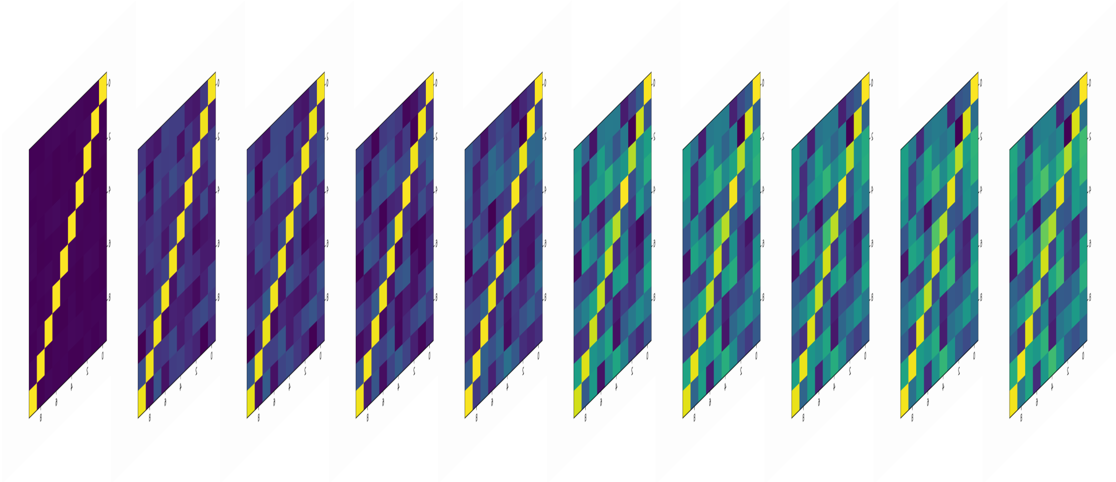


Figure 5: Illustration of path development in $O(10)$

Chen's relation To have that $D_M(x * y) = D_M(x)D_M(y)$, it is sufficient to have unicity for the following equation :

$$\begin{cases} Z_0 = D_M(x)_t \\ dZ_v = Z_v M dy_v \end{cases}$$

Because both $v \mapsto D_M(x * y)_{t+v}$ and $v \mapsto D_M(x)_t D_M(y)_v$ are solutions.

Continuity The product topology also ensures path developments are continuous.

Property 4.3 (Path development as a representation)**.** *The last two statements ensure that path developments define a representation of Π_n onto $\mathbb{R}^n$ or $\mathbb{C}^n$.*

Reparameterisation invariance Ensures well posedness of the definition as an exponential (which is parameterized). Reference article with proof.

Universality One has a similar diagram to the path signature :

$$\begin{tikzcd} \Pi_n \arrow[r, "f"] \arrow[dr, dashed, "D_{G_\cdot}"'] & \mathbb{C} \\ & G_\cdot \arrow[u, "F"'] \end{tikzcd}$$

Where $G_\cdot$ is a matrix Lie group of arbitrary dimension and $D_{G_\cdot}$ represents the developments D_M as M spans G_n for any n. An important result from [CL16] is that the subalbegra of $D_{G_\cdot}$ of albegra morphisms that stem from a linear map $V \to \mathfrak{g}$ seperate points on the image of the signature :

$$\begin{tikzcd} \Pi_n \arrow[d, "S"'] & \mathbb{C} \\ T((V)) \arrow[r, "M \in \mathrm{Hom}(\mathbb{R}^n)"'] & G \arrow[u, "F"'] \end{tikzcd}$$

The Stone-Weierstrass completes the universality of path developments : f can be approximated uniformly and arbitrarily close by $F \circ D$.

Property 4.4 (Link with the signature [LLN22]). *If $M \in L(\mathbb{R}^n), \mathfrak{g})$, one can extend it naturally to an algebra homomorphism $\tilde{M}$. Then :*

$$\forall x \in \Pi_n,\ D_M(x) = \tilde{M}(S(x))$$

Eventually as the signature spans group-like elements of $T((V))$, the correspondance is unambiguous.

Characteristic function on signature elements

4.1.1 Interpretability

There is no canonical choice of representation for path development. One can imagine the choice of matrix Lie group greatly influences its expressiveness and properties. Moreover, once the matrix Lie group G is chosen, the matrix Lie algebra embedding M needs to be defined and is not canonical.

While [CT24a] randomizes the choice of M and uses a kernel trick to avoid computing expectations on matrix groups, the authors in [LLN22] take a data-dependant approach. In this case multiple channels take different M and are considered as trainable parameters. The correspondence between a matrix Lie group and it's Lie algebra simplify the training and enable efficient backpropagation [Lez19].

The following example path developments can provide some insight on possible actions on paths.

Hyperbolic development :
From [LLN22]. M is chosen to have the action on $\mathbb{R}^2$:

$$MdX = \begin{pmatrix} & & dx \\ & & dy \\ dx & dy & \end{pmatrix}$$

Then define Z_t be the path development of some path X in $\mathbb{R}^2$. The path in $\mathbb{R}^3$ defined by $\gamma(0) = (0,0,1)^T$ and $\gamma(t) = Z_t\gamma(0)$ belongs to :

$$\mathbb{H}^2 = \left\{(x,y,z) \in \mathbb{R}^3 | x^2 + y^2 - z^2 = -1\right\}$$

Heisenberg development
For a given path in $\mathbb{R}^2$:

$$MdX = \begin{pmatrix} 0 & dx & \\ & 0 & dy \\ & & 0 \end{pmatrix}$$

Spherical development :
M is chosen to have the action on $\mathbb{R}^2$:

$$dZ = MdX = \begin{pmatrix} & & dx \\ & & dy \\ -dx & -dy & \end{pmatrix} \in \mathfrak{o}(3)$$

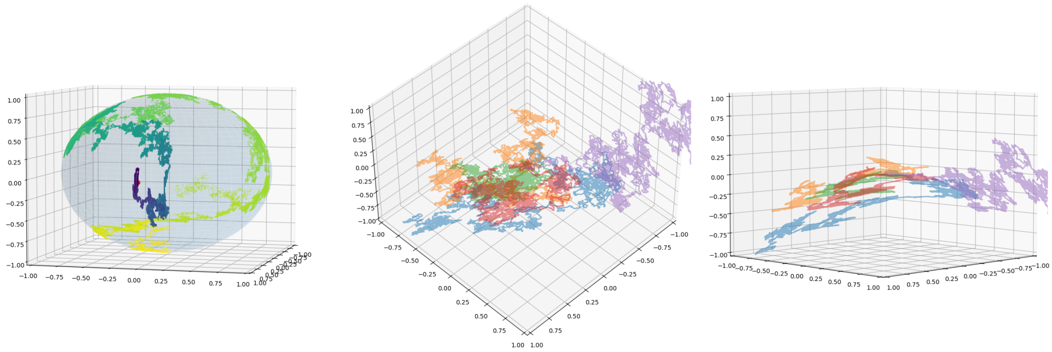
Finding symmetries would enable one to understand better the action of path developments. These symmetries are expected to play a role in the harmonic analysis stemming from path development as is the case for periodic function on real Fourier analysis. Lowering dimensions to 1-dimensional paths, the spherical development gives :

$$\exp\left(MdX\right) = \begin{pmatrix} \cos(dX) & \sin(dX) \\ -\sin(dX) & \cos(dX) \end{pmatrix} \in O(2)$$

and $Z_t^M = \mathrm{Rot}_2(X_t - X_0)$ the rotation in $\mathbb{R}^2$ of angle $X_t - X_0$ which is 2π-periodic.

In the 3-d rotation development, the lines of Z_t seen as vectors in $\mathbb{R}^3$ follow the same evolution equation :

$$dZ_t^i = \begin{pmatrix} -z_t^{i3}dx \\ -z_t^{i3}dy \\ z_t^{i1}dx + z_t^{i2}dy \end{pmatrix}$$



(a) Spherical (b) Heisenberg view 1 (c) Heisenberg view 2

Figure 6: Example developments of Brownian motion

Take the simple linear path : $t \mapsto \begin{pmatrix} t \\ t \end{pmatrix}$, the equation has the closed form :

$$Z_t^i = \exp{(At)}\, e_i$$

Where :

$$A = \begin{pmatrix} 0 & 0 & -1 \\ 0 & 0 & -1 \\ 1 & 1 & 0 \end{pmatrix}$$

After a simple computation :

$$\exp{(At)} = \begin{pmatrix} \frac{1}{2}(1 + \cos\sqrt{2}t) & \frac{1}{2}(\cos\sqrt{2}t - 1) & -\frac{1}{\sqrt{2}}\sin\sqrt{2}t \\ \frac{1}{2}(\cos\sqrt{2}t - 1) & \frac{1}{2}(1 + \cos\sqrt{2}t) & -\frac{1}{\sqrt{2}}\sin\sqrt{2}t \\ \frac{1}{\sqrt{2}}\sin\sqrt{2}t & \frac{1}{\sqrt{2}}\sin\sqrt{2}t & \cos\sqrt{2}t \end{pmatrix} \in O(3)$$

Which is $\sqrt{2}\pi$-periodic.

4.2 Expected development

Challenge to the definition The Fourier transform for compact groups defined in section 1.2 on page 9 is expressed as :

$$\tau(f) = \int_{x \in G} f(x)\tau(x)dx$$

Where τ is a finite-dimensional irreducible unitary representation.

In our case, the compact group G is only an embedding. It would be natural to integrate on the space of paths and use the pushforward to the matrix group in the integrand. However, unlike the signature, the path development for fixed M is not injective on the space of tree-reduced, unparametrized paths Π_n. Composing with this contraction is the point of the following section.

Definition of the characteristic function with the Peter-Weyl theorem
Let $\mu \in \mathcal{P}\left(\Pi_n, \mathcal{B}(\Pi_n)\right)$ be a probability measure, where the Borel sets of Π_n are defined as in the appendix 2.8 on page 19.

Denote with $D : \Pi_n \to G$ the path development driven by some $M : \mathbb{R}^n \to \mathfrak{g}$. It is continuous therefore measurable, and one can define the pushforward of the measure μ :

$$\begin{tikzcd} \Pi_n \arrow[r, "D"] \arrow[d, "\mu"'] & G \arrow[dl, "D_*(\mu)"] \\ \mathbb{R} & \end{tikzcd}$$

The probability measure $\tilde{\mu} = D_*(\mu) \in \mathcal{P}(G, \mathcal{B}(G))$ is well-defined by composition of measurable maps, and finite if μ is integrable.
Define the equivalence class on paths defined by equality of their development :

$$x \sim_D y \Leftrightarrow D^M(x) = D^M(y)$$

Denote by $\tilde{\Pi}_n = \Pi_n / \sim_D$ the set of equivalence classes. The sets :

$$\left\{D^{-1}(X)/X \in G\right\}$$

partition Π_n (with conditions on M to ensure development is surjective) and is a subset of the Borel algebra of Π_n (the development is continuous). One deduces the measure $\hat{\mu} \in \mathcal{P}(\tilde{\Pi}_n)$ by $\hat{\mu}(\hat{x}) = \mu(D^{-1}(D(x))$. Then $D : \tilde{\Pi}_n \to G$ is an isomorphism. Choose $\tilde{\tau}$ as the trivial identity representation of G. Taking G as $O(n)$ or $U(n)$ ensures it is unitary. Further assume that $\tilde{\mu} \in L^1(G)$.

$$\tilde{\tau}(\tilde{\mu}) = \int_{X \in G} X \tilde{\mu}(dX)$$

The integral being Lebesgue as the integrand is in $\mathcal{M}_n(\mathbb{R})$ up to isomorphism. Rename the path development D as τ to coincide with the theory of representation literature. Expliciting the push-forward of the measure, the bijection on the equivalence classes gives :

$$\int_{X \in G} X \tilde{\mu}(dX) = \int_{\tilde{\Pi}_n} \tau(x) \hat{\mu}(dx)$$

Given $\tau(x)$ is constant on the equivalence classes $\sim_D$, the integral is eventually :

$$\int_{\tilde{\Pi}_n} \tau(x) \hat{\mu}(dx) = \int_{\Pi_n} \tau(x) \mu(dx)$$

Therefore information is lost in the quotienting necessary to apply the Peter-Weyl theorem and it it clear that the transformation is not characteristic for any finite dimensional group G. A key to understanding the behaviour of this equivalent characteristic function and understand the role of the choice of embedding will be to have intuition on what equivalence classes $\sim_D$ are. Another interesting aspect to delve into would be in the limit of n [Voi91], see [CT24a] for an example with kernelized path developments.

From there it is immediate that taking only one path development will not make the representation characteristic. A further step would therefore be to take multiple such M, each corresponding to a development, to separate points in $\mathcal{P}(\Pi_n)$. It is done via randomizing the embedding in [CT24a] or optimising the lift coefficients [LLN22] for learning task applications.

Property 4.5. *The development embedding of the convolution of measures is not in general the product of both. Instead, it is the case for the pushforward of the measures in the following sense : Let $\mu, \nu \in \mathcal{P}(\Pi_n)$ with corresponding push-forward $\tilde{\mu}, \tilde{\nu} \in \mathcal{P}(\tilde{\Pi}_n)$,*

$$\tau(\tilde{\mu} * \tilde{\nu}) = \tau(\tilde{\mu}) \tau(\tilde{\nu})$$

Proof. It relies on the isomorphism property of D from $\tilde{\Pi}_n$ to G :

$$\forall Y \in G, D^{-1}(Y)^{-1} = D^{-1}(Y^{-1})$$

and

$$\forall dX \in \mathcal{B}(G), D^{-1}(Y^{-1}) * D^{-1}(dX) = D^{-1}(Y^{-1} dX)$$

Moreover the change of variable $Y = D(y)$ for $y \in \tilde{\Pi}_n$ is a one-one correspondence.

$$\begin{aligned}
\tau(\tilde{\mu} * \tilde{\nu}) &= \int_{X\in G} X(\tilde{\mu} * \tilde{\nu})(D^{-1}(dX)) \\
&= \int_{X\in G} X \int_{y\in\tilde{\Pi}_n} \tilde{\mu}(D^{-1}(dX) * y^{-1})\tilde{\nu}(D^{-1}(dY)) \\
&= \int_{X\in G} X \int_{Y\in G} \tilde{\mu}(D^{-1}(dX) * D^{-1}(Y)^{-1})\tilde{\nu}(D^{-1}(dY)) \\
&= \int_{X\in G}\int_{Y\in G} X\tilde{\mu}(D^{-1}(dX.Y^{-1}))\tilde{\nu}(D^{-1}(dY)) \\
&= \int_{X\in G}\int_{Y\in G} XY\tilde{\mu}(D^{-1}(dX))\tilde{\nu}(D^{-1}(dY)) \\
&= \int_{X\in G} X\tilde{\mu}(D^{-1}(dX)) \int_{Y\in G} Y\tilde{\nu}(D^{-1}(dY)) \\
&= \tau(\tilde{\mu})\tau(\tilde{\nu})
\end{aligned}$$

□

As the proof mainly relies on the separability of paths given their matrix development, one could wonder if the result would hold in the limit of the size of M for well-chosen coefficients, which could relate to free probability theory and results on limits of KME from [CT24a].

Finding closed form Fourier transforms Having in mind that the expected signature of the Gaussian measure on paths is the gaussian in $T((V))$ defined as $\exp\left(\sum e_i^{\otimes 2}\right)$, an interesting result would be to have a similar form for $\tau(\mu)$ when μ is a gaussian probability measure on Π_n.
A toy example is obtained by taking the development on the torus :

$$\begin{aligned}
M_\lambda : \mathbb{R}^n &\to \mathfrak{gl}_n(\mathbb{C}) \\
x &\mapsto i\lambda \mathrm{diag}(x)
\end{aligned}$$

resulting in $\tau_\lambda(\mu) \sim \phi_\mu(\lambda)$, the characteristic function of the endpoint random variable.
The brownian lift on the Heisenberg group from section 2.3 can be seen as a path development with :

$$M(dB) = \begin{pmatrix} 0 & dB^1 & 0 \\ 0 & 0 & dB^2 \\ 0 & 0 & 0 \end{pmatrix}$$

4.2.1 Numerical considerations

An implementation of path developments is implemented in the library [LLN22] :

`https://github.com/PDevNet/DevNet`

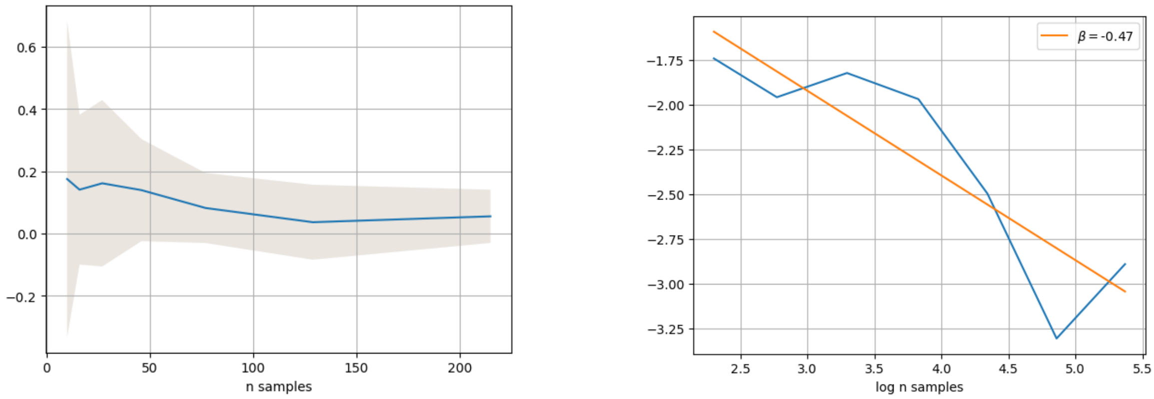
and has been used for illustrations and simulations. An optimisation called trivialization [Lez19] is also provided to backpropagate the gradient to the choice of matrix M for data-driven applications.

Computing the Peter-Weyl transform of a measure
Applying a naive Monte-Carlo methods to compute :

$$\int_{\Pi_n} \tau(x)\mu(dx)$$

would be intensive given the size of the group of paths and the computational cost of path developments.

Example 4.6. *With $n = 1000$ paths of length $m = 1000$ in dimension $d = 10$, solving the path development equation for each costs at least $\mathcal{O}(nmd) \approx 10^7$ FLOPS.*
The convergence rate for Monte-Carlo is $\mathcal{O}\left(\frac{1}{\sqrt{n}}\right)$ and suffers from the dimensionality as errors accumulate.



(a) Convergence plot and confidence intervals

(b) Convergence speed

Figure 7: Monte Carlo convergence of Peter-Weyl representation for the standard brownian measure

Computing instead :

$$\int_{X \in G} X \tilde{\mu}(dX)$$

requires the knowledge of the pushdown measure and an efficient integration method on groups like $O(n), SO(n)$.

Detailing the spherical development earlier in section 4.1.1 on page 32 was motivated by the following two considerations :

- the Cayley map is a one-to-one mapping from antisymmetric matrices to the orthogonal group :

  $$A \mapsto (I + A)(I - A)^{-1}$$

  Which is potentially quite cheap using an adapted stateless inversion method (it can also be used to compute the whole path development [Ise01]).

- efficient computations for matrix exponential $\exp : \mathfrak{g} \to G$ [AH10] also enables sampling from $\mathfrak{o}(n) = \mathcal{A}_n$:

  $$\mathcal{A}_n \to O(n)$$

Finding well defined symmetries would allow one to use quadrature methods or to uniformly sample from a compact of the Lie algebra of dimension $\frac{n(n-1)}{2}$, and use a Monte Carlo scheme to compute an integral on $O(n)$.
But as one of my lecturers once put it :

"Monte Carlo is depressingly slow. You should only use it if you hate yourself"

- eventually adding *"or if you have no other choice eg in high dimension"*.

One can kernelize the development to obtain an efficient computation solving a PDE [CT24a] - the cost is even independent of the dimension on the path.

Questions then remain :

- What are equivalence classes like ?
- How to compute push-forward ?
- How to integrate efficiently on compact matrix groups ?

A numerical experiment have been conducted on CPU with 16Gb RAM. With 2-dimensional paths of length 1000 each, with development on $O(10)$, having a few thousand samples was already too much to be reasonably applied to a simple problem. Still, the task is embarrassingly parallel and the implementation in [LLN22] can be used on GPU.

4.2.2 Applications and prospects

Instead of finding F directly, one interested in some of its properties - to be defined - can focus on the dual group of G :

$$\begin{tikzcd}
\Pi_n \arrow[r, "f"] \arrow[d, dashed, "d"'] & \mathbb{C} \\
G \arrow[ur, "F"] \arrow[r, "\mathcal{F}"'] & \hat{G}
\end{tikzcd}$$

The idea is that elements in $\hat{G}$ could capture symmetries and steady-state properties of the function in a compressed format.
One can as well consider the final step of finding a function $\hat{G} \to \mathbb{C}$ which would make the diagram commute :

$$\begin{tikzcd}
\Pi_n \arrow[r, "f"] \arrow[d, dashed, "d"'] & \mathbb{C} \\
G \arrow[r, "\mathcal{F}"'] & \hat{G} \arrow[u, "\tilde{F}"']
\end{tikzcd}$$

It is akin to finding the inverse Fourier transform. This question is mentioned in [Kap09] for the expected signature, mentioning the difficulties it involves.

As $\hat{G}$ is the direct sum of finite-dimensional unitary representations, it is infinite-dimensional. In fact, this situation has a different purpose. Indeed to compute $\tau(D * f)$, one needs to integrate on the whole space and so is not a point-wise operation. Instead, it could be applied to the problem of distribution regression (DR), where a function f_i is associated to a scalar target y_i.

Examples include regressing the temperature of a gas from particle trajectories [Lem+21].

Following data-driven incentives, one could think of pipelining the Fourier transform of a distribution on paths in a supervised learning method sending paths onto their Cartan developments on one - or more - compact matrix Lie group, and then regressing on the Peter-Weyl matrix Fourier coefficients to a scalar value. Furthermore, the Cartan development can be optimised with backpropagation [Lez19].

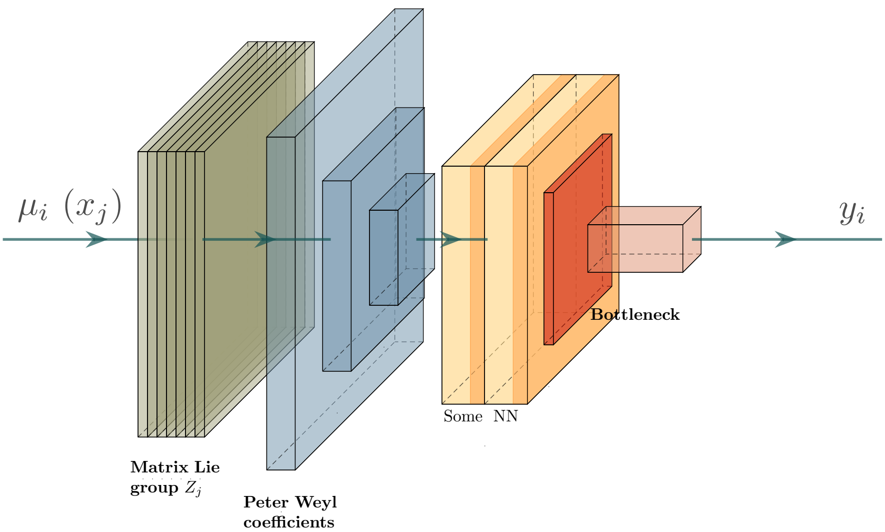


Figure 8: Path development embedding and Fourier transform for Distribution Regression

Covariance-like estimators can be sought, for instance using the orthogonal group $O(n)$ as compact matrix Lie group, for $\mu, \nu \in \mathcal{P}(\Pi_n)$, define the Hilbert-Schmidt inner product on matrices as :

$$d(\mu, \nu) = \operatorname{tr}\left(\tau(\mu)^* \tau(\nu)\right)$$

With a given dataset the empirical estimator is then :

$$d_{n,m}(\mu, \nu) = \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} \overline{\mu(x_i)} \nu(y_j) \operatorname{tr}\left[\tau(x_i)^* \tau(y_j)\right]$$

The motivation is that as the matrix in $O(n)$ defines an orthonormal basis of $\mathbb{R}^n$, taking the trace of the product of representations would keep the order of basis elements that each define and compare how they project onto each other position-wise. As their norm is one, the representations don't interfere with the measure scale (otherwise it would be ill-posed).

It is reminiscent of the celebrated method of Hilbert-Schmidt kernelized covariance coined in [Gre+07], defining an independence criterion (HSIC).

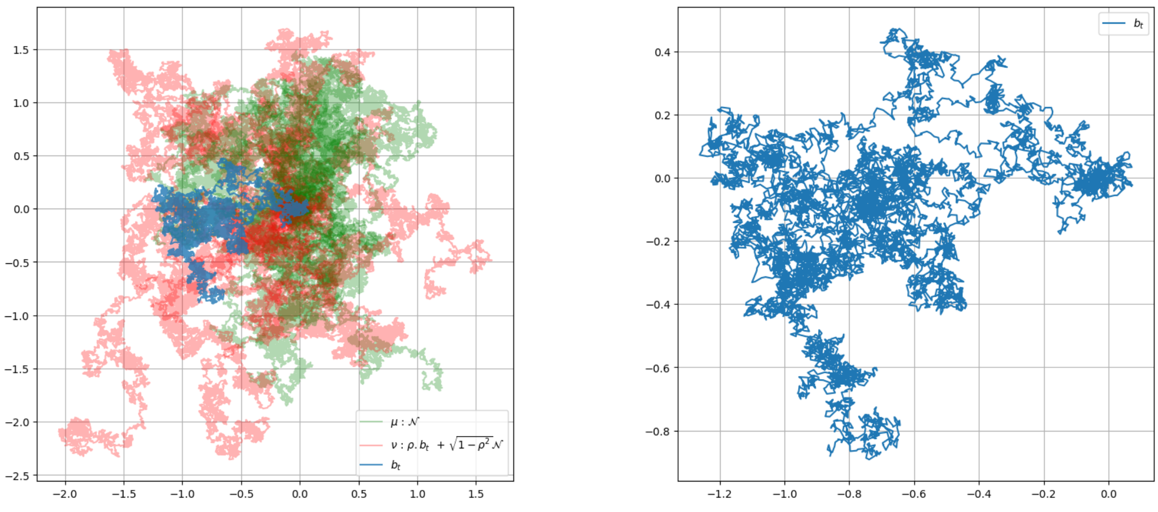
We also refer to [CT24a] for the kernelized MMD on paths developments.

As displayed in figure 8, paths developments can be condensed in matrices of different sizes, and the development can have different channels into different matrix groups. This is potentially very expressive, and can be combined with off-the-shelf algorithms for classification or regression. The difficulty is transferred to having intuition on which development and representation of the latter to choose.

Example Hypothesis Testing Having two slightly different gaussian measures, can their matrix representation tell if a new sample has been drawn from one or the other ?

The distributions that are taken as example are :

- the standard Brownian motion parameterized on $[0, 1]$
- A Brownian motion with distribution $\mathcal{N}(\rho b_t, \sqrt{1-\rho^2} I)$, where b_t is a fixed path, drawn from a standard BM.



(a) Example paths from the pair of distributions

(b) Common trend for ν paths

Figure 9: Samples from simulated dataset

It's not so obvious to distinguish a path from those two distribution visually. Then as path development can be considered as a trainable layer - which is expected to provide more adaptability than the parameter-free signature - the following method has been used to train the model :

The action of paths on the group of positive rotations SO would be an orthogonal change of coordinates.

The speed and direction of rotation into the group is expected to be controlled during training - as a branch cut. A full rotation equals the identity so only the residual will count.

Algorithm 1 Training methodology

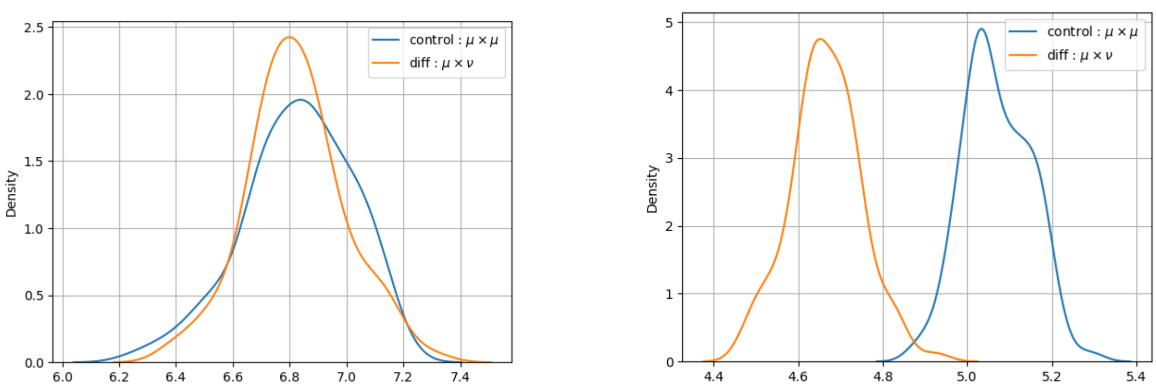
Require: Development group
Require: Representation on compact, default=Id
  $(x^1, x^2), l^1 \leftarrow \mu \times \mu$, -1 ▷ independent or bootstrap
  $(y^1, y^2), l^2 \leftarrow \mu \times \nu$, 1 ▷ independent or bootstrap
  concatenate and shuffle dataset
  initialize dev layer
  optimizer $\leftarrow$ Adam (params = dev, $lr = 5e^{-2}$)
  while $N_{\text{epochs}} \neq 0$ **do**
    sample $(x, y), l$ from batch
    $X, Y \leftarrow \text{dev}(x), \text{dev}(y)$ ▷ + optional representation compact group
    loss $\leftarrow l \times \langle X, Y \rangle_{HS}$
    backpropagate development parameters
    $N_{\text{epochs}} \leftarrow N_{\text{epochs}} - 1$
  end while

Example 4.7. *In the experiment, $\rho = \frac{3}{10}$, paths have 100 time-steps, sample distributions for training have 100 paths each, all independently drawn.*

The training dataset is then composed of 30 couple (μ, μ) and 30 couple (μ, ν). It is equivalent to have a balanced training set of 6000 pairs of paths. The batch size was 20 and the optimizer was Adam with naive learning rate $5e^{-2}$.

Test sample distributions are taken independent of the train sets.

The training optimisation has not been the primary focus. Still, the effect is visible after only a few epochs :



(a) before training

(b) after 10 epochs

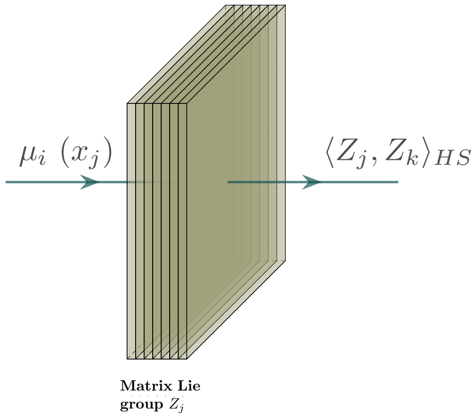
Figure 10: Training separability of distributions

These KDE plots have been obtained with 100 samples of pairs of distributions of 50 paths for each distribution. The Hilbert-Schmidt inner product seperates linearly the problem. One could estimate type-I and type-II errors to carry out a hypothesis test for a new sample.

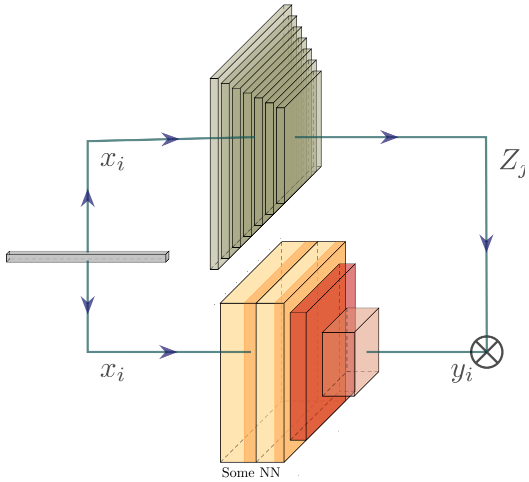
Applicability Note that this problem already has a simple solution : the covariance of the increments is an estimator that can separate the two distributions. It is a lot more powerful and requires a lot less compute and data as each increment counts as an independent draw. However when the increments are not stationary, or not normally distributed, or that the trend is not a brownian path itself, the assumptions to use Pearson's covariance estimator do not hold.

Using path development does not make assumptions on the increments of the paths, only that they are drawn independently from the same distribution. Monte-Carlo Markov-Chain methods could also be considered to compute the integral if the density is kwown.

Benchmark Other methods for this task include KNN - where different metrics can be used. There is no training time but evaluation is costly.



(a) Standalone



(b) Layer in a larger model

Figure 11: Different integration of path development

Hidden size	Train (%)	Test (%)
2	52.2	56.0
3	**98.5**	94.8
4	**98.5**	95.1
5	97.0	94.7
6	97.0	95.6
7	**98.5**	**96.2**
8	97.0	94.7
9	97.0	95.4
10	97.0	95.9

Figure 12: SO Accuracy vs hidden size

A missing part of this application would be to compare performance with the expected signature method [Lem+21], and KNN as an off-the-shelf baseline.

Then, studying the effect of the number of channels, the size of the embedding and more importantly the choice of group could be explored. The code used to compute simulations and plot all illustrations in this paper is available as a fork of the repository from [LLN22] at :

`https://github.com/chataignault/DevNet`

Application to time-series classification : Italy Power Demand dataset
The data is available from the *UAE time-series classification website* `https://www.timeseriesclassification.com/` .

With a reasonable development size (2 channels of dimension 5 - which totals 20 parameters on SO), path developments can tally the logistic regression considering time-series as tabular data (in terms of accuracy). The latter having more parameters (24).

Group	Train (%)	Test (%)
SO	**98.5**	**96.2**
Sp	92.5	92.0
Se	94.0	90.3

Figure 13: h=6 accuracy vs group

N channels	Train (%)	Test (%)
1	97.0	95.9
2	**98.5**	**96.7**
3	**98.5**	96.4
4	**98.5**	96.2
5	97.0	94.8

Figure 14: h=5 accuracy vs n channels

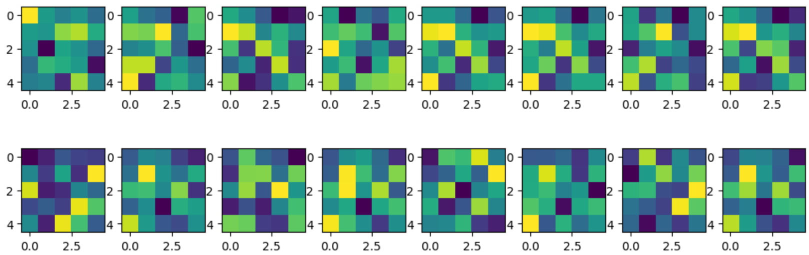


Figure 15: Samples of path development per class

Towards a local characterisation Both the signature and Cartan development of a path have non-local properties : they are a form of integration on the whole path and thus inverse problems are impractical.

Other methods including Dynamic Time Warping (DTW) [Kru83] or shapelets [YK09] (classification) tend to take the opposite approach in characterising paths.

To this extent and in continuation to Fourier analysis on the real line, harmonic analysis defines wavelet theory [DE14] which from a first approach seem promising to encode local information of functions or measures on the space of paths.

A Theorems and properties

A.1 Functional Analysis properties

- The existence of a Schauder basis implies the space is separable
- An **absorbing set** can be scaled up to include any point in a vector space. Every neighbourhood of the origin is an absorbing subset
- **Minkowski functionnal** : $K \subset X$, $p_K(x) = \inf\{r \in \mathbb{R}_+ / x \in rK\} \in \bar{\mathbb{R}}_+$. K often assumed to be an absorbing disk in X. Every semi-norm p on X is the Minkowski functionnal with respect to some set K (absorbing disk).
- A **semi-norm** need not be positive definite. A norm is a semi-norm that separates points.

Theorem A.1 (Bessel, real analysis). *$G = (\mathbb{R}/\mathbb{Z}, +)$ The Fourier series of $f \in L^2([0,1])$ is defined as :*

$$S_n(f) = \sum_{k=-n}^{n} c_k(f)\chi_k$$

Then :

$$||f - S_n(f)||_2^2 = ||f||_2^2 - ||S_n(f)||_2^2$$

Which can be rewritten :

$$||f - \sum_{k=-n}^{n} c_k(f)\chi_k||_2^2 = ||f||_2^2 - \sum_{k=-n}^{n} |c_k(f)|^2$$

Proof. It is sufficient to state that $(\chi_k)_{k\in\mathbb{Z}}$ form an orthonormal basis of the Hilbert space $L^2(0,1)$, and that $S_n(f)$ is the orthogonal projector of f onto $\text{Span}(e_k/k \in [-n,n])$. ☐

Theorem A.2 (Bochner - groups). *Let G be a locally compact abelian group, with dual group $\hat{G}$. For any normalised ($f(1_G) = 1$), continuous, positive-definite function f on G, there exists a unique probability measure μ on $\hat{G}$, such that :*

$$\forall g \in G, f(g) = \int_{\hat{G}} \chi(g) d\mu(\chi)$$

That is, f is the Fourier transform of a probability measure on $\hat{G}$, μ. Conversely, the Fourier transform of a probability measure on $\hat{G}$ is a normalised, continuous positive-definite function.

Corollary A.2.1 (Bochner - kernel). *For a continuous, translation invariant kernel $k(x,y) = k(x-y)$,*

$$k \text{ is positive definite} \Leftrightarrow \forall \Delta \in \mathbb{R}^d, \quad k(\Delta) = \int_{\mathbb{R}_D} p(\omega) \exp^{i\Delta\omega} d\omega$$

where p is a non-negative measure.

Characterising a kernel function (positive definite) with it's Fourier transform underpins scalable approximation techniques such at Random Fourier Features [RR07]. Unfortunately the kernels with either the signature or the path development are not translation invariant so the approximation does not directly apply.

Corollary A.2.2 (Herglotz - autocorrelation functions of stationary time series). *Let $\gamma : \mathbb{Z} \to \mathbb{C}$.*

$$\gamma \text{ semi-positive definite } \Leftrightarrow \exists \mu \in \mathcal{P}([a,b]), \forall h \in \mathbb{Z}, \gamma(h) = \int_0^{2\pi} e^{iwh} \mu(dw)$$

μ is the spectral distribution function and if μ is absolutely continuous with respect to the Lebesgue measure, then $\exists f$ spectral density :

$$\gamma(h) = \int e^{iwh} f(w) \lambda(dw)$$

Theorem A.3 (Parseval identity)**.** *Result on the summability of Fourier series. Equality of energy of periodic signal and it's frequency domain representation. Generalised Pythagorean theorem for inner-product spaces.*

$$||f||^2_{L^2([0,2\pi])} = 2\pi \sum_{n\in\mathbb{Z}} |c_n|^2$$

With :

$$c_n = \frac{1}{2\pi}\int_0^{2\pi} f(x)e^{inx}dx$$

It can be seen as an application of the Plancherel theorem.

A.1.1 Convolution of measures

Let G be a group, μ and ν two distributions on G. The convoluted distribution is a distribution on G defined as :

$$\mu * \nu = m_*(\mu \boxtimes \nu)$$

Where $m : G \times G \to G$ is the multiplication in G and $\boxtimes$ is the Cartesian product.

In this simple case, the convoluted measure can be described by :

$$\mu * \nu(dz) = \int_G \mu(dzy)\nu(d(y^{-1}))$$

Similarly, one can define the convolution of pro-distributions on $\mathcal{S}$ as follows. Let $\mu_\bullet$ and $\nu_\bullet$ be two pro-distributions. The convolution can be defined in the same way :

$$\mu_\bullet * \nu_\bullet = m_*(\mu_\bullet \boxtimes \nu_\bullet)$$

Where the multiplication and Cartesian product are in the algebra $T((V))$. Then, to understand better the action of convoluting one can describe the pushdown with the multiplication writing :

$$\begin{aligned} m^{-1}(d\psi) &= \left\{ X, Y \in T((V)), \forall p \in \mathbb{N},\ \sum_{k=0}^{p} X_k \otimes Y_{p-k} \in d\psi_p \subset V^{\otimes p} \right\} \\ &= \bigcap_{p\in\mathbb{N}} \left\{ X, Y \in T((V)),\ \sum_{k=0}^{p} X_k \otimes Y_{p-k} \in d\psi_p \subset V^{\otimes p} \right\} \end{aligned}$$

The convoluted distribution acts on level p of the tensor algebra in the following way :

$$(\mu_\bullet * \nu_\bullet)_p(d\psi) = \sum_{k=0}^{p} \int_{X^k \otimes Y^{p-k} \in d\psi_p} \mu_k(dX^k)\nu_{p-k}(dY^{p-k})$$

Therefore the proof of the property is a direct application of the definition :

$$\begin{aligned} \hat{\mathcal{F}}(\mu_\bullet).\hat{\mathcal{F}}(\nu_\bullet) = & \left(\sum_{p=0}^{\infty} \sum_{i_1,\ldots,i_p} \left(\int_{\psi\in\mathcal{S}^p} \psi_{i_1,\ldots,i_p}.d\mu_p \right) Z_{i_1}\ldots Z_{i_p} \right) \\ & \otimes \left(\sum_{p=0}^{\infty} \sum_{i_1,\ldots,i_p} \left(\int_{\psi\in\mathcal{S}^p} \psi_{i_1,\ldots,i_p}.d\nu_p \right) Z_{i_1}\ldots Z_{i_p} \right) \\ = & \sum_{p=0}^{\infty} \sum_{i_1,\ldots,i_p} \sum_{k=0}^{p} \left(\int_{\psi\in G_{n,k}(\mathbb{R})} \psi_{i_1,\ldots,i_k}.d\mu_p \right) \\ & \times \left(\int_{\psi\in G_{n,p-k}(\mathbb{R})} \psi_{i_{k+1},\ldots,i_p}.d\nu_{p-k} \right) Z_{i_1}\ldots Z_{i_p} \\ = & \sum_{p=0}^{\infty} \sum_{i_1,\ldots,i_p} \underbrace{\sum_{k=0}^{p} \left(\int_{\psi\in\mathcal{S}^k} \int_{\phi\in\mathcal{S}^{p-k}} \psi_{i_1,\ldots,i_k}\phi_{i_{k+1},\ldots,i_p}.d\mu_p(\psi)d\nu_{p-k}(\phi) \right)}_{\int_{\psi\in\mathcal{S}^p} \psi_{i_1,\ldots,i_p}.d(\mu_p * \nu_p)} Z_{i_1}\ldots Z_{i_p} \\ = & \hat{\mathcal{F}}(\mu_\bullet * \nu_\bullet) \end{aligned}$$

A.2 Topology properties

Definition A.4 (Frechet / T1 space). *Topological space where for each pair of distinct points, each have a neighbourhood that does not contain the other.*

Definition A.5 (Hausdorff / T2 space). *Topological space where for each pair of distinct points, there exist a neighbourhood of each that don't intersect.*

Definition A.6 (Polish space). *Topologocal space that is separable and completely metrizable.*

Definition A.7 (Lusin space). *A Hausdorff space is Lusin if some topology makes it Polish.*

Definition A.8 (Baire space). *The following definitions are equivalent :*

- *For a countable family (U_n) of open dense subsets of X, then $D = \cap U_n$ is also dense in X*
- *If $X = \cup A_n$ with A_n closed subset, then $\exists n_0 \in \mathbb{N}$ such that $int(A_{n_0}) \neq \varnothing$*
- *Any countable union of meagre sets is meagre*

Definition A.9 (Initial topology). *(induced/ weak / limit/ projective) $f_i : X \to Y_i$ is the coarsest topology that makes those functions continuous, eg subspace of product topology.*

Definition A.10 (Final topology). *(strong / inductive) $f_i : Y_i \to X$ on X is the finest / strongest topology that makes those functions continuous, eg quotient topology.*

Definition A.11 (Projective tensor product). *given two locally convex topological vector spaces X and Y, the projective topology on $X \otimes Y$ is the strongest topology that makes it a locally convex topological vector space and such that the canonical map $(x, y) \mapsto x \otimes y$ is continuous. Denoted $X \otimes_\pi Y$*

Theorem A.12 (Heine-Borel). *X subset of a Euclidean space is compact $\Leftrightarrow$ every open cover has a finite subcover*

Other relevant properties

- If X is compact and $C \subset X$ is closed, then it is compact
- If X is Hausdorff and $C \subset X$ is compact then it is closed
- Every metric space is Hausdorff
- If $f : X \to Y$ is continuous and $C \subset X$ is compact, then so is $f(C) \subset Y$
- A space is T_1 if and only if all singletons are closed subspaces, every Hausdorff space is T_1
- Let G be a LCA group. There exists an absorbing exhaustion of G [Dei05]
- Any locally compact Hausdorff space and any complete metric space is a Baire space
- A topological vector space is locally convex if it's topology is induced by a family of semi-norms

A.2.1 Haar integration

Let G be a topological group and $\mathcal{B}$ it's Borel σ-algebra.
A locally finite Borel measure μ on $\mathcal{B}$ is called *outer Radon* if :

- μ is continuous from above with open sets : $\mu(A) = \inf_{A \subset U} \mu(U)$ for any Borel set A, with U open sets.
  This property is also called *outer regularity.*
- μ is continuous from below with compact sets : $\mu(A) = \sup_{K \subset A} \mu(K)$ for any Borel set A, with K compact sets.
  This property is also called *weak inner regularity*

A measure μ is called left-invariant is $\forall A \in \mathcal{B}, \forall x \in G$:

$$\mu(xA) = \mu(A)$$

Theorem A.13 (Existence of a Haar measure)**.** *Let G be a locally compact group. There exists a non-zero left-invariant outer-Radon measure on G, which is uniquely determined up to a positive multiplicative constant.*

See [DE14] for a complete proof. The following example completes those in the aforementioned reference.

Example A.14. *Let* $G = \left\{ \begin{pmatrix} x & z \\ & y \end{pmatrix} / (x, y) \in (\mathbb{R}^*)^2, z \in \mathbb{R}, \right\}$ *The Haar measure can be defined as :*

$$\mu = \frac{dx}{x^2}\frac{dy}{y}dz$$

Indeed it is left-invariant :

$$\begin{aligned}
\mu\left(\begin{pmatrix} x' & z' \\ & y' \end{pmatrix} A\right) &= \iiint_A \begin{pmatrix} x' & z' \\ & y' \end{pmatrix}\begin{pmatrix} x & z \\ & y \end{pmatrix} d^*(x,y,z) \\
&= \iiint_A \begin{pmatrix} xx' & x'z + z'y \\ & yy' \end{pmatrix} \frac{dx}{x^2}\frac{dy}{y}dz \\
&= \iiint_A \begin{pmatrix} x & z \\ & y \end{pmatrix} \frac{dx}{x^2}\frac{dy}{y}dz \\
&= \mu(A)
\end{aligned}$$

Then the outer-Radon property is entailed by the fact that the Lebesgue measure is outer-Radon, and the measure is locally finite on G.
In this case, the volume of a cube is given by :

$$\begin{aligned}
\iiint_{[a,b]^3} d^*(x,y,z) &= \iiint_{[a,b]^3} \frac{dx}{x^2}\frac{dy}{y}dz \\
&= \frac{1}{6a^6b^6}(b-a)^3(b+a)(b^2+ab+a^2)
\end{aligned}$$

Which explodes as a tends to zero.

The Haar measure then defines the *Haar integral.* To name a few, Haar integral satisfies the usual properties of integration :

- Every compact set has finite measure,
- for a positive continuous function, $\int_G f d\mu = 0 \Rightarrow f = 0$ pointwise.

A.3 Algebra properties

The tensor algebra is naturally graded. It has two different coalgebra structures, but only one making it a bialgebra. To complete the definition of the Hopf algebra structure which is to have :

$$\begin{tikzcd}
T(V) \boxtimes T(V) \arrow[rr, "\delta\boxtimes id"] & & T(V) \boxtimes T(V) \arrow[d, "\nabla"] \\
T(V) \arrow[u, "\Delta"] \arrow[d, "\Delta"'] \arrow[r, "\epsilon"'] & \mathbb{C} \arrow[r, "\eta"'] & T(V) \\
T(V) \boxtimes T(V) \arrow[rr, "id\boxtimes\delta"'] & & T(V) \boxtimes T(V) \arrow[u, "\nabla"']
\end{tikzcd}$$

to commute, one defines :

- the coproduct is defined as :
$$\begin{cases} \Delta x = 1 \boxtimes + x \boxtimes 1 \\ \Delta 1 = 1 \boxtimes 1 \end{cases}$$
and extends as an algebra homomorphsim.
- the antipode : defined by induction as $\alpha(1) = 1$ and $\alpha(x) = -x \; \forall x \in V$
- unit $\epsilon : T((V)) \to \mathbb{C}$ the projection on level zero of the tensor algebra.
- counit : the inclusion map of $\mathbb{C}$ into the tensor algebra.

A.3.1 The signature as group-like elements of Hopf algebra

An element is group-like if it verifies :

$$\Delta x = x \boxtimes x$$

Property A.15. *For a path x in Π_n, $S(x)$ is a group-like element.*

A group-like element in the tensor algebra verifies :

$$\forall k \in \mathbb{N}, \ \sum_{p=0}^{k} a^p \boxtimes a^{(k-p)} = \Delta a^k = \Delta a_1^k \otimes \cdots \otimes \Delta a_k^k$$

Expanding the right-hand side gives :

$$\begin{aligned}
\sum_{p=0}^{k} a^p \boxtimes a^{(k-p)} &= \sum_{p=0}^{k} \sum_{\substack{1<i_1<\cdots<k \\ j\neq i}} \left(a_{i_1}^k \otimes \cdots \otimes a_{i_p}^k\right) \boxtimes \left(a_{j_1}^k \otimes \cdots \otimes a_{j_{k-p}}^k\right) \\
&= \sum_{p=0}^{k} \sum_{S\in \mathrm{Sh}(p,k)} \left(\prod_{i\in S} a_i^k\right) \boxtimes \left(\prod_{i\notin S} a_i^k\right) \\
&= \left(a^k \overset{\boxtimes}{\sqcup\!\sqcup} a^k\right)^{(k)}
\end{aligned}$$

Where $\overset{\boxtimes}{\sqcup\!\sqcup}$ defines an outer shuffle product. See [Reu03] proposition 1.10 page 29 for more details.

- At level zero, begin group-like implies $a^0 = 1$.
- At level one, no condition is obtained. For a in $T(V)$, one always has :

$$(a \boxtimes a)^1 = 1 \boxtimes a^1 + a^1 \boxtimes 1 = (\Delta a)^1$$

- At level two, a property similar to a non-commutative integration by parts is entailed :

$$a^1 \boxtimes a^1 = a_2^2 \boxtimes a_1^2 + a_1^2 \boxtimes a_2^2$$

Proof. The result extends the integration by parts formula with the shuffle identity that makes the tensor algebra a Hopf algebra.

Note that the coproduct preserves the natural grading of the tensor algebra. Let $x \in \Pi_n$. Bilinearity of the coproduct yields :

$$\begin{aligned}
S(x) \boxtimes S(x) &= \sum_{k=0}^{\infty} \sum_{j=0}^{k} S^{(j)}(x) \boxtimes S^{(k-j)}(x) \\
&= 1 \boxtimes 1 + x_t \boxtimes 1 + 1 + \boxtimes x_t + S^{(2)}(x) \boxtimes 1 + S^{(1)}(x) \boxtimes S^{(1)}(x) + 1 \boxtimes S^{(2)}(x) + \ldots
\end{aligned}$$

Let $k \in \mathbb{N}$,

$$\sum_{j=0}^{k} S^{(j)}(x) \boxtimes S^{(k-j)}(x) = \sum_{j=0}^{k} \int_{0\leq t_1<\cdots<t_k\leq t} \cdots \int \sum_{S\in \mathrm{Sh}(p,k)} \left(\prod_{i\in S} dx_{t_i}^k\right) \boxtimes \left(\prod_{i\notin S} dx_{t_i}^k\right)$$

Is ensured by [Reu03] $\square$

The set of group-like elements will be denoted $G(V) \subset T(V)$.

The image of the signature $S(\Pi_n)$ is at least dense in the set of group-like elements [CS24].

B Harmonic analysis on algebras

One can start from different points to apply harmonic analysis on paths, given the initial group structure can be represented in the tensor algebra. A celebrated application of Harmonic analysis on algebra is the spectral decomposition of compact linear operators on Hilbert spaces, as a generalisation of matrix eigenvalue decomposition.

Noticeably, the tensor algebra can be turned into a Hilbert canonically if the underlying space is one itself.

Considered normed algebras are :

Space	Operations	Norm
$C(K)$ for K Hausdorff compact	$+,.$	$\|\|.\|\|_\infty$
$\mathcal{B}(H)$ for H hilbert	$+,\circ$	$\|\|.\|\|_{HS}$
$L^1(G)$ for G topological group	$+,*$	$\|\|.\|\|_1$ with Haar measure

Table 3: Example of Banach algebras

Besides, continuous functions on a compact Hausdorff space form an algebra. Thus, one could think of using the tensor algebra as a Banach algebra, or applying harmonic analysis to continuous functions on paths instead of using the group structure of paths or on a representation of the later - which will be explored in the last section.

B.1 Gelfand theory for commutative Banach algebras

Let $\mathcal{A}$ be a commutative unital Banach algebra.

Definition B.1. *For $a \in \mathcal{A}$, the resolvent is :*

$$Res(a) = \{\lambda \in \mathbb{C}/\lambda 1 - a \quad invertible\}$$

It is open in $\mathbb{C}$.

Definition B.2. *For $a \in \mathcal{A}$, the spectrum is defined as :*

$$\sigma_{\mathcal{A}}(a) = \mathbb{C} \backslash Res_{\mathcal{A}}(a)$$

It is closed and bounded by $||a||$.

Theorem B.3 (Gelfand-Mazur)**.** *For unital Banach algebra, all elements invertible implies $\mathcal{A} = \mathbb{C}1$.*

Theorem B.4 (Gelfand-Naimark)**.** *commutative C^*-algebra isomorphic to algebra of continuous functions vanishing at infinity on a locally compact Hausdorff space.*

Property B.5. *$\mathcal{A}^\times$, which is open topological group.*

Definition B.6. *The* **structure space** *of $\mathcal{A}$ is defined as :*

$$\Delta_{\mathcal{A}} = \{m : \mathcal{A} \to \mathbb{C}/ \text{ non-zero continuous algebra homomorphism}\}$$

Also called **maximal ideal space** *.*
$m \in \Delta_{\mathcal{A}}$ is called a **multiplicative linear functional***, $m(1_{\mathcal{A}}) = 1$ necessarily.*

Example B.7.

- *$\mathcal{A} = \mathcal{C}_0(X)$, X Hausdorff, then $\forall x \in X, m_x : f \mapsto f(x)$*
- *$\mathcal{A} = L(G)$, G LCA group, then $\chi \in \hat{G}, m_\chi : f \mapsto \hat{f}(\chi) = \int_{\mathcal{A}} f(x)\overline{\chi(x)}dx$*
- *$\mathcal{A} = T((V))$, V Banach, then $V' \cong \Delta_{\mathcal{A}}$:*

$$\forall \phi \in V', m_\phi : x \in T((V)) \mapsto \sum_{n \geq 0} \prod_{k=1}^{n} \phi(x_k^{(n)}) \in \mathbb{C}$$

This last example will be taken at the end of this section, applying the notorious spectral decomposition on bounded operators. Furthermore, it is also used in [CL16] that will be presented later on.

Having in mind to apply the Stone-Weierstrass theorem, one needs to define a topology on $\Delta_{\mathcal{A}}$ that is weak enough to guarantee continuity but maintaining separability properties.

finding a topology to have $\Delta_{\mathcal{A}}$ locally compact : $m \in \Delta_{\mathcal{A}}$, m continuous and $||m|| \leq 1$, $||m|| = 1$ if $\mathcal{A}$ unital. (Using the topology induced by the dual space norm.

$$\Delta_{\mathcal{A}} \subset \overline{B'} = \{f \in \mathcal{A}'/||f|| \leq 1\} \subset \mathcal{A}'$$

Weak-*-topology : For V Banach, V' defined by $\{\delta_v/v \in V\}$ evaluation functions. Therefore defines pointwise convergence.

Tychonov's theorem then implies :

Theorem B.8 (Banach Alaoglu). *With V a normed space,*

$$\overline{B'} = \{f \in V'/||f|| \leq 1\} \subset V'$$

is a compact Hausdorff space with respect to the weak--topology.*

So $\Delta_{\mathcal{A}}$ with respect to the weak-*-topology is locally compact and Hausdorff.

Theorem B.9 (Gelfand transform). *For $a \in \mathcal{A}$:*

$$\begin{aligned} \hat{a} : \Delta_{\mathcal{A}} &\longrightarrow \mathbb{C} \\ m &\longmapsto \hat{a}(m) = m(a) \end{aligned}$$

The Gelfand transform is the map :

$$a \mapsto \hat{a} \in \mathcal{C}_0(\Delta_{\mathcal{A}})$$

and is an algebra homomorphism.

- *$\mathcal{A}$ unital $\Rightarrow \Delta_{\mathcal{A}}$ compact*
- *$\forall a \in \mathcal{A}$, $\hat{a}$ continuous and vanishes at infinity*
- *$\forall a \in \mathcal{A}, ||\hat{a}||_{\Delta_{\mathcal{A}}} \leq ||a||$ therefore the Gelfand transform is continuous.*

Property B.10 (Maximal ideals). *If $\mathcal{A}$ is commutative, then any non-invertible element of $\mathcal{A}$ lies in some maximal ideal. For $I \subset \mathcal{A}$ a closed ideal, $\mathcal{A}/I$ is still a Banach algebra with $||a + I|| = \inf_{d \in I} ||a + d||$*

Theorem B.11.

- *$m \mapsto ker(m)$ bijection from $\Delta_{\mathcal{A}}$ to the set of maximal ideals of $\mathcal{A}$*
- *$a \in \mathcal{A}$ invertible $\Leftrightarrow m(a) \neq 0, \forall m \in \Delta_{\mathcal{A}}$*
- *$a \in \mathcal{A}, \sigma(a) = Im(\hat{a})$*

Definition B.12 (*-algebra). *A* **Banach *-algebra** *is a Banach algebra with an involution operation *.*
It is a **C-*-algebra** *if $||a^*a|| = ||a||^2$.*

For example $\mathcal{B}(H)$ the space of bounded operators on a Hilbert space. In such space, if $a \in \mathcal{A}$ is self-adjoint, then $r(a) = ||a||$ (in general $r(a) \leq ||a||$).

Property B.13.

- *a Banach-*-algebra is* **symmetric** *if :*

$$m(a^*) = \overline{m(a)} \Leftrightarrow \hat{a^*} = \overline{\hat{a}} \Leftrightarrow m(a) \in \mathbb{R}, \forall a \textit{self-adjoint}$$

- *every commutative C^*-algebra is symmetric*
- *for $\mathcal{A}$ commutative Banach algebra, the image of the Gelfand map $\hat{\mathcal{A}} = \{\hat{a}/a \in \mathcal{A}\}$ striclty separates points of $\Delta_{\mathcal{A}}$*

The later point lays the conditions to apply the Stone-Weierstrass theorem, which is included in the following theorem.

Theorem B.14 (Gelfand-Naimark theorem). *$\mathcal{A}$ symmetric commutative Banach $*$-algebra, then $\hat{\mathcal{A}}$ dense subalgebra of $\mathcal{C}^0(\Delta_{\mathcal{A}})$. If $\mathcal{A}$ commutative C^*-algebra, then Gelfand transform is isometric $*$-isomorphism, $||a|| = ||\hat{a}||_{\Delta_{\mathcal{A}}}$, $m(a^*) = \overline{m(a)}$.*
$\Delta_{\mathcal{A}}$ compact $\Leftrightarrow \mathcal{A}$ unital, then Gelfand transform isomorphism :

$$\mathcal{A} \cong \mathcal{C}(\Delta_{\mathcal{A}})$$

Continuous Functional Calculus Let $\mathcal{A}$ be a C^*-algebra, $a \in \mathcal{A}$ and f continuous on $\sigma(a) = \mathrm{Im}(\hat{a})$. Then $f \circ \hat{a}$ can be seen as belonging to $\mathcal{A}$ and is denoted $f(a)$ ($\mathcal{A} \cong C(\Delta_{\mathcal{A}})$ The map $f \mapsto f(a)$ is called the (continuous) functional calculus for a.

An element a is called **normal** if $a^*a = aa^*$. $C^*(1,a)$ unital C^*-subalgebra generated by a (is smallest), similar definition for $C^*(a)$. Then a is normal $\Leftrightarrow C^*(a)$ is commutative.

Theorem B.15. *Let $\mathcal{A}$ be a unital C^*-algebra, and let $a \in \mathcal{A}$ be normal. Then there exists a unique *-homomorphism $\Phi_a : C(\sigma_{\mathcal{A}}(a) \to \mathcal{A}$ such that $\Phi_a(1_{\sigma(a)}) = 1_{\mathcal{A}}$ and $\Phi_a(Id_{\sigma(a)}) = a$. Denote : $f(a) = \Phi_a(f)$, $\Phi = \Phi_a$ has the following properties :*

- $\Phi(C(\sigma_{\mathcal{A}})) = C^*(1,a) \subset \mathcal{A}$
- $||\Phi(f)|| = ||f||_{\sigma_{\mathcal{A}}(a)}$ *for every* $f \in C(\sigma_{\mathcal{A}}(a))$
- *If $f = \sum c_n z^n$ is absolutely uniformly convergent on $\sigma_{\mathcal{A}}(a)$, then :*

$$f(a) = \sum c_n a^n$$

where the series converges in $\mathcal{A}$

The proof is based on the Stone-Weiertrass and Gelfand-Naimark theorems. An important corollary of this theorem is the following :

Suppose that $a = a^*$ is a self-adjoint element of the C^*-algebra $\mathcal{A}$. Then $\sigma_{\mathcal{A}}(a) \subset \mathbb{R}$.

Property B.16. *Let H be a Hilbert space. The space of bounded operators on H, noted $\mathcal{B}(H)$ is a C^*-algebra.*

The previous results can therefore be applied, to eventually define a spectral decomposition of bounded compact operators.

B.2 Application to paths

One can try applying Gelfand-Naimark theorem to continuous functions on paths. To ensure the supremum norm is finite, one can take a compact subset $K \subset \Pi_n$.

Property B.17. $\mathcal{A} = \big(C^0(K, \mathbb{C}), +, .\big)$ *with the supremum norm on K is a unital C-*-algebra.*

Proof. Taking the product topology on K it is Hausdorff therefore the algebra structure and norm are defined as on any other compact Hausdorff space so it remains to show it is Banach.
Let $(f_n)_{n\in\mathbb{N}}$ be a Cauchy sequence in $\mathcal{A}$. Let $\epsilon > 0$. There exists $n_0 \in \mathbb{N}$ such that $\forall n, m \geq n_0, ||f_n - f_m||_\infty < \epsilon$. For any $\gamma \in K$,

$$|f_n(\gamma) - f_m(\gamma)| \leq \sup_{\eta \in K} |f_n(\eta) - f_m(\eta)| = ||f_n - f_m||_\infty < \epsilon$$

So $(f_n(\gamma))_{n\in\mathbb{N}}$ is a Cauchy sequence in $\mathbb{R}$. Define $f : \gamma \in K \mapsto f(\gamma)$ the map of limits of every path. To complete the proof one needs f continuous on K. Let (γ_n) a sequence in K with limit $\gamma \in K$. Let also $m \geq n_0$ defined above.

$$\begin{aligned}|f(\gamma_n) - f(\gamma)| \leq& |f(\gamma_n) - f_m(\gamma_n)| + |f_m(\gamma_n) - f_m(\gamma)| + |f_m(\gamma) - f(\gamma)| \\ \leq& 2||f - f_m||_\infty + |f_m(\gamma_n) - f_m(\gamma)| \\ \leq& 3\epsilon\end{aligned}$$

Given f_m is continuous for every m.
Then the unit is the constant function 1_K and the involution is the complex conjugate. □

$$T_\mu : x \mapsto \mathbb{E}_{y\sim\mu}\left[S(y)\right] x$$

Is a bounded normal linear operator on $T((V))$, but is not compact.

Eventually given a locally compact group G, one can associate the $C - *$ algebra $C^*(G)$ defined as the completion of $L^1(G)$ with the supremum norm on representations :

$$||f||_{C^*} = \sup\left\{||\pi(f)||/\pi \;\; \text{unitary on} \;\; G\right\}$$

References

[AB12] Pascal Auscher and Lashi Bandara. *Real harmonic analysis*. ANU Press, 2012.

[Bah+23] Hajer Bahouri et al. "Spectral summability for the quartic oscillator with applications to the Engel group." In: *Journal of Spectral Theory* 13.2 (2023).

[Bau04] Fabrice Baudoin. *An introduction to the geometry of stochastic flows*. World Scientific, 2004.

[CS24] Thomas Cass and Cristopher Salvi. *Lecture notes on rough paths and applications to machine learning*. 2024. arXiv: 2404.06583.

[CT24a] Thomas Cass and William F Turner. "Free probability, path developments and signature kernels as universal scaling limits". In: *arXiv preprint arXiv:2402.12311* (2024).

[CT24b] Thomas Cass and William F Turner. "Topologies on unparameterised path space". In: *Journal of Functional Analysis* 286.4 (2024), p. 110261.

[Che54] Kuo-Tsai Chen. "Iterated integrals and exponential homomorphisms". In: *Proceedings of the London Mathematical Society* 3.1 (1954), pp. 502–512.

[CL16] Ilya Chevyrev and Terry Lyons. "Characteristic functions of measures on geometric rough paths". In: (2016).

[CO18] Ilya Chevyrev and Harald Oberhauser. "Signature moments to characterize laws of stochastic processes". In: *arXiv preprint arXiv:1810.10971* (2018).

[Dei05] Anton Deitmar. *A first course in harmonic analysis*. Springer, 2005.

[DE14] Anton Deitmar and Siegfried Echterhoff. *Principles of harmonic analysis*. Springer, 2014.

[Faw03] Thomas Fawcett. "Problems in stochastic analysis. Connections between rough paths and non-commutative harmonic analysis". PhD thesis. University of Oxford, 2003.

[Gre+07] Arthur Gretton et al. "A kernel statistical test of independence". In: *Advances in neural information processing systems* 20 (2007).

[Ise01] Arieh Iserles. "On Cayley-transform methods for the discretization of Lie-group equations". In: *Foundations of Computational Mathematics* 1 (2001), pp. 129–160.

[Kap09] Mikhail Kapranov. *Noncommutative geometry and path integrals*. Springer, 2009.

[Kru83] Joseph B Kruskal. "An overview of sequence comparison: Time warps, string edits, and macromolecules". In: *SIAM review* 25.2 (1983), pp. 201–237.

[Lem+21] Maud Lemercier et al. "Distribution regression for sequential data". In: *International Conference on Artificial Intelligence and Statistics*. PMLR. 2021, pp. 3754–3762.

[Lez19] Mario Lezcano Casado. "Trivializations for gradient-based optimization on manifolds". In: *Advances in Neural Information Processing Systems* 32 (2019).

[LLN22] Hang Lou, Siran Li, and Hao Ni. "Path development network with finite-dimensional Lie group representation". In: *arXiv preprint arXiv:2204.00740* (2022).

[LLN24] Hang Lou, Siran Li, and Hao Ni. "PCF-GAN: generating sequential data via the characteristic function of measures on the path space". In: *Advances in Neural Information Processing Systems* 36 (2024).

[LX17] Terry J Lyons and Weijun Xu. "Hyperbolic development and inversion of signature". In: *Journal of Functional Analysis* 272.7 (2017), pp. 2933–2955.

[LX18] Terry J Lyons and Weijun Xu. "Inverting the signature of a path". In: *Journal of the European Mathematical Society* 20.7 (2018), pp. 1655–1687.

[McK24] Henry P McKean. *Stochastic integrals*. Vol. 353. American Mathematical Society, 2024.

[AH10] Awad H Al-Mohy and Nicholas J Higham. "A new scaling and squaring algorithm for the matrix exponential". In: *SIAM Journal on Matrix Analysis and Applications* 31.3 (2010), pp. 970–989.

[Pan82] Pierre Pansu. "Geometrie du group d'Heisenberg". In: *These de doctorat, Université Paris VII* (1982).

[RR07] Ali Rahimi and Benjamin Recht. "Random features for large-scale kernel machines". In: *Advances in neural information processing systems* 20 (2007).

[Reu03] Christophe Reutenauer. “Free lie algebras”. In: *Handbook of algebra*. Vol. 3. Elsevier, 2003, pp. 887–903.

[SS18] Carl-Johann Simon-Gabriel and Bernhard Schölkopf. “Kernel distribution embeddings: Universal kernels, characteristic kernels and kernel metrics on distributions”. In: *Journal of Machine Learning Research* 19.44 (2018), pp. 1–29.

[Tay86] Michael Eugene Taylor. *Noncommutative harmonic analysis*. 22. American Mathematical Soc., 1986.

[Voi91] Dan Voiculescu. “Limit laws for random matrices and free products”. In: *Inventiones mathematicae* 104.1 (1991), pp. 201–220.

[Wil95] Michel Willem. *Analyse harmonique réelle*. 1995.

[YK09] Lexiang Ye and Eamonn Keogh. “Time series shapelets: a new primitive for data mining”. In: *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*. 2009, pp. 947–956.