UNIQUENESS FOR THE SIGNATURE OF A PATH OF BOUNDED VARIATION AND THE REDUCED PATH GROUP

B.M. HAMBLY AND TERRY J. LYONS

Abstract. We introduce the notions of tree-like path and tree-like equivalence between paths and prove that the latter is an equivalence relation for paths of finite length. We show that the equivalence classes form a group with some similarity to a free group, and that in each class there is a unique path that is tree reduced. The set of these paths is the Reduced Path Group. It is a continuous analogue of the group of reduced words. The signature of the path is a power series whose coefficients are certain tensor valued definite iterated integrals of the path. We identify the paths with trivial signature as the tree-like paths, and prove that two paths are in tree-like equivalence if and only if they have the same signature. In this way, we extend Chen's theorems on the uniqueness of the sequence of iterated integrals associated with a piecewise regular path to finite length paths and identify the appropriate extended meaning for parameterisation in the general setting. It is suggestive to think of this result as a non-commutative analogue of the result that integrable functions on the circle are determined, up to Lebesgue null sets, by their Fourier coefficients. As a second theme we give quantitative versions of Chen's theorem in the case of lattice paths and paths with continuous derivative, and as a corollary derive results on the triviality of exponential products in the tensor algebra.

1. Introduction

1.1. **Paths with finite length.** Paths, that is to say (right) continuous functions γ mapping a non-empty interval $J \subset \mathbb{R}$ into a topological space V, are fundamental objects in many areas of mathematics, and capture the concept of an ordered evolution of events.

If (V, d_V) is a metric space, then one of γ's most basic properties is its length $|\gamma|_J$. This can be defined as

$$|\gamma|_J := \sup_{\mathcal{D} \subset J} \sum_{\substack{t_i \in \mathcal{D} \\ i \neq 0}} d_V\left(\gamma_{t_{i-1}}, \gamma_{t_i}\right)$$

where the supremum is taken over all finite partitions $\mathcal{D} = \{t_0 < t_1 < \cdots < t_r\}$ of the interval J. It is clear that $|\gamma|$ is positive (although possibly infinite) and independent of the parameterisation for γ. Any continuous path of finite length can

Date: July, 2006.
2000 *Mathematics Subject Classification.* Primary 70G45, 16S10; Secondary 11L07, , 16S30, 16S32, 16S34, 93C15.
Key words and phrases. non-linear exponential, reduced path group, bounded variation, signature, Chen series, Magnus series, classifying space, paths of finite length, Cartan development.
The authors gratefully acknowledge EPSRC support: GR/R29628/01, GR/S18526/01.

always be parameterised to have unit speed by letting $\tau(t) = |\gamma|_{[0,t]}$ and setting $\eta(\tau(t)) = \gamma(t)$.

Paths of finite length are often said to be those of bounded or finite variation. We denote the set of paths of bounded variation by BV, BV-paths with values in V by $BV(V)$, and those defined on J by $BV(J,V)$.

We first observe any path $\gamma \in BV(J,V)$ can be factored into two paths by splitting γ at a point in J. If V is a vector space, then we can go the other way in that for any $\gamma \in BV([0,t],V)$ and $\tau \in BV([0,s],V)$ we can form the concatenation $\gamma * \tau \in BV([0,s+t],V)$

$$\begin{aligned}\gamma * \tau(u) &= \gamma(u), \ u \in [0,s] \\ \gamma * \tau(u) &= \tau(u-s) + \gamma(s) - \tau(0), \ u \in [s,s+t].\end{aligned}$$

The operation * is associative, and if $V$is a normed space, then $|\gamma| + |\tau| = |\gamma * \tau|$.

1.2. **Differential Equations.** One reason for looking at $BV(V)$ is that one can do calculus with these paths, while at the same time the set of paths with $|\gamma|_J \leq l$ is closed under the topology of pointwise convergence (uniform convergence, ...). Differential equations allow one to express relationships between paths in BV. If f^i are Lipschitz vector fields on a space W and $\gamma_t = (\gamma_1(t), \ldots, \gamma_d(t)) \in BV(\mathbb{R}^d)$, then the differential equation

$$\begin{aligned}\frac{dy}{dt} &= \sum_i f^i \frac{d\gamma_i}{dt} = f(y) \cdot \frac{d\gamma}{dt}, \\ y_0 &= a\end{aligned} \tag{1.1}$$

has a unique solution for each γ. BV is a natural class here, for unless the vector fields commute, there is no meaningful way to make sense of this equation if the path γ is only assumed to be continuous.

If (y,γ) solves the differential equation and $(\tilde{y},\tilde{\gamma})$ are simultaneous reparameterisations, then they also solve the equation and so it is customary to drop the dt and write

$$dy = \sum_i f^i \dot{\gamma}_i dt = \sum_i f^i d\gamma_i = f(y) \cdot d\gamma.$$

We can regard the location of y_s as a variable and consider the diffeomorphism π_{st} defined by $\pi_{st}(y_s) := y_t$. Then π_{st} is a function of $\gamma|_{[s,t]}$. One observes that the map $\gamma|_{[s,t]} \to \pi_{st}$ is a homomorphism from $(BV(V), *)$ to the group of diffeomorphisms of the space W.

1.3. **Iterated integrals and the signature of a path.** One could ask which are the key features of $\gamma|_{[s,t]}$, which, with y_s, accurately predict the value y_t in equation (1.1). The answer to this question can be found in a map from BV into the free tensor algebra!

Definition 1.1. Let γ be a path of bounded variation on $[S,T]$ with values in a vector space V. Then its signature is the sequence of definite iterated integrals

$$\begin{aligned}\mathbf{X}_{S,T} &= \left(1 + X^1_{S,T} + \ldots + X^k_{S,T} + \ldots\right) \\ &= \left(1 + \int_{S<u<T} d\gamma_u + \ldots + \int_{S<u_1<\ldots<u_k<T} d\gamma_{u_1} \otimes \ldots \otimes d\gamma_{u_k} + \ldots\right)\end{aligned}$$

regarded as an element of an appropriate closure of the tensor algebra $T(V) = \bigoplus_{n=0}^{\infty} V^{\otimes n}$.

The signature is the *definite* integral over the fixed interval where γ is defined; re-parameterising γ does not change its signature. The first term $X^1_{[S,T]}$ produces the path γ (up to an additive constant). For convenience of notation, when we have many paths, we will sometimes use a symbol such as Y_t (instead of γ_t) for our path, $Y^i_{S,T}$ for the tensor coordinate of degree i of the signature of Y_t, and $\mathbf{Y}_{S,T}$ for the signature of the path. In some circumstances we will drop the time interval and just write Y for the path and $\mathbf{Y}$ for its signature. We call this map the signature map and sometimes denote it by $S : Y \to S(Y)$ when this helps our presentation.

The signature of a path X is a natural object to study. The map $X \to \mathbf{X}$ is a homomorphism (c.f. Chen's identity [7]) from the monoid of paths with concatenation to (a group embedded in) the algebra $T(V)$. The signature $\mathbf{X}(= \mathbf{X}_{0,T})$ can be computed by solving the differential equation

$$\begin{aligned} d\mathbf{X}_{0,u} &= \mathbf{X}_{0,u} \otimes dX_u \\ \mathbf{X}_{0,0} &= (1,0,0,\ldots), \end{aligned}$$

and, in particular, paths with different signatures will have different effects for some choice of differential equation.

There is a converse, although this is a consequence of our main theorem. If X controls a system through a differential equation

$$\begin{aligned} dY_u &= f(Y_u)dX_u, \\ Y_0 &= a, \end{aligned}$$

and f is Lipschitz, then the state Y_T of the system after the application of $X|_{[0,T]}$ is completely determined by the signature $\mathbf{X}_{0,T}$ and Y_0. In other words the signature $\mathbf{X}_{S,T}$ is a truly fundamental representation for the bounded variation path defined on $[S,T]$ that captures its effect on any non-linear system.

This paper explores the relationship between a path and its signature. We determine a precise *geometric* relation $\sim$ on bounded variation paths. We prove that two paths of finite length are $\sim$-equivalent if and only if they have the same signature:

$$X|_J \sim Y|_K \iff \mathbf{X}_J = \mathbf{Y}_K$$

and hence prove that $\sim$ is an equivalence relation and identify the sense in which the signature of a path determines the path.

The first detailed studies of the iterated integrals of paths are due to K. T. Chen. In fact Chen [2] proves the following theorems which are clear precursors to our own results:

Chen Theorem 1: Let $d\gamma_1, \cdots, d\gamma_d$ be the canonical 1-forms on $\mathbb{R}^d$. If $\alpha, \beta \in [a,b] \to R^d$ are irreducible piecewise regular continuous paths, then the iterated integrals of the vector valued paths $\int_{\alpha(0)}^{\alpha(t)} d\gamma$ and $\int_{\beta(0)}^{\beta(t)} d\gamma$ agree if and only if there exists a translation T of $\mathbb{R}^d$, and a continuous increasing change of parameter $\lambda : [a,b] \to [a,b]$ such that $\alpha = T\beta\lambda$.

Chen Theorem 2: Let G be a Lie group of dimension d, and let $\omega_1 \cdots \omega_d$ be a basis for the left invariant 1-forms on G. If $\alpha, \beta \in [a,b] \to G$ are irreducible piecewise regular continuous paths, then the iterated integrals of the vector valued paths $\int_{\alpha(0)}^{\alpha(t)} d\omega$ and $\int_{\beta(0)}^{\beta(t)} d\omega$ agree if and only if there exists a translation T of G, and a continuous increasing change of parameter $\lambda : [a,b] \to [a,b]$ such that $\alpha = T\beta\lambda$.[1]

[1]We borrow these formulations from the Math Review of the paper [2] but include the precise smoothness assumptions.

In particular, Chen characterised piecewise regular paths in terms of their signatures.

1.4. **The main results.** There are two essentially independent goals in this paper.

(1) To provide quantitative versions of some of Chen's results. If γ is continuous, of bounded variation and parameterised at unit speed, then we will obtain lower bounds on the coefficients in the signature in terms of the modulus of continuity of γ' and the length of the path and show how one can recover the length of a path γ using the asymptotic magnitudes of these coefficients (c.f. Tauberian theorems in Fourier Analysis).
(2) To prove a uniqueness theorem characterising paths of bounded variation in terms of their signatures (c.f. the characterisation of integrable functions in terms of their Fourier series) extending Chen's theorem to the bounded variation setting.

For our first goal we provide quantitative versions of Chen's results in two settings. In the discrete setting we can show the following result.

Theorem 1. *If the path in the two dimensional integer lattice corresponding to a given word of length L has its first $\lfloor e\log(1+\sqrt{2})L\rfloor$ $GL(2,\mathbb{C})$-iterated integrals zero, then all its iterated integrals (in the tensor algebra) are zero and, in its reduced form, the word is trivial.*

A key tool is the development of paths into hyperbolic space. We also obtain a continuous version of Theorem 1.

Theorem 2. *Let γ be a path of length l parameterised at unit speed. Suppose that the modulus of continuity δ for γ' is continuous. There is an integer $N(l,\delta)$ such that at least one of the first $N(l,\delta)$ terms in the signature must be non-zero.*

These results represent a sort of rigidity; the first result concerns the complexity of algebraic exponential products, the second result shows that any path with bounded local curvature and with the first N terms in the signature zero must be rather long or trivial. We can obtain explicit bounds depending only on curvature estimates and N. However, our result is far from sharp: consider the figure of eight, it is a path with curvature at most 4π and length one. It is clear that the first two terms in its signature are zero. Our results indicate that it cannot have all of the first 115 terms in the signature zero!

In the continuous setting, and under quite weak hypotheses, one can recover the length of the path from the asymptotic behaviour of $\|X_k\|$ where X is the signature. Even in the case where path is in $\mathbb{R}^d$ the term X_k lies in a kd dimensional space and the magnitude of $\|X_k\|$ depends on the choice of cross norm used on the tensor product. Interestingly there is a phase transition. For the biggest (the projective) norm, and for some others, we see the value of $\lim_{k\to\infty}\left\|l^{-k}k!X_k\right\|=1$ while for other cross norms such as the Hilbert Schmidt (or Hilbert space) norm this limit still exists but is in general strictly less than one:

Theorem 3. *If γ is a C^3 path of length l parameterised at unit speed with signature $X=(1,X_1,\ldots,)$, and if $V^{\otimes k}$ is given the projective norm, then*

$$\lim_{k\to\infty}\left\|l^{-k}k!X_k\right\|=1.$$

If $V^{\otimes k}$ is given the Hilbert Schmidt norm, then the limit

$$\lim_{k\to\infty} \left\| l^{-k} k! X_k \right\|^2 = \mathbb{E}\left(\exp\left(\int_{s\in[0,1]} \left| W_s^0 \right|^2 \langle \gamma'(s), \gamma'''(s) \rangle ds \right) \right) \leq 1$$

exists, where W_s^0 is a Brownian bridge starting at zero and finishing at zero at time 1[2]*. The inequality is strict if γ is not a straight line.*

In Theorem 9 of Section 3 we strengthen Theorem 3 by showing $\lim_{k\to\infty} \left\| l^{-k} k! X_k \right\| = 1$ for a wide class of tensor norms, including the projective norm and for paths γ satisfying a Hölder condition on their derivative. Under even weaker assumptions on γ we can still recover the length of the path by considering $\left\| l^{-k} k! X_k \right\|^{1/k}$.

For the second goal of the paper we need a notion of tree-like path. Our definition codes R-trees by positive continuous functions on the line, as developed, for instance, in [5].

Definition 1.2. $X_t,\ t \in [0,T]$ is a tree-like path in V if there exists a positive real valued continuous function h defined on $[0,T]$ such that $h(0) = h(T) = 0$ and such that

$$\|X_t - X_s\|_V \leq h(s) + h(t) - 2 \inf_{u\in[s,t]} h(u).$$

The function h will be called a height function for X. We say X is a Lipschitz tree-like path if h can be chosen to be of bounded variation.

Definition 1.3. Let $X, Y \in BV(V)$. We say $X \sim Y$ if the concatenation of X and Y 'run backwards' is a Lipschitz tree-like path.

We now focus on $\mathbb{R}^d$ and state our main results.

Theorem 4. *Let $X \in BV(\mathbb{R}^d)$. The path X is tree-like if and only if the signature of X is $\mathbf{0} = (1, 0, 0, \dots)$.*

As the map $X \to \mathbf{X}$ is a homomorphism, and running a path backwards gives the inverse for the signature in $T(V)$, an immediate consequence of Theorem 4 is

Corollary 1.4. *If $X, Y \in BV(\mathbb{R}^d)$, then $\mathbf{X} = \mathbf{Y}$ if and only if the concatenation of X and 'Y run backwards' is a Lipschitz tree-like path.*

Corollary 1.5. *For $X, Y \in BV(\mathbb{R}^d)$ the relation $X \sim Y$ is an equivalence relation. Concatenation respects $\sim$ and the equivalence classes Σ form a group under this operation.*

There is an analogy between the space of paths of finite length in $\mathbb{R}^d$ and the space of words $a^{\pm 1} b^{\pm 1} \dots c^{\pm 1}$ where the letters $a, b, \dots, c$ are drawn from a d-letter alphabet A. Every such word has a unique reduced form. This reduction respects the concatenation operation and projects the space of words onto the free group. We extend this result from paths on the integer lattice (words) to the bounded variation case.

Corollary 1.6. *For any $X \in BV(\mathbb{R}^d)$ there exists a unique path of minimal length, $\bar{X}$, called the reduced path, with the same signature $\mathbf{X} = \bar{\mathbf{X}}$.*

---

[2] This result was suggested to us by the referee

Taking these results together we see that the reduced paths form a group. The multiplication operation is to concatenate the paths and then reduce the result. One should note that this reduction process is not unique (although we have proved that the reduced word one ultimately gets is). This group is at the same time rather natural and concrete (a collection of paths of finite length), but also very different to the usual finite dimensional Lie groups. It admits more than one natural topology, and multiplication is not continuous for the topology of bounded variation.

We can restate these results in different language. The space BV with $*$, the operation of concatenation, is a monoid. Let $\mathcal{T}$ be the set of tree-like paths in BV. Then $\mathcal{T}$ is also closed under concatenation. If $\gamma \in BV$, and we use the notation γ^{-1} for γ run backwards, it is clear from the definition that $\gamma^{-1}\mathcal{T}\gamma \subset \mathcal{T}$ for all $\gamma \in BV$. As we have proved that tree-like equivalence is an equivalence relation, $BV/\mathcal{T}$ is well defined, closed under multiplication, has inverses and is thus a group.

We have the following picture

$$0 \to \mathcal{T} \to BV \overset{\dashleftarrow}{\to} \Sigma \to 0$$

where one can regard Σ as the $\sim$-equivalence classes of paths or as the subgroup of the tensor algebra. The map $\dashleftarrow$ takes the class to the reduced path which is an element of BV. As $\mathcal{T}$ has no natural BV-normal sub-monoids, one should expect that any continuous homomorphism of BV into a group will factor through Σ if it is trivial on the tree-like elements. It is clear that the set

$$\hat{\mathcal{T}} = \{(\gamma, h), \gamma \in \mathcal{T},\ h \text{ a height function for } \gamma\}$$

is contractible. An interesting question is whether $\mathcal{T}$ itself is contractible.

We prove in Lemma 6.3 that any $\gamma \in \mathcal{T}$ is the limit of weakly piecewise linear tree-like paths and hence $\mathcal{T}$ is the *smallest* multiplicatively closed and topologically closed set containing the trivial path. This universality suggests that Σ has similarities to the free group. One characterising property of the free group is that every function from the alphabet A into a group can be extended to a map from words made from A into paths in the group. The equivalent map for bounded variation paths is Cartan development. Let θ be a linear map of $\mathbb{R}^d$ to the Lie algebra $\mathfrak{g}$ of a Lie group G and let $X_t|_{t\leq T}$ be a bounded variation path. Then Cartan development provides a canonical projection of $\theta(X)$ to a path Y starting at the origin in G and we can define $\tilde{\theta} : X \to Y_T$. This map $\tilde{\theta}$ is a homomorphism from Σ to G.

It is an exercise to prove that this map $\tilde{\theta}$ takes all tree-like paths to the identity element in the group G. As a consequence, $\tilde{\theta}$ is a map from paths of finite variation to G which is constant on each $\sim$ equivalence class and so defines a map from Σ to G.

Let $X_t|_{t\leq T}$ be a path of bounded variation in $\mathbb{R}^d$ and suppose that for every linear map θ into a Lie algebra $\mathfrak{g}$, $\tilde{\theta}(X)$ is trivial. As the computation of the first n terms in the signature is itself a development (into the free n-step nilpotent group) we conclude that $\mathbf{X}_{0,T} = (1, 0, 0, \ldots)$ and so from Theorem 4 X is tree-like.

Corollary 1.7. *A path of bounded variation is tree-like if and only if its development into every finite dimensional Lie group is trivial.*

The observation that any linear map of $\mathbb{R}^d$ to the Lie algebra $\mathfrak{g}$ defines a map from Σ to the Lie group is a universal property of a kind giving further evidence that Σ is some sort of continuous analogue of the free group. However, Σ is not a Lie group although it has a Lie algebra and it is not characterised by this universal property. (Chen's piecewise regular paths provide another example since they are paths of bounded variation and are dense in the unit speed paths of finite length).

1.5. **Questions and Remarks.** How important to these results is the condition that the paths have finite length? Does anything survive if one only insists that the paths are continuous?

The space of continuous paths with the uniform topology is another natural generalisation of words - certainly concatenation makes them a monoid. However, despite their popularity in homotopy theory, it seems possible that no natural closed equivalence relation could be found on this space that transforms it into a continuous 'free group' in the sense we mapped out above. The notion of tree-like makes good sense (one simply drops the assumption that the height function h is Lipschitz). With this relaxation,

Problem 1.8. *Does $\sim$ define an equivalence relation on continuous paths?*

Homotopy is the correct deformation of paths if one wants to preserve the line integral of a path against a closed one-form. On the other hand tree-like equivalence is the correct deformation of paths if one wants to preserve the line integral of a path against any one form. As we mention elsewhere in this paper, integration of continuous functions against general one forms makes little sense. This is perhaps evidence to suggest the answer to the problem is in the negative. The difficulty lies in establishing the transitivity of the relation.

Problem 1.9. *Is there a unique tree reduced path associated to any continuous path?*

For smooth paths $\gamma = (\gamma_1, \gamma_2)$ in $\mathbb{R}^2$, Cartan development into the Heisenberg group is the map $(\gamma_1, \gamma_2) \to \left(\gamma_1, \gamma_2, \int \gamma_1 d\gamma_2\right)$. One knows [9, Proposition 1.29] that there is no continuous bilinear map extending this definition to any Banach space of paths which carries the Wiener measure. We also know from Levy, that there are many "almost sure" constructions for this integral made in similar ways to "Levy area". All are highly discontinuous and can give different answers for the same Brownian path in $\mathbb{R}^2$. This wide choice for the case of Brownian paths (which have finite p-variation for every $p > 2$) makes it clear there cannot be a canonical development for all continuous paths.

The paper [7] sets out a close relationship between differential equations, the signature, and the notion of a geometric rough path. These "paths" also form a monoid under concatenation. Also, as any linear map from $\mathbb{R}^d$ into the $(p+\varepsilon)$-Lipschitz vector fields on a manifold M induces a canonical homomorphism of the p-rough paths with concatenation into the group of diffeomorphisms of M, they certainly have the analogy to the Cartan development property. Similarly, every rough path has a signature, and the map is a homomorphism.

Problem 1.10. *Given a path γ of finite p-variation for some $p > 1$, is the triviality of the signature of γ equivalent to the path being tree-like?*

Our theorem establishes this in the context of $p = 1$ or bounded variation paths but our proof uses the one dimensionality of the image of the path in an essential way. An extension to p-rough paths with $p > 1$ would require new ideas to account for the fact that these rougher paths are of higher "dimension".

There seem to be many other natural questions.

By Corollary 1.6, among paths of finite length with the same signature, there is a unique shortest one - the reduced path. Successful resolution of the following question could have wide ramifications in numerical analysis and beyond. The question is interesting even for lattice paths.

Problem 1.11. *How does one effectively reconstruct the reduced path from its signature?*

A related question is to:

Problem 1.12. *Identify those elements of the tensor algebra that are signatures of paths and relate properties of the paths (for example their smoothness) to the behaviour of the coefficients in the signature.*

Some interesting progress in this direction can be found in [3]. We conclude with some wider comments.

(1) There is an obvious link between these reduced paths and geometry since each connection defines a closed subgroup of the group of reduced paths (the paths whose developments are loops).
(2) It seems natural to ask about the extent to which the intrinsic structure of the space of reduced paths (with finite length) in $d \geq 2$ dimensions changes as d varies.

1.6. **Outline.** We begin in Section 2 by discussing the lattice case. In this setting we can obtain our first quantitative result on the signature, Theorem 1. The case of words in d generators is also treated and if the first $c(d)L$ terms in the signature are zero, the word is reducible, where the constant $c(d)$ grows logarithmically in d.

In Section 3 we extend these quantitative estimates to finite length paths. In order to do this we need to discuss the development of a path into a suitable version of hyperbolic space - a technique that has more recently proved useful in [8]. Using this idea we obtain a quantitative estimate on the difference between the length of the developed path and its chord in terms of the modulus of continuity of the derivative of the path. This allows us to obtain, in the case where the derivative is continuous, some estimates on the coefficients in the signature and also shows how to recover the length of the path from the signature.

After this we return to the proof of our uniqueness result, the extension of Chen's theorem. Our proof relies on various analytic tools (the Lebesgue differentiation theorem, the area theorem), and particularly we introduce a mollification of paths that retain certain deeply non-linear properties of these paths to reduce the problem to the case where γ is piecewise linear. Piecewise linear paths are irreducible piecewise regular paths in the sense of Chen and thus the result follows from Chen's Theorem. The quantitative estimates we obtained give an independent proof for this piecewise linear result.

In Section 4 we establish the key properties for tree-like paths that we need. In Section 5 we prove that any path $X_t|_{t\in[0,T]} \in BV$ and with trivial signature can, after re-parameterisation, be uniformly approximated by (weakly) piecewise linear paths with trivial signature. This is an essentially non-linear result as the constraint of trivial signature corresponds to an infinite sequence of polynomial constraints of increasing complexity. In Section 6 we show that, by our quantitative version of Chen's theorem, such piecewise linear paths must be reducible and so tree-like in our language.

This certainly gives us enough to show, in Section 7, that any weakly piecewise linear path with trivial signature is tree-like. It is clear from the definitions that uniform limits of tree-like paths with uniformly bounded length are themselves tree-like. Thus, applying the results of Section 5, the argument is complete. Finally, in Section 8, we draw together the results to give the proofs of Theorem 4 and its Corollaries.

2. Paths on the integer lattice

2.1. **A discrete case of Chen's theorem.** Consider an alphabet A and new letters $A^{-1} = \left\{a^{-1}, a \in A\right\}$. Let Ω be the set of words in $A \cup A^{-1}$. Then Ω has a natural multiplication (concatenation) and an equivalence relation that respects this multiplication.

Definition 2.1. A word $w \in \Omega$ is reducible to the empty word if, by applying successive applications of the rule

$$a \ldots bcc^{-1}d \ldots e \to a \ldots bd \ldots e, \qquad a, b, c, d, e, \ldots \in A \cup A^{-1}$$

one can reduce w to the empty word. We will say that $(a \ldots b)$ is equivalent to $(e \ldots f)$

$$(a \ldots b) \sim (e \ldots f)$$

if $\left(a \ldots bf^{-1} \ldots e^{-1}\right)$ is reducible to the empty word.

An easy induction argument shows that $\sim$ is an equivalence relation. It is well known that the free group F_A can be identified as $\Omega/\sim$. There is an obvious bijection between words in Ω, and lattice paths, that is to say the piecewise linear paths x_u which satisfy $x_0 = 0$ and $\|x_k - x_{k+1}\| = 1$, are linear on each interval $u \in [k, k+1]$, and have $x_k \in \mathbb{Z}^{|A|}$ for each k. The length of the path is an integer equal to the number of letters in the word. The equivalence relation between words can be re-articulated in the language of lattice paths: Consider two lattice paths x and y, and let z be the concatenation of x with y traversed backwards. Clearly, if x and y are equivalent then, keeping its endpoints fixed, z can be "retracted" step by step to a point while keeping the deformations inside what remains of the graph of z. The converse is also true: if U is the universal cover of the lattice, and we identify based path segments in the lattice with points in U, then the words equivalent to the empty word correspond with paths x in the lattice that lift to loops in U. They are the paths that can be factored into the composition of a loop in a tree with a projection of that tree into the lattice. A loop in a tree is a tree-like path, as one can use the distance from the basepoint of the loop as a height function.

Chen's theorem tells us that any path that is not retractable to a point in the sense of the previous paragraph has a non-trivial signature. Our quantitative approach allows us to prove an algebraic version of this result.

Let γ_w be the lattice path associated to the word $w = a_1 \ldots a_L$, where $a_i \in A \cup A^{-1}, i = 1, \ldots, L$. As the signature is a homomorphism, we have $S\left(\gamma_w\right) = S\left(\gamma_{a_1}\right) \ldots S\left(\gamma_{a_L}\right)$. Since γ_{a_i} is a path that moves a unit in a straight line in the a_i direction, its signature is the exponential and $S\left(\gamma_w\right) = e^{a_1} \otimes \ldots \otimes e^{a_L}$.

Our quantitative approach will show in Theorem 5 that for a word of length L in a two letter alphabet, if

$$e^{a_1} \otimes \ldots \otimes e^{a_L} = \left(1, 0, 0, \ldots, 0, X^{N(L)+1}, X^{N(L)+2} \ldots\right),$$

where $N(L) = \lfloor e \log\left(1+\sqrt{2}\right) L \rfloor$, then there is an i for which $a_i = a_{i+1}^{-1}$ and by induction the reduced word is trivial.

The proof is based on regarding $\mathbb{R}^d$ as the tangent space to a point in d-dimensional hyperbolic space $\mathbb{H}$, scaling the path γ and developing it into hyperbolic space. There are two ways to view this development of the path, one of which yields analytic information out of the iterated integrals, the other geometric information. Together they quickly give the result. We work in two dimensional hyperbolic space and, at the end, show that the general case can be reduced to this one.

2.2. **The universal cover as a subset of $\mathbb{H}$.** Let X be a lattice path in $\mathbb{R}^2$, $\theta \geq 0$, and $X^\theta = \theta X$ be the re-scaled lattice path. The development Y^θ of X^θ into $\mathbb{H}$ is easily described. It moves along successive geodesic segments of length θ in $\mathbb{H}$; each time X^θ turns a corner, so does Y^θ, and angles are preserved.

For a fixed choice of θ we can trace out in $\mathbb{H}$ the four geodesic segments from the origin, the three segments out from each of these, and the three from each of these, and so on. It is clear that if the scale θ is large enough, the negative curvature forces the image to be a tree. This will happen exactly when the path that starts by going along the real axis and then always turns anti-clockwise never hits its reflection in the line $x = y$.

The successive moves can be expressed as iterations of a Mobius transform,

$$m(x) := \frac{-ir + x}{-i - rx}$$
$$x_n = m^n(0),$$

and if $r = 1/\sqrt{2}$, then the trajectory eventually ends at $(1+i)/\sqrt{2}$. Hyperbolic convexity ensures that all these trajectories are (after the first linear step) always in the region contained by the geodesic from $(1+i)/\sqrt{2}$ to $(1-i)/\sqrt{2}$. In particular they never intersect the trajectories whose first move is from zero to i, to $-i$, or to -1. Now, there is nothing special about zero in this discussion, and using conformal invariance it is easy to see that

Lemma 2.2. *If θ is at least equal to the hyperbolic distance from 0 to $1/\sqrt{2}$ in $\mathbb{H}$, then the path Y_t^θ takes its values in a tree. This value $1/\sqrt{2}$ is sharp.*

We have developed X^θ into a tree in $\mathbb{H}$; we have already observed that a loop in a tree is tree-like. If we can prove that $Y_T^\theta = Y_0^\theta$, the Y^θ will be tree-like and hence so will $X^\theta$and X. To achieve this we must use the assumption that the path has finite length and that all its iterated integrals are zero from a different perspective.

2.3. **Cartan development as a linear differential equation.** If G is a finite dimensional Lie group represented as a closed subset of a matrix algebra, and $X_t|_{t\leq T}$ is a path in its Lie algebra $\mathfrak{g}$, then the equation for the Cartan development $M_T \in G$ of $X_t|_{t\leq T} \in \mathfrak{g}$ is given by the differential equation

$$M_{t+\delta t} \approx M_t \exp(\delta X_t) \text{ or equivalently } dM_t = M_t dX_t.$$

The development of a smooth path in the tangent space to 0 in

$$\mathbb{H} = \{z \in \mathbb{C} |\, \|z\| < 1\},$$

is also expressible as a differential equation. However, it is easier to express this development in terms of Cartan development in the group of isometries regarded as matrices in $GL(2,\mathbb{C})$ rather than on the points of $\mathbb{H}$. We identify $\mathbb{R}^2$ with the Lie subspace

$$\begin{pmatrix} 0 & x+iy \\ x-iy & 0 \end{pmatrix}.$$

In this representation, the equation for M_t is linear and so we have an expansion for M_T:

$$\begin{aligned} M_T &= M_0 \left(I + \int_{0<u<T} dX_u + \int_{0<u_1<u_2<T} dX_{u_1} dX_{u_2} + \dots \right) \\ &= M_0 \times \begin{pmatrix} a & b \\ \bar{b} & \bar{a} \end{pmatrix}, \end{aligned}$$

where

$$\begin{aligned} a &= 1 + \sum_k \int_{0<u_1<u_2<\ldots<u_{2k}<T} dX_{u_1} d\bar{X}_{u_2} \dots dX_{u_{2k-1}} d\bar{X}_{u_{2k}} \\ b &= \sum_k \int_{0<u_1<u_2<\ldots<u_{2k-1}<T} dX_{u_1} d\bar{X}_{u_2} \dots dX_{u_{2k-1}} \end{aligned}$$

and $\int_{0<u_1<\cdots<u_{2k}<T} dX_{u_1} d\bar{X}_{u_2} \dots dX_{u_{2k-1}} d\bar{X}_{u_{2k}}$ is now, with an abuse of notation, a complex number. We have an a priori bound:

Lemma 2.3. *If X is a path of length exactly θL, then*

$$\left| \int_{0<u_1<u_2<\ldots<u_{2k}<T} dX_{u_1} d\bar{X}_{u_2} \dots dX_{u_{2k-1}} d\bar{X}_{u_{2k}} \right| \le \frac{(\theta L)^{2k}}{(2k)!}.$$

To use this lemma we need to be able to estimate the tail of an exponential series. The following lemma (based on Stirling's formula) articulates a convenient inequality.

Lemma 2.4. *Let $x \ge 1/e$ and set $\xi_0 = \lim_{y\to\infty} \frac{e^{-y} y^{\frac{1}{2}+y}}{y!} \le \frac{e^{-1/2}}{\sqrt{2\pi}}$.*
(1) $\frac{x^m}{m!} < \frac{\xi_0}{m^{1/2}}$ holds for all $m \ge ex$.
(2) If for any k one has $m \ge ex + k$, then

$$\sum_{r\ge m} \frac{x^r}{r!} \le \frac{e^{\frac{1}{2}}}{\sqrt{2\pi}(e-1)} e^{-k} x^{-1/2} \simeq 0.38\, e^{-k} x^{-1/2}.$$

Proof. By Stirling's formula $\lim_{y\to\infty} \frac{e^{-y} y^{\frac{1}{2}+y}}{y!} = \frac{1}{\sqrt{2\pi}}$ and is approached monotonely from below. It is an upper bound and also a good global approximation to $\frac{e^{-y} y^{\frac{1}{2}+y}}{y!}$ valid for all $y \ge 1$. Putting $y = ex$ gives

$$\frac{e^{-ex} (ex)^{\frac{1}{2}+ex}}{(ex)!} < \frac{1}{\sqrt{2\pi}}$$
$$\frac{x^{ex}}{(ex)!} < e^{-\frac{1}{2}} x^{-\frac{1}{2}} \frac{1}{\sqrt{2\pi}}.$$

Moreover the recurrence relation for the ! function implies, for every $k \in \mathbb{Z}$ with $ex + k > 0$, that

$$\frac{x^{ex}}{(ex)!} \ge e^k \frac{x^{ex+k}}{(ex+k)!}$$

and so

$$\frac{x^{ex+k}}{(ex+k)!} < e^{-k-\frac{1}{2}} x^{-\frac{1}{2}} \frac{1}{\sqrt{2\pi}},$$

establishing the first claim. Now summing this bound we have

$$\begin{aligned}\sum_{k\geq 0} \frac{x^{ex+k}}{(ex+k)!} &\leq e^{-\frac{1}{2}} x^{-\frac{1}{2}} \frac{1}{\sqrt{2\pi}} \sum_{k\geq 0} e^{-k} \\ &= \frac{e^{\frac{1}{2}}}{e-1} \frac{1}{\sqrt{2\pi}} x^{-\frac{1}{2}}.\end{aligned}$$

Since for $ex > 0$ the function $k \to \frac{x^{ex+k}}{(ex+k)!}$ is monotone decreasing on $\mathbb{R}^+$, we see that

$$\sum_{m\geq ex} \frac{x^m}{m!} \leq \frac{e^{\frac{1}{2}}}{e-1} \frac{1}{\sqrt{2\pi}} x^{-\frac{1}{2}},$$

completing the proof of the lemma. Finally we note the approximate value of the constant:

$$\frac{e^{\frac{1}{2}}}{\sqrt{2\pi}(e-1)} \simeq 0.38.$$

□

2.4. **The signature of a word of length** L**.** We deduce the following totally algebraic corollary for paths X that have traversed at most L vertices.

Theorem 5. *If a path of length L in the two dimensional integer lattice (corresponding to a word with L letters drawn from a two letter alphabet and its inverse), has the first* $\lfloor e\log(1+\sqrt{2})L\rfloor$ $GL(2,\mathbb{C})$*-iterated integrals* [3] *zero, then all iterated integrals (in the tensor algebra) are zero, the path is tree-like, and the corresponding reduced word is trivial.*

Proof. Any Mobius transformation preserving the disk can be expressed as

$$M = \begin{pmatrix} z & 0 \\ 0 & \bar{z} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{1-r^2}} & \frac{r}{\sqrt{1-r^2}} \\ \frac{r}{\sqrt{1-r^2}} & \frac{1}{\sqrt{1-r^2}} \end{pmatrix} \begin{pmatrix} \omega & 0 \\ 0 & \bar{\omega} \end{pmatrix},$$

where $|z| = |\omega| = 1$ and r is the Euclidean distance from 0 to $M0$. Now

$$\begin{aligned} Tr\left[A\overline{B}\right] &= \sum_i \sum_j a_{ij}\bar{b}_{ji} \\ &= \overline{Tr\left[B\overline{A}^t\right]} \end{aligned}$$

and

$$\begin{pmatrix} \omega & 0 \\ 0 & \bar{\omega} \end{pmatrix} \overline{\begin{pmatrix} \omega & 0 \\ 0 & \bar{\omega} \end{pmatrix}}^T = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

[3] $GL(2,\mathbb{C})$-iterated integrals: since our path is in a vector subspace of the algebra $GL(2,\mathbb{C})$ we may compute the iterated integrals in the algebra $GL(2,\mathbb{C})$ or in the tensor algebra over the vector subspace. There is a natural algebra homomorphism of the tensor algebra onto $GL(2,\mathbb{C})$. The $GL(2,\mathbb{C})$-iterated integrals are the images of those in the tensor algebra under this projection and a priori contain less information.

hence

$$\begin{aligned} Tr\left[M\overline{M}^{t}\right] &= Tr\left[\left(\begin{array}{cc} \frac{1}{\sqrt{1-r^2}} & \frac{r}{\sqrt{1-r^2}} \\ \frac{r}{\sqrt{1-r^2}} & \frac{1}{\sqrt{1-r^2}} \end{array}\right)^2\right] \\ &= \frac{2\left(1+r^2\right)}{\left(1-r^2\right)}. \end{aligned}$$

Letting $r = 1/\sqrt{2}$ we see that if

$$Tr\left[M\overline{M}^{t}\right] < 6,$$

then the image of 0 under the Mobius transformation must lie in the circle of radius $1/\sqrt{2}$. On the other hand

$$Tr\left[\left(\begin{array}{cc} a & b \\ \bar{b} & \bar{a} \end{array}\right)\left(\begin{array}{cc} \bar{a} & b \\ \bar{b} & a \end{array}\right)\right] = 2\left(|a|^2 + |b|^2\right)$$

and in our context, where the first N iterated integrals are zero, this gives the inequality

$$\begin{aligned} &\left|1 + \sum_{k>N} \int_{0<u_1<\cdots<u_{2k}<T} dX_{u_1} d\bar{X}_{u_2} \ldots dX_{u_{2k-1}} d\bar{X}_{u_{2k}}\right|^2 + \\ &\left|\sum_{k>N} \int_{0<u_1<\cdots<u_{2k-1}<T} dX_{u_1} d\bar{X}_{u_2} \ldots dX_{u_{2k-1}}\right|^2 < 6. \end{aligned} \tag{2.1}$$

Using our a priori estimate from Lemma 2.3, the inequality will hold if

$$\left(1 + \sum_{k>N} \frac{(\theta L)^{2k-1}}{(2k-1)!}\right)^2 + \left(\sum_{k>N} \frac{(\theta L)^{2k}}{(2k)!}\right)^2 < 6.$$

Now observe that, as we will choose $N > \theta L$, the terms in the sums are decreasing. Hence we have

$$\sum_{k>N} \frac{(\theta L)^{2k-1}}{(2k-1)!} < s$$

and also

$$\sum_{k>N} \frac{(\theta L)^{2k}}{(2k)!} < s.$$

We see that (2.1) will always be satisfied if we choose s such that

$$2\left(s + s^2\right) < 5 \text{ and } \sum_{k>N} \frac{(\theta L)^{2k-1}}{(2k-1)!} < s$$

Hence, if

$$\sum_{k>N} \frac{(\theta L)^{2k-1}}{(2k-1)!} < \frac{\sqrt{11}-1}{2},$$

then $Y_T = 0$. By Lemma 2.4 (2) with $x = \log(1+\sqrt{2})L$, we have if $N \geq e\log(1+\sqrt{2})L$,

$$\sum_{m \geq N} \frac{\left(\log(1+\sqrt{2})L\right)^m}{m!} \leq 0.38..(\log(1+\sqrt{2})L)^{-1/2} < \frac{\sqrt{11}-1}{2}$$

for all $L \geq 1$.

Observe that if $\theta \geq \log\left[1+\sqrt{2}\right]$ then $Y_t = M_t 0$ lies in a tree, and the development of Y is such that every vertex of the tree is at least a distance θ from the origin except the origin itself. By our hypotheses and the above argument $d\left[Y_T, 0\right] < \theta$ and hence $Y_T = 0$. Therefore Y is tree-like and the reduced word is trivial. □

Finally we note that the case of the free group with two generators is enough to obtain a general result, as the free group on d generators can be embedded in it.

Lemma 2.5. *Suppose that Γ_d is the free group on d letters e_i and that Γ is the free group on the letters a, b. Then we can identify $f_i \in \Gamma$ so that the homomorphism induced by $e_i \to f_i$ from Γ_d to Γ is an isomorphism and so that the length of the reduced words f_i are at most $|f_i| \leq 2\left\lceil \log_3 \frac{d}{2} \right\rceil + 3$.*

Proof. It is enough to show that we can embed $\Gamma_{23^{l-1}}$ into Γ so that each f_i has length $2l+1$. Consider the collection of all reduced words of length l in Γ. There are 43^{l-1} of them if $l > 0$. Partition them into pairs, so that the left most letter of each of the words in a pair is the same up to inverses. Order them lexicographically. Now consider the space which is the ball in the Cayley graph of Γ comprising reduced words with length at most l. It is obvious that this is a contractible space. Now adjoin new edges connecting the ends of our pairs. Associate with each of the new edges the alternate letter and orient the edge to point from the lower to the higher word in the lexicographic order. Then this new space Δ is contractible to 23^{l-1} loops and so has the free group $\Gamma_{23^{l-1}}$ as its fundamental group. On the other hand, we can obviously lift any path in Δ to the Cayley graph of Γ; the map from loops in Δ to Γ is a homomorphism. The monodromy theorem tells us that this homomorphism induces a homomorphism of the homotopy group of Δ to Γ. As Γ is a tree, any two lifts of paths with the same endpoint in Γ are homotopic relative to those endpoints in graph Γ. Therefore we can associate every point in the homotopy group of Δ with a unique element of Γ and see that the homomorphism is injective. So we see that the image is a copy of the free group $\Gamma_{23^{l-1}}$. The generators of the classes in Δ clearly lift to paths of length $2l+1$ in Γ and we take the end points of these paths to be the f_i. □

Using this and Theorem 5 we have the following general result.

Theorem 6. *If X is a path of length L in the d-dimensional integer lattice and the projections into $GL(2,\mathbb{C})$ of the first $\lfloor (2\left\lceil \log_3 \frac{d}{2} \right\rceil + 3)e\log(1+\sqrt{2})L \rfloor$ iterated integrals are zero, then the path is tree-like.*

In this section, our arguments depended on the tree-like nature of the development of the path in the lattice and little else - this is a property of the development into any rank one symmetric space but is still plausible, if less obvious for general homogeneous spaces. Each space will give rise to a different class of iterated integrals that are sufficient to determine the tree-like nature of a path in a 'jungle gym'. One should note that computing the iterated integrals is not the most efficient way to determine if a word is reducible if the word, as opposed to its signature, is presented.

3. Quantitative versions of Chen's Theorem

We now work with paths in $\mathbb{R}^d$ and aim to obtain a similar result to that in Section 2. We will also discuss the recovery of information about the path from its signature. An important tool is the development of the path into the hyperboloid model for the d-dimensional hyperbolic space $\mathbb{H}$ (which embeds $\mathbb{H}$ into a $d+1$-dimensional Lorentz space) because the isometries of $\mathbb{H}$ extend to linear maps.

Consider the quadratic form on $\mathbb{R}^{d+1}$ defined by

$$I_d\left(x,y\right)=\sum_1^d x_i y_i - x_{d+1}y_{d+1},\ \ \forall x,y\in\mathbb{R}^{d+1}$$

and the surface

$$\mathbb{H}=\left\{x|\ I_d\left(x,x\right)=-1\right\}.$$

Then $\mathbb{H}$ is hyperbolic space with the metric obtained by restricting I_d to the tangent spaces to $\mathbb{H}$. If $x\in\mathbb{H}$, then $\{y|I_d\left(y,x\right)=0\}$ is the tangent space at x to $\mathbb{H}$ in $\mathbb{R}^{d+1}$, and moreover I_d is positive definite on $\{y|I_d\left(y,x\right)=0\}$. Thus this inner product is a Riemannian structure on $\mathbb{H}$. In fact, (see [1], p83) distances in $\mathbb{H}$ can be calculated using I_d

$$-\cosh d\left(x,y\right)=I_d\left(x,y\right) \tag{3.1}$$

If $SO\left(I_d\right)$ denotes the group of matrices with positive determinant preserving[4] the quadratic form I_d then one can prove this is exactly the group of orientation preserving isometries of $\mathbb{H}$. The Lie algebra of $SO\left(I_d\right)$ is easily recognised as the $d+1$ dimensional matrices that are antisymmetric in the top left $d\times d$ block, symmetric in the last column and bottom row and zero in the bottom right corner. Then the development of a path $\gamma\in\mathbb{R}^d$ to $SO\left(I_d\right)$ and $\mathbb{H}$ (chosen to commute with the action of multiplication on the right in $SO\left(I_d\right)$) is given by solving the following differential equation

$$d\Gamma_t=\begin{pmatrix} 0 & \cdots & 0 & d\gamma_t^1 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & d\gamma_t^d \\ d\gamma_t^1 & \cdots & d\gamma_t^d & 0 \end{pmatrix}\Gamma_t. \tag{3.2}$$

We define X to be the development of the path γ to the path in $\mathbb{H}$ starting at $o=\left(0,\cdots,0,1\right)^t$ and given by

$$X_t=\Gamma_t o.$$

---

[4]Precisely, $M\in SO\left(I_n\right)$ if $I_d\left(\left(My^t\right)^t,\left(Mx^t\right)^t\right)\equiv I_d\left(y,x\right)$

Now we can write $d\Gamma_t = F(d\gamma_t)\Gamma_t$ where

$$F : x \to \begin{pmatrix} 0 & \cdots & 0 & x_1 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & x_d \\ x_1 & \cdots & x_d & 0 \end{pmatrix}$$

is a map from $\mathbb{R}^d$ to $\operatorname{Hom}\left(\mathbb{R}^{d+1}, \mathbb{R}^{d+1}\right)$, where for precision we choose the Euclidean norm on $\mathbb{R}^d$ and $\mathbb{R}^{d+1}$ and the operator norm on $\operatorname{Hom}\left(\mathbb{R}^{d+1}, \mathbb{R}^{d+1}\right)$.

Lemma 3.1. *In fact* $||F||_{Hom(\mathbb{R}^{d+1},\mathbb{R}^{d+1})} = 1$.

Proof. Let $e \in \mathbb{R}^d$ and $f \in \mathbb{R}$. Then for $x \in \mathbb{R}^d$

$$F(x) \begin{pmatrix} e \\ f \end{pmatrix} = \begin{pmatrix} fx \\ e.x \end{pmatrix}$$

and computing norms

$$\left|\left| \begin{pmatrix} fx \\ e.x \end{pmatrix} \right|\right|^2 \leq f^2 ||x||^2 + ||e||^2 ||x||^2 = \left|\left| \begin{pmatrix} e \\ f \end{pmatrix} \right|\right|^2 ||x||^2$$

and hence $\|F\|_{Hom(\mathbb{R}^{d+1},\mathbb{R}^{d+1})} = 1$. □

3.1. **Paths close to a geodesic.** We are interested in developing paths γ of fixed length l into paths Γ in $SO(I_d)$, and in the function

$$\varrho(\gamma) := d(o, \Gamma o)$$

giving the length of the chord connecting the beginning and end of the development of γ into Hyperbolic space. Amongst paths γ of fixed length, straight lines maximise ϱ as their developments are geodesics. The function ϱ is a smooth function on path space [6]. Therefore one would expect that for some constant K

$$\varrho(\gamma) \geq l - K\varepsilon^2$$

whenever γ is in the ε-neighbourhood (for the appropriate norm) of a straight line. We will make this precise using Taylor's theorem.

Suppose our straight line is in the direction of a unit vector v. If our path γ is parameterised at unit speed we can represent it by

$$d\gamma_t = \Theta_t v dt,$$

where Θ_t is a path in the linear isometries of $\mathbb{R}^d$. In this discussion we assume that Θ_t is continuous and has modulus of continuity δ. Of course, γ is close to $t \to tv$ if Θ is uniformly close to the identity. Consider the development $\Gamma_t o = (\hat{x}_t, x_t)$ of γ into the hyperboloid model of $\mathbb{H}$ defined by

$$\begin{aligned} x_t &\in \mathbb{R} \\ \hat{x}_t &\in \mathbb{R}^d \\ dx_t &= \hat{x}_t \cdot \Theta_t v dt \qquad x_0 = 1 \\ d\hat{x}_t &= x_t \Theta_t v dt \qquad \hat{x}_0 = 0 \end{aligned}$$

We know that $\|\hat{x}_t\|^2 + 1 = |x_t|^2$ and that

$$\begin{aligned} \cosh d(\Gamma_t o, o) &= -I_d\left(\begin{pmatrix} \hat{x}_t \\ x_t \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right) \\ &= x_t \end{aligned}$$

in other words $\cosh \varrho\left(\gamma|_{[0,t]}\right) = x_t$.

Proposition 3.2. *Suppose that one can express Θ_t in the form e^{A_t} where A_t is a continuously varying anti-symmetric matrix and that $\|A\|_\infty \leq \eta < 1$. Then*

$$|\cosh T - x_T| \leq 4^T \frac{\|A\|_\infty^2}{2}$$

Proof. Suppose $\varepsilon \in [-1,1]$. We introduce a family of paths γ_t^ε with $\gamma_t^1 \equiv \gamma_t$ and with γ_t^0 the straight line tv. We set

$$\begin{aligned} d\gamma_t^\varepsilon &= e^{\varepsilon A_t} v dt \\ \gamma_0^\varepsilon &= 0. \end{aligned}$$

We can then consider the real valued function f on $[-1,1]$ comparing the length of the development of γ^ε and the straight line

$$f(\varepsilon) := \cosh \varrho\left(\gamma_t^\varepsilon|_{t\in[0,T]}\right) - \cosh T.$$

Of course $f(0) = 0$ and $f \leq 0$. Now [6, Theorem 2.2] proves that development of a path γ is Frechet differentiable as a map from paths to paths in all p-variation norms with $p \in [1,2)$.

It is elementary that

$$\begin{aligned} d\left(\gamma_t^\varepsilon - \gamma_t^{\varepsilon+h}\right) &= e^{\varepsilon A_t}\left(1 - e^{hA_t}\right) v dt \\ &= hA_t e^{\varepsilon A_t} v dt + \frac{1}{2}\tilde{h}_t^2 A_t^2 e^{\varepsilon A_t} v dt \end{aligned}$$

where $\tilde{h}_t \in [0,h]$. Working towards the 1-variation derivative

$$\begin{aligned} \int_{t\in[0,T]} \left| d\left(\gamma_t^\varepsilon - \gamma_t^{\varepsilon+h}\right) - hA_t e^{\varepsilon A_t} v dt \right| &\leq \int_{t\in[0,T]} \left| \frac{1}{2}\tilde{h}_t^2 A_t^2 e^{\varepsilon A_t} v dt \right| \\ &\leq \frac{h^2}{2} \int_{t\in[0,T]} \left|A_t^2\right| dt. \end{aligned}$$

Thus $\varepsilon \to \gamma^\varepsilon$ is differentiable with derivative

$$d\gamma^{(1),\varepsilon} := A_t e^{\varepsilon A_t} v dt,$$

providing $\int_{t\in[0,T]} \left|A_t^2\right| dt < \infty$. A similar estimate shows that the derivative of $\gamma^{(1),\varepsilon}$ exists and is

$$d\gamma^{(2),\varepsilon} := A_t^2 e^{\varepsilon A_t} v dt,$$

providing $\int_{t\in[0,T]} \left|A_t^3\right| dt < \infty$. From [6, Theorem 2.2] we know that the development map is certainly twice differentiable in the 1-variation norm and applying the chain rule it follows that f is a twice differentiable function on $[-1,1]$. On the other hand $f(0) = 0$ and $f(\varepsilon) \leq 0$ for $\varepsilon \in [-1,1]$ so that $f'(0) = 0$, and applying Taylor's theorem

$$0 \geq f(1) \geq \inf_{\varepsilon\in[0,1]} \frac{\varepsilon^2}{2} f''(\varepsilon).$$

In fact the derivatives in ε form a simple system of differential equations. If

$$\begin{pmatrix} \hat{x}_t^{\varepsilon+h} \\ x_t^{\varepsilon+h} \end{pmatrix} = \begin{pmatrix} \hat{x}_t^\varepsilon \\ x_t^\varepsilon \end{pmatrix} + h \begin{pmatrix} \hat{y}_t^\varepsilon \\ y_t^\varepsilon \end{pmatrix} + \frac{h^2}{2} \begin{pmatrix} \hat{z}_t^\varepsilon \\ z_t^\varepsilon \end{pmatrix} + o\left(h^2\right),$$

then

$$\begin{pmatrix} d\hat{x}_t^\varepsilon \\ d\hat{y}_t^\varepsilon \\ d\hat{z}_t^\varepsilon \end{pmatrix} = \begin{pmatrix} e^{\varepsilon A_t} v dt & 0 & 0 \\ A_t e^{\varepsilon A_t} v dt & e^{\varepsilon A_t} v dt & 0 \\ A_t^2 e^{\varepsilon A_t} v dt & 2A_t e^{\varepsilon A_t} v dt & e^{\varepsilon A_t} v dt \end{pmatrix} \begin{pmatrix} x_t^\varepsilon \\ y_t^\varepsilon \\ z_t^\varepsilon \end{pmatrix}$$

$$\begin{pmatrix} dx_t^\varepsilon \\ dy_t^\varepsilon \\ dz_t^\varepsilon \end{pmatrix} = \begin{pmatrix} e^{\varepsilon A_t} v dt & 0 & 0 \\ A_t e^{\varepsilon A_t} v dt & e^{\varepsilon A_t} v dt & 0 \\ A_t^2 e^{\varepsilon A_t} v dt & 2A_t e^{\varepsilon A_t} v dt & e^{\varepsilon A_t} v dt \end{pmatrix} \bullet \begin{pmatrix} \hat{x}_t^\varepsilon \\ \hat{y}_t^\varepsilon \\ \hat{z}_t^\varepsilon \end{pmatrix}$$

with the initial conditions

$$\begin{array}{ll} \hat{x}_0^\varepsilon = 0 & x_0^\varepsilon = 1 \\ \hat{y}_0^\varepsilon = 0 & y_0^\varepsilon = 0 \\ \hat{z}_0^\varepsilon = 0 & z_0^\varepsilon = 0. \end{array}$$

The simple exponential bound on the solution of a linear equation shows that

$$|z_t^\varepsilon| \leq 4^{\max\left\{T, \int_{t\in[0,T]} |A_t| dt, \int_{t\in[0,T]} \left|A_t^2\right| dt\right\}}.$$

Applying Taylor's theorem we have that

$$f(\varepsilon) \geq -\frac{\varepsilon^2}{2} 4^{\max\left\{T, \int_{t\in[0,T]} |A_t| dt, \int_{t\in[0,T]} \left|A_t^2\right| dt\right\}}.$$

If $\|A\|_\infty \leq 1$ then $f(\varepsilon) > -\frac{\varepsilon^2}{2} 4^T$, and as $\|A\|_\infty \leq \eta < 1$, then we can replace A by $\eta^{-1}A$ and evaluate $f_{\eta^{-1}A}$ at η to deduce that $f_A(1) > -\frac{\eta^2}{2} 4^T$ giving us the uniform estimate we seek. □

3.2. **Some estimates from hyperbolic geometry.** We require some simple hyperbolic geometry. Fix a point A (in hyperbolic space), and consider two other points B and C. Let θ_A, θ_B, and θ_C be the angles at A, B, and C respectively. Let a, b, and c be the hyperbolic lengths of the opposite sides. Recall the hyperbolic cosine rule

$$\sinh(b)\sinh(c)\cos(\theta_A) = \cosh(b)\cosh(c) - \cosh(a)$$

and note the following simple lemmas:

Lemma 3.3. *If the distance c from A to B is at least* $\log\left(\frac{\cos|\theta_A|+1}{1-\cos|\theta_A|}\right)$, *then*

$$|\theta_B| \leq |\theta_A|.$$

Proof. Fix c and the angle θ_A, the angle θ_B is zero if $b = 0$ and monotone increasing as $b \to \infty$. Suppose that $|\theta_B| > |\theta_A|$. We may reduce b so that $|\theta_B| = |\theta_A|$, now the triangle has two equal edges and applying the cosine rule to compute the base length:

$$\begin{aligned} \sinh(a)\sinh(c)\cos(\theta_A) &= \cosh(a)\cosh(c) - \cosh(a) \\ c &= \log\left(-\frac{(\cos|\theta_A|)e^{2a} + e^{2a} - \cos|\theta_A| + 1}{-e^{2a} + (\cos|\theta_A|)e^{2a} - \cos|\theta_A| - 1}\right) \\ &< \lim_{a\to\infty} \log\left(-\frac{(\cos|\theta_A|)e^{2a} + e^{2a} - \cos|\theta_A| + 1}{-e^{2a} + (\cos|\theta_A|)e^{2a} - \cos|\theta_A| - 1}\right) \\ &= \log\left(\frac{\cos|\theta_A| + 1}{1 - \cos|\theta_A|}\right). \end{aligned}$$

□

Lemma 3.4. *We have* $a \geq b + c - \log \frac{2}{1-\cos\theta_A}$. *Thus if* $\max(b,c) \geq \log \frac{2}{1-\cos\theta_A}$, *then* $a > \min(b,c)$.

Proof. Consider triangles with fixed angle θ_A and with side lengths λb, λc and resulting length $a(\lambda)$ for the side opposite θ_A. Then

$$\lambda b + \lambda c - a(\lambda)$$

is monotone increasing in λ with a finite limit. Now

$$\begin{aligned}
\sinh(\lambda b)\sinh(\lambda c)\cos(\theta_A) &= \cosh(\lambda b)\cosh(\lambda c) - \cosh(a(\lambda)) \\
\frac{\cosh(\lambda b)\cosh(\lambda c)}{\sinh(\lambda b)\sinh(\lambda c)} - \cos(\theta_A) &= \frac{\cosh(a(\lambda))}{\sinh(\lambda b)\sinh(\lambda c)} \\
\lim_{\lambda\to\infty} \log \frac{\cosh(a(\lambda))}{\sinh(\lambda b)\sinh(\lambda c)} &= \lim_{\lambda\to\infty} (a(\lambda) - \lambda b - \lambda b) + \log 2 \\
\lambda b + \lambda c - a(\lambda) &\leq \lim_{\lambda\to\infty} (\lambda b + \lambda c - a(\lambda)) \\
&= \log \frac{2}{1-\cos\theta_A}.
\end{aligned}$$

Thus

$$a \geq b + c - \log \frac{2}{1-\cos\theta_A}.$$

Also, providing $\max(b,c) \geq \log \frac{2}{1-\cos\theta_A}$, one has $a \geq \min(b,c)$. □

Corollary 3.5. *If the distance c from A to B is at least* $\log\left(\frac{2}{1-\cos|\theta_A|}\right)$, *then*

$$|\theta_B| \leq |\theta_A|,$$

and $a \geq b$.

The above lemma is useful in the case where the angles of interest are acute. However we are also interested in the case where one angle is very obtuse, where the following lemma gives much better information.

Lemma 3.6. *Suppose that* $\theta_A > \pi/2$ *and that the distance c from A to B is at least* $\log\left(\sqrt{2}+1\right)$ *then* $\theta_B < (\pi - \theta_A)/2$.

Proof. Fix $\theta_A > \pi/2$. By our assumptions $\cosh(c) \geq \sqrt{2}$, and so applying the second hyperbolic cosine rule

$$\sin(\theta_B)\sin(\theta_A)\cosh(c) = \cos(\theta_C) + \cos(\theta_B)\cos(\theta_A)$$

we have

$$\cosh(c) = \frac{\cos(\theta_C) + \cos(\theta_B)\cos(\theta_A)}{\sin(\theta_B)\sin(\theta_A)} \geq \sqrt{2}.$$

Since the sum of interior angles in a Hyperbolic triangle is less than π one can conclude that $\theta_B = \alpha(\pi - \theta_A)$ where $0 < \alpha < 1$ and that θ_B and θ_C are in $[0, \pi/2)$. To prove this lemma we need to show further that $\alpha \leq \frac{1}{2}$. It is enough to demonstrate that, in the case $\theta_A > \pi/2$, and $\frac{1}{2} < \alpha < 1$, we have

$$\frac{\cos(\theta_C) + \cos(\theta_B)\cos(\theta_A)}{\sin(\theta_B)\sin(\theta_A)} < \sqrt{2}.$$

It is enough to prove that

$$\frac{1 + \cos(\theta_B)\cos(\theta_A)}{\sin(\theta_B)\sin(\theta_A)} < \sqrt{2}.$$

Replacing $(\pi - \theta_A)$ by τ and rewriting

$$f(\alpha,\tau) := \frac{1-\cos(\alpha\tau)\cos(\tau)}{\sin(\alpha\tau)\sin(\tau)}$$

it is enough to prove that $f(\alpha,\tau) < \sqrt{2}$ if $\tau < \pi/2$ and $\frac{1}{2} < \alpha < 1$. As

$$\frac{\partial f}{\partial \alpha}(\alpha,\tau) = \tau \frac{(\cos(\tau) - \cos(\alpha\tau))}{\sin(\tau)\sin(\alpha\tau)^2} \leq 0,$$

f is strictly decreasing in α in our domain. Hence, if $\alpha > 1/2$, then

$$f(\alpha,\tau) < f\left(\frac{1}{2},\tau\right).$$

As

$$\frac{\partial f}{\partial \tau}(1/2,\tau) = \frac{1}{8}\frac{(1+2\cos(\tau/2))\tan(\tau/4)}{\cos(\tau/4)^2\cos(\tau/2)^2} \geq 0,$$

we have

$$f(\alpha,\tau) < f\left(\frac{1}{2},\tau\right) < f\left(\frac{1}{2},\pi/2\right) = \sqrt{2},$$

which completes the argument. □

Lemma 3.7. *Let $0 = T_0 < \ldots < T_i < \ldots T_n = T$ be a partition of $[0,T]$. Let $(X_t)_{t\in[0,T]}$ be a continuous path, geodesic on the intervals $[T_i,T_{i+1}]\,|_{i=0,\ldots,n-1}$ in hyperbolic space with $n \geq 1$ where, at each T_i, the angle between the two geodesic segments: $\angle X_{i-1}X_{T_i}X_{T_{i+1}}$ is in $[2\theta,\pi]$. Suppose that each geodesic segment has length at least $K(\theta) = \log\left(\frac{2}{1-\cos|\theta|}\right)$. Then*

(1) *$d(X_0,X_{T_i})$ is increasing in i and for each $i \leq n$*

$$d(X_0,X_{T_i}) \geq d\left(X_0,X_{T_{i-1}}\right) + d\left(X_{T_{i-1}},X_{T_i}\right) - K(\theta) \geq K(\theta) \tag{3.3}$$

and the angle between $\overrightarrow{X_{T_{i-1}},X_{T_i}}$ and $\overrightarrow{X_0X_{T_i}}$ is at most θ.

(2) *We also have*

$$0 \leq \sum_{i=1}^{n} d\left(X_{T_{i-1}},X_{T_i}\right) - d(X_0,X_{T_n}) \leq (n-1)K(\theta).$$

Proof. We proceed by induction. Suppose $d(X_0,X_{T_i}) \geq K(\theta)$ and the angle $\angle X_{T_{i-1}}X_{T_i}X_{T_0}$ is at most θ. Now the angle $\angle X_{T_{i-1}}X_{T_i}X_{T_{i+1}}$ is at least 2θ so that the angle $\angle X_{T_0}X_{T_i}X_{T_{i+1}}$ is at least θ. As $d(X_0,X_{T_i}) \geq K(\theta)$ and our supposition $d\left(X_{T_i},X_{T_{i+1}}\right) \geq K(\theta)$, Lemma 3.4 and Corollary 3.5 imply

$$d\left(X_0,X_{T_{i+1}}\right) \geq d(X_0,X_{T_i}) + d\left(X_{T_i},X_{T_{i+1}}\right) - K(\theta),$$

and that $\angle X_0X_{T_{i+1}}X_{T_i} \leq \theta$, proving the main inequality. Using the induction one also has the second part of the inequality

$$d(X_0,X_{T_{i+1}}) \geq K(\theta).$$

The second claim is obtained by iterating (3.3),

$$d\left(X_0,X_{T_{i+1}}\right) \geq d(X_0,X_{T_1}) + \sum_{j=1}^{i} d\left(X_{T_j},X_{T_{j+1}}\right) - iK(\theta).$$

Now rearrange to get the result. □

3.3. **The main quantitative estimate.** Let γ in $\mathbb{R}^d$ be a continuous path of finite length l, and parameterised at unit speed. With this parameterisation γ' can be regarded as a path on the unit sphere in $\mathbb{R}^d$. We consider the case where $u \to \gamma'(u)$ is continuous with modulus of continuity δ_γ. If $\alpha \in \mathbb{R}$, then the path $\gamma_\alpha := t \to \alpha\gamma(t/\alpha)$ is also parameterised at unit speed, its length is αl and its derivative has modulus of continuity $\delta_{\gamma_\alpha}(\alpha h) = \delta_\gamma(h)$. Its development (defined in (3.2)) from the identity matrix into $SO(I_d)$ is denoted by Γ_α.

The goal of this section is to provide a quantitative understanding for Γ_a as we let $\alpha \to \infty$. Our estimates will only depend on δ and the length of the path. We let $R_0 = \log(1+\sqrt{2})$.

Proposition 3.8. *Let γ in $\mathbb{R}^d$ be a continuous path of length l. For each $C<1$ and $1 \le M \in \mathbb{N}$, for any α chosen large enough that $\alpha l \ge MR_0$ and $\delta_\gamma\left(\frac{M+1}{M}\frac{R_0}{\alpha}\right) < \sqrt{2\left(\sqrt{2}-\sqrt{1+C^2}\right)4^{-\frac{M+1}{M}R_0}}$, we have*

$$|d(o,\Gamma_\alpha o) - \alpha l| \le \left(\frac{4^{\frac{M+1}{M}R_0}}{2C} + \frac{16\log 2}{\pi^2}\right)\frac{\alpha l}{R_0}\delta_\gamma\left(\frac{M+1}{M}\frac{R_0}{\alpha}\right)^2.$$

In particular for paths of any length $l \ge MR_0$, with modulus of continuity for the derivative $\delta_\gamma\left(\frac{M+1}{M}R_0\right) < \sqrt{2\left(\sqrt{2}-\sqrt{1+C^2}\right)4^{-\frac{M+1}{M}R_0}}$, we have

$$|d(o,\Gamma o) - l| \le \left(\frac{4^{\frac{M+1}{M}R_0}}{2C} + \frac{16\log 2}{\pi^2}\right)\frac{l}{R_0}\delta_\gamma\left(\frac{M+1}{M}R_0\right)^2.$$

We set $D_1(C,M) = \left(\frac{4^{\frac{M+1}{M}R_0}}{2C} + \frac{16\log 2}{\pi^2}\right)\frac{l}{R_0}$, and $D_2(M) = \frac{M+1}{M}R_0$ so that the inequality becomes

$$|d(o,\Gamma_\alpha o) - l\alpha| \le D_1\delta_\gamma(D_2/\alpha)^2\alpha l. \tag{3.4}$$

We note that $R_0 \approx .881374$, $4^{R_0} \approx 3.34393 \le 4^{(M+1)R_0/M} \le 4^{2R_0} \approx 11.5154$ and $\frac{16\log 2}{\pi^2} \approx 1.12369$. Fixing $M=1$, one immediately sees that the distance $d(o,\Gamma_\alpha o)$ grows linearly with the scaling and the chordal distance $d(o,\Gamma_\alpha o)$ behaves like the length of the path γ as $\alpha \to \infty$. We also note that the shape of this result is reminiscent of the elegant result of Fawcett[3, Lemma 68]: that among C^2-curves γ with modulus of continuity $\delta_\gamma(h) \le \kappa h$ one has sharp estimates on the minimal value of $d(o,\Gamma o)$ given by

$$\inf_\gamma \cosh(d(o,\Gamma o)) = \frac{\cosh\left(\alpha l\sqrt{1-\kappa^2}\right) - \kappa^2}{1-\kappa^2}.$$

A natural question to ask is whether our estimate (which is non-infinitesimal and only needs information about $\delta_\gamma(2R_0)$) can be improved to this shape and even to this sharp form.

Proof. The path $\gamma_\alpha := t \to \alpha\gamma(t/\alpha)$ is of length αl and parameterised at unit speed; its derivative has modulus of continuity $\delta_{\gamma_\alpha} : t \to \delta_\gamma(t/\alpha)$. Because $\alpha l \ge MR_0$ we can fix $R = \alpha l/N$, where $R \in \left[R_0, \frac{M+1}{M}R_0\right]$ and N is a positive integer depending on α. Let $t_i = iR$ where $i \in [0,N]$. Let $G_i \in SO(I_d)$ be the development of the path segment $\gamma_\alpha|_{[t_{i-1},t_i]}$ into $SO(I_d)$ and let $\Gamma_{\alpha,t}$ be the development of the path segment $\gamma_\alpha|_{[0,t]}$. We define $X_0 := o \in \mathbb{H}$ and $X_j := G_{t_j}X_{j-1} \in \mathbb{H}$. Then X_j are the points $\Gamma_{\alpha,t_j}o$ on the path $\Gamma_{a,t}o$.

As the length of the path is greater than any chord

$$\begin{aligned} |\alpha l - d(o, \Gamma_\alpha o)| &= \alpha l - \sum_{i=1}^{N} d(X_{i-1}, X_i) \\ &\quad + \sum_{i=1}^{N} d(X_{i-1}, X_i) - d(X_0, X_N), \end{aligned} \tag{3.5}$$

where

$$\alpha l - \sum_{i=1}^{N} d(X_{i-1}, X_i) \geq 0 \text{ and } \sum_{i=1}^{N} d(X_{i-1}, X_i) - d(X_0, X_N) \geq 0.$$

We now estimate each of these terms from above.

For the first term we use our result on paths close to a geodesic. By Proposition 3.2 we have

$$\begin{aligned} \cosh d(X_{i-1}, X_i) &\geq \cosh R - \frac{\delta_{\gamma_\alpha}(R)^2}{2} 4^R \\ &= \cosh R - \frac{\delta_\gamma(\frac{R}{\alpha})^2}{2} 4^R \end{aligned}$$

Thus, using the convexity of cosh and hyperbolic trig identities,

$$\begin{aligned} \frac{\delta_\gamma(\frac{R}{\alpha})^2}{2} 4^R &\geq \cosh R - \cosh d(X_{i-1}, X_i) \\ &\geq (R - d(X_{i-1}, X_i)) \sinh d(X_{i-1}, X_i) \\ &= (R - d(X_{i-1}, X_i)) \sqrt{\cosh d(X_{i-1}, X_i)^2 - 1} \\ &\geq (R - d(X_{i-1}, X_i)) \sqrt{\left(\cosh R - \frac{\delta_\gamma(\frac{R}{\alpha})^2}{2} 4^R\right)^2 - 1} \\ &\geq (R - d(X_{i-1}, X_i)) \sqrt{\left(\sqrt{2} - \frac{\delta_\gamma(\frac{R}{\alpha})^2}{2} 4^R\right)^2 - 1}. \end{aligned}$$

Now, for $C < 1$, providing

$$\left(\sqrt{2} - \frac{\delta_\gamma(\frac{R}{\alpha})^2}{2} 4^R\right)^2 \geq 1 + C^2,$$

we have

$$\frac{\delta_\gamma(\frac{R}{\alpha})^2}{2C} 4^R \geq (R - d(X_{i-1}, X_i)).$$

This will follow if we choose α large enough such that our condition

$$\delta_\gamma(\frac{M+1}{M}\frac{R_0}{\alpha})^2 \leq 2\left(\sqrt{2} - \sqrt{1 + C^2}\right) 4^{-\frac{M+1}{M} R_0}$$

holds.

Hence, summing over all the pieces, we have

$$\left(\frac{\alpha l}{R}\right) \frac{\delta_\gamma(\frac{R}{\alpha})^2}{2C} 4^R \geq \left(\alpha l - \sum_{i=1}^{N} d(X_{i-1}, X_i)\right).$$

We now use our bounds on R to obtain

$$\left(\frac{\alpha l}{R_0}\right)\frac{\delta_\gamma(\frac{M+1}{M}\frac{R_0}{\alpha})^2}{2C}4^{\frac{M+1}{M}R_0} \geq \left(\alpha l - \sum_{i=1}^{N} d\left(X_{i-1}, X_i\right)\right). \tag{3.6}$$

Applying Lemma 3.6 and Lemma 3.7 we see that as $R \geq R_0 = \log(1+\sqrt{2})$, then the angle $X_0 X_{T_n} X_{T_{n+1}}$ is at least $\pi - 2\delta_\gamma\left(\frac{R}{\alpha}\right)$ for each $0 < n < N-1$.

Thus

$$\begin{aligned}\sum_{i=1}^{N} d\left(X_{i-1}, X_i\right) - d\left(X_0, X_{T_N}\right) &\leq (N-1)K(\pi - 2\delta_\gamma(\frac{R}{\alpha})) \\ &= \left(\frac{\alpha l}{R} - 1\right)\log\left(\frac{2}{1-\cos\left(\pi - 2\delta_\gamma\left(\frac{R}{\alpha}\right)\right)}\right).\end{aligned}$$

Since $\left(\frac{1}{u^2}\log\left(\frac{2}{1-\cos|\pi-u|}\right)\right)$ is increasing in u for $u \leq \pi$, we have

$$\log\left(\frac{2}{1-\cos\left|\pi - 2\delta_\gamma\left(\frac{R}{\alpha}\right)\right|}\right) \leq 4\delta_\gamma\left(\frac{R}{\alpha}\right)^2\left(\frac{1}{(\pi/2)^2}\log\left(\frac{2}{1-\cos|\pi/2|}\right)\right).$$

As

$$\delta_\gamma\left(\frac{R}{\alpha}\right) \leq \sqrt{2\left(\sqrt{2}-\sqrt{1+C^2}\right)4^{-\frac{M+1}{M}R_0}} \leq \pi/4$$

for all C and M, we have

$$\log\left(\frac{2}{1-\cos\left|\pi - 2\delta_\gamma\left(\frac{R}{\alpha}\right)\right|}\right) \leq \delta_\gamma\left(\frac{R}{\alpha}\right)^2\frac{16\log 2}{\pi^2}.$$

Observing that $\delta_\gamma(R)$ is increasing and that $\frac{M+1}{M}R_0 \geq R$ gives the second part of our estimate,

$$\begin{aligned}\sum_{i=1}^{n} d\left(X_{i-1}, X_i\right) - d\left(X_0, X_{T_n}\right) &\leq \frac{16\log 2}{\pi^2}\left(\frac{\alpha l}{R} - 1\right)\delta_\gamma\left(\frac{R}{\alpha}\right)^2 && (3.7)\\ &\leq \frac{16\log 2}{\pi^2}\left(\frac{\alpha l}{R_0} - 1\right)\delta_\gamma\left(\frac{M+1}{M}\frac{R_0}{\alpha}\right)^2. && (3.8)\end{aligned}$$

Combining the estimates (3.6) and (3.8) completes the proof. □

3.4. **Recovering information about the path from its signature.** Our main quantitative result in this section proves that a smooth path imposes strong restrictions on even the first few terms of its signature. Before proving it, we explain how relatively crude quantities such as the asymptotic magnitudes of the terms in the signature X of a path γ contain important and detailed information about γ. We show how information about the length l of the path can sometimes be recovered by looking at the exponential behaviour of $\left\|X^k\right\|$ which in practise seems to be independent of the choice of tensor norm.

We also give more refined results relating to the order of convergence; these are quite sensitive to this choice of norm and to the continuity properties of the derivative of γ and interestingly exhibit a phase transition (We are grateful to the referee for showing us the result we give here for the Hilbert Schmidt norm).

The iterated integral of a path in a vector space V is a tensor in $V^{\otimes k}$. Even if there is a canonical norm on V, there are many natural cross norms on $V^{\otimes k}$

associated with the given norm on vectors [11] ; the differences between these norms become accentuated as $k \to \infty$. This is true even in the case where the underlying vector space is $\mathbb{R}^d$. We begin by discussing the Hilbert Schmidt tensor norm.

Lemma 3.9. *Suppose that $\gamma \in V$ is parameterised at unit speed on $[0,l]$, and that $\{u_1 < \ldots < u_k\}$ are the order statistics from k points chosen uniformly from $[0,1]$ then*

$$k!X^k = l^k \mathbb{E}\left(\otimes_{i=1}^k \gamma'(u_i)\right).$$

If, further, V is an inner product space, and $V^{\otimes k}$ is given the Hilbert Schmidt norm then

$$\left\|k!X^k\right\|^2 = l^k \mathbb{E}\left(\prod_{i=1}^k \left\langle \gamma'(u_i), \gamma'(v_i)\right\rangle\right)$$

where $\{v_1 < \ldots < v_k\}$ is a second independently sampled set of order statistics.

Proof. The first claim in this lemma follows the observation that uniform measure on the simplex $\{0 < v_1 < \ldots < v_k < l\}$, when normalised to be a probability measure, is the distribution of the order statistics from k points chosen uniformly from $[0,l]$.

The second claim is only slightly more subtle; if e_ρ is an orthonormal basis for $V^{\otimes k}$, then

$$\begin{aligned}
\left\|\mathbb{E}\left(\otimes\gamma'(u_i)\right)\right\|^2 &= \sum_{e_\rho} \mathbb{E}\left(\left\langle e_\rho, \otimes\gamma'(u_i)\right\rangle\right)^2 \\
&= \sum_{e_\rho} \mathbb{E}\left(\left\langle e_\rho, \otimes\gamma'(u_i)\right\rangle \left\langle e_\rho, \otimes\gamma'(v_i)\right\rangle\right) \\
&= \sum_{e_\rho} \mathbb{E}\left(\left\langle e_\rho, \otimes\gamma'(u_i)\right\rangle \left\langle e_\rho, \otimes\gamma'(v_i)\right\rangle \left\langle e_\rho, e_\rho\right\rangle\right) \\
&= \sum_{e_\rho}\sum_{e_{\rho'}} \mathbb{E}\left(\left\langle e_\rho, \otimes\gamma'(u_i)\right\rangle \left\langle e_{\rho'}, \otimes\gamma'(v_i)\right\rangle \left\langle e_\rho, e_{\rho'}\right\rangle\right) \\
&= \mathbb{E}\left(\sum_{e_\rho}\sum_{e_{\rho'}} \left\langle e_\rho, \otimes\gamma'(u_i)\right\rangle \left\langle e_{\rho'}, \otimes\gamma'(v_i)\right\rangle \left\langle e_\rho, e_{\rho'}\right\rangle\right) \\
&= \mathbb{E}\left(\left\langle \otimes\gamma'(u_i), \otimes\gamma'(v_i)\right\rangle\right) \\
&= \mathbb{E}\left(\prod_{i=1}^k \left\langle \gamma'(u_i), \gamma'(v_i)\right\rangle\right).
\end{aligned}$$

□

In [12] it is shown that the quantile process, the linearly interpolated process of the order statistics $\{0 < u_1 < \cdots < u_k < 1\}$ from a sample of k independent uniform [0,1] random variables, converges to a Brownian bridge. Let $U_k(t) = u_i$ for $t = i/(k+1)$ and interpolate linearly for other values of $t \in [0,1]$.

Lemma 3.10. *The process $\hat{U}_k$ defined by*

$$\hat{U}_k(t) := \sqrt{k}(U_k(t) - t),$$

converges weakly as $k \to \infty$ to the Brownian bridge $W^0 = \{W_t^0 : 0 \le t \le 1\}$ which goes from 0 to 0 in time 1.

Theorem 7. *Suppose that V is an inner product space and that γ is a continuously differentiable path in V of length 1 parameterised by arc-length, then*

$$\lim_{k\to\infty} b_k^{1/k} = \lim_{k\to\infty} \left\| k! X^k \right\|^{1/k} = 1$$

where $V^{\otimes k}$ is given the Hilbert Schmidt norm.

Proof. Let $\{0 < u_1 < \cdots < u_k < 1\}$ and $\{0 < v_1 < \cdots < v_k < 1\}$ be the order statistics from two independent samples of k independent uniform [0,1] random variables. For a path parameterised at unit speed, it is almost surely true that for each i, $\langle \gamma'(u_i), \gamma'(v_i)\rangle \le 1$.

By Lemma 3.10 we have

$$\begin{aligned} p_k : &= \mathbb{P}\left(\max_{1\le i\le k} |u_i - v_i| > \left(\frac{c}{k\log k}\right)^{\frac{1}{2}} \right) \\ &= \mathbb{P}\left(\max_{1\le i\le k} \frac{1}{\sqrt{k}} \left| \hat{U}_k(\frac{i}{k+1}) - \hat{V}_k(\frac{i}{k+1}) \right| > \left(\frac{c}{k\log k}\right)^{\frac{1}{2}} \right) \\ &= \mathbb{P}\left(\sup_t \left| \hat{U}_k(t) - \hat{V}_k(t) \right| > \left(\frac{c}{\log k}\right)^{\frac{1}{2}} \right) \end{aligned}$$

converges to one as $k \to \infty$. If δ is the modulus of continuity of γ' then, since $|\gamma'(u)| \equiv 1$, simple trigonometry leads to the estimate

$$|1 - \langle \gamma'(u_i), \gamma'(v_i)\rangle| \ \le \delta(u_i - v_i)$$

and hence from Lemma 3.9

$$\begin{aligned} E\left(\prod_{i=1}^{k} \langle \gamma'(u_i), \gamma'(v_i)\rangle \right)^{\frac{1}{k}} &\ge E\left(\left(1 - \max_i \delta\left(u_i - v_i\right)\right)^k \right)^{\frac{1}{k}} \\ &\ge \left(p_k \left(1 - \delta\left(\left(\frac{2}{k\log k}\right)^{\frac{1}{2}} \right) \right)^k \right)^{\frac{1}{k}} \\ &\ge p_k^{\frac{1}{k}} \left(1 - \delta\left(\left(\frac{2}{k\log k}\right)^{\frac{1}{2}} \right) \right) \\ &\to 1 \text{ as } k \to \infty. \end{aligned}$$

□

Theorem 8. *If γ is a C^3 path of length l and signature $X = (1, X_1, \ldots,)$, and if $V^{\otimes k}$ is given the Hilbert Schmidt norm, then the limit*

$$\lim_{k\to\infty} \left\| l^{-k} k! X_k \right\|^2 = \mathbb{E}\left(\exp\left(\int_{s\in[0,1]} \left| W_s^0 \right|^2 \langle \gamma'(s), \gamma'''(s)\rangle ds \right) \right) \tag{3.9}$$

exists, where W_s^0 is a Brownian bridge starting at zero and finishing at zero at time 1. The limit is bounded above by 1 and, except for the case where γ is a straight line, the limit is strictly less than one.

Proof. A path parameterised at unit speed has

$$\langle \gamma'(u), \gamma'(u)\rangle = 1, \quad \langle \gamma'(u), \gamma''(u)\rangle = 0$$

so, applying Taylors theorem to third order one has

$$\langle \gamma'(u), \gamma'(v)\rangle = 1 + \langle \gamma'(u), \gamma'''(u)\rangle (v-u)^2/2 + O\left((v-u)^3\right).$$

We make two observations. Firstly that $\langle \gamma'(u), \gamma'(v)\rangle \le 1$ for all v. Secondly the continuous function $\langle \gamma'(u), \gamma'''(u)\rangle \le 0$ and only when γ is a straight line, is it identically equal to zero.

Using Lemma 3.9 we can write

$$\begin{aligned}
& l^{-k}\left\|k!X^k\right\|^2 \\
= \; & \mathbb{E}\left(\prod_i \langle \gamma'(u_i), \gamma'(v_i)\rangle\right). \\
= \; & \mathbb{E}\left(\exp\left(\sum_{i=1}^k \log\langle \gamma'(u_i), \gamma'(v_i)\rangle\right)\right) \\
= \; & \mathbb{E}\left(\exp\left(\frac{1}{2}\sum_{i=1}^k \langle \gamma'(u_i), \gamma'''(u_i)\rangle (\sqrt{k}(v_i-u_i))^2\frac{1}{k} + O\left((v_i-u_i)^3\right)\right)\right) \\
= \; & \mathbb{E}\left(\exp\left(\frac{1}{2}\sum_{i=1}^k \langle \gamma'(u_i), \gamma'''(u_i)\rangle (\hat{V}_k(\frac{i}{k+1}) - \hat{U}_k(\frac{i}{k+1}))^2\frac{1}{k}\right.\right. \\
& \qquad \left.\left. + O\left(\left(\hat{V}_k(\frac{i}{k+1}) - \hat{U}_k(\frac{i}{k+1})\right)^3 \frac{1}{\sqrt{k}}\right)\right)\right).
\end{aligned}$$

As $u_i = i/(k+1) + \hat{U}_k(i/(k+1))/\sqrt{k}$ we can apply Lemma 3.10. The weak convergence of the quantile process to the Brownian bridge ensures that expectations of continuous bounded functions of the path converge and thus, by continuity of γ', γ''', we have that as $k \to \infty$,

$$l^{-k}\left\|k!X^k\right\|^2 \to \mathbb{E}\left(\exp\left(\frac{1}{2}\int_0^1 \langle \gamma'(s), \gamma'''(s)\rangle (W_s^0 - \tilde{W}_s^0)^2 ds\right)\right),$$

where $\tilde{W}^0$ is an independent Brownian bridge going from 0 to 0 in time 1. Now, using the fact that the difference of the two Brownian bridges is a bridge with twice the variance, we have the result. □

We now consider the same questions but with a different norm on $V^{\otimes k}$. From the previous section we know that if α is large enough then there are constants D_1, D_2 such that

$$|d(o, \Gamma_\alpha o) - l\alpha| \le D_1 \delta_\gamma (D_2/\alpha)^2 \alpha l \tag{3.10}$$

and in particular the left hand side will go to zero as $\alpha \to \infty$ if $\delta_\gamma(\varepsilon) = o\left(\varepsilon^{1/2}\right)$.

The lower bound on $d(o, \Gamma_\alpha o)$ implicit in (3.10) leads to a lower bound on the norm of Γ_α as a matrix. We will compare it with the upper bound that comes from expressing the matrix Γ_α as a series whose coefficients are iterated integrals. We have an upper bound for the norm of each coefficient in the series, and their sum provides an upper bound for the matrix norm of Γ_α. We can prove that this upper bound is so close to the lower bound on the norm of Γ_α that it allows us to conclude lower bounds, term by term, for the norms of the coefficients as tensors over $\mathbb{R}^d$ and hence to relate the decay rate for the norms of the iterated integrals directly to the length of γ.

Although identifying $\left(\mathbb{R}^d\right)^{\otimes k}$ with $\mathbb{R}^{dk}$ makes the Euclidean norm a plausible choice of cross norm for $\left(\mathbb{R}^d\right)^{\otimes k}$, in many ways the projective and injective cross norms are more significant. The biggest cross norm is the projective tensor norm.

Definition 3.11. If V is a Banach space and $x \in V^{\otimes k}$, then the projective norm of x is

$$\inf\left\{ \sum_i \|v_{1,i}\| \dots \|v_{k,i}\| \,\middle|\, x = \sum_i v_{1,i} \otimes \dots \otimes v_{k,i} \right\}.$$

Recall that if W is a Banach space, then $\mathrm{Hom}(W,W)$ with the usual operator norm is a Banach algebra. The norm is obtained by considering multilinear maps from V into $\mathrm{Hom}(W,W)$.

Definition 3.12. If V is a Banach space, A is a Banach algebra and $F_1, \dots, F_k \in Hom(V,A)$, then writing $F_1 \otimes \dots \otimes F_k$ for the canonical linear extension of the multilinear map

$$(v_1, v_2 \dots, v_k) \to F_1(v_1) F_2(v_2) \dots F_k(v_k),$$

from $V^{\otimes k} \to A$, we can define

$$\|x\|_{\to A} = \sup_{\substack{F_i \in Hom(V,A) \\ \|F_i\|_{Hom(V,A)} = 1}} \|F_1 \otimes F_2 \dots \otimes F_k(x)\|_A$$

The case $A = \mathbb{R}$ yields the injective (the smallest) cross norm. In practise we will consider the case where A is $\mathrm{Hom}\left(\mathbb{R}^{d+1}, \mathbb{R}^{d+1}\right)$ with the operator norm where $\mathbb{R}^{d+1}$ is given the Euclidean norm; we will give our lower bounds in terms of $\|x\|_{\to Hom(\mathbb{R}^{d+1}, \mathbb{R}^{d+1})}$ and so they hold in the projective tensor norm as well. As Theorem 8 shows, our lower bounds do not in general apply in the injective norm.

Proposition 3.13. *Let $G \in SO\left(I_d\right)$. Then $\|G\|_{Hom(\mathbb{R}^{d+1}, \mathbb{R}^{d+1})} \geq e^{d(o,Go)}$ where $\mathbb{R}^{d+1}$ has the Euclidean norm.*

Proof. If

$$F_\rho := \begin{pmatrix} 1 & 0 & \cdots & \cdots & 0 & 0 & 0 \\ 0 & 1 & \ddots & \ddots & \vdots & \vdots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 1 & 0 & \vdots & \vdots \\ 0 & \cdots & \cdots & 0 & 1 & 0 & 0 \\ 0 & \cdots & \cdots & \cdots & 0 & \cosh\rho & \sinh\rho \\ 0 & \cdots & \cdots & \cdots & 0 & \sinh\rho & \cosh\rho \end{pmatrix},$$

then $F_\rho F_\tau = F_{\rho+\tau}$ and the set of such elements forms a (maximal) Abelian subgroup of $SO\left(I_d\right)$. Any element G of $SO\left(I_d\right)$ can be factored into a Cartan Decomposition $K F_\rho \tilde{K}$ where K and $\tilde{K}$ are built out of rotations Θ of $\mathbb{R}^d$

$$\begin{pmatrix} \Theta & \mathbf{0} \\ \mathbf{0}^t & 1 \end{pmatrix}$$

and $\rho \in \mathbb{R}_+$. As an operator on Euclidean space, G has norm $\|G\| = \left\|K F_\rho \tilde{K}\right\| = \|F_\rho\|$ since K, $\tilde{K}$ are isometries. In addition, the matrix F_ρ is symmetric and hence has a basis comprising eigenfunctions; its norm is at least as large as its largest eigenvalue. Computation shows that the eigenvalues of F_ρ are $\{e^\rho, e^{-\rho}, 1, \cdots, 1\}$ so that, given $\rho > 0$, one has

$$\|G\| \geq e^\rho.$$

On the other hand

$$\begin{aligned} -\cosh d(o, Go) &= I_d\left(\begin{pmatrix} x_1 \\ \vdots \\ x_d \\ \cosh\rho \end{pmatrix}, \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}\right) \\ &= -\cosh\rho \end{aligned}$$

and so $\|G\| \geq e^{d(o,Go)}$. □

If γ is a path of finite length, then the development (3.2) into hyperbolic space $\mathbb{H}$ is defined by

$$d\Gamma_t = F(d\gamma_t)\,\Gamma_t,$$

where by Lemma 3.1 $F : \mathbb{R}^d \to \operatorname{Hom}\left(\mathbb{R}^{d+1}, \mathbb{R}^{d+1}\right)$ has norm one as a map from Euclidean space to the operators on Euclidean space. As a result the development of γ is given by

$$G = I + \int_{0<u<T} F(d\gamma_u) + \ldots + \int_{0<u_1<\ldots<u_k<T} F\left(d\gamma_{u_1}\right) \otimes \ldots \otimes F\left(d\gamma_{u_k}\right) + \ldots$$

and, as in Lemma 2.3, we have an a priori bound which holds for all cross norms and in particular for $\|\cdot\|_{\to Hom(\mathbb{R}^{d+1},\mathbb{R}^{d+1})}$. If l is the length of the path γ, then

$$\left\|\int_{0<u_1<\ldots<u_k<T} d\gamma_{u_1} \otimes \ldots \otimes d\gamma_{u_k}\right\|_{\to Hom(\mathbb{R}^{d+1},\mathbb{R}^{d+1})} \leq \frac{l^k}{k!}.$$

Applying this to $\alpha\gamma$, we conclude that

$$\begin{aligned} e^{d(o,\Gamma_\alpha o)} &\leq \|\Gamma_\alpha\|_{Hom(\mathbb{R}^{d+1},\mathbb{R}^{d+1})} \\ &\leq \sum \alpha^k \left\|\int_{0<u_1<\ldots<u_k<1} F\left(d\gamma_{u_1}\right)\ldots F\left(d\gamma_{u_k}\right)\right\|_{Hom(\mathbb{R}^{d+1},\mathbb{R}^{d+1})} \\ &\leq \sum \alpha^k \|F\|^k_{Hom(\mathbb{R}^d, Hom(\mathbb{R}^{d+1},\mathbb{R}^{d+1}))} \left\|\int_{0<u_1<\ldots<u_k<1} d\gamma_{u_1} \otimes \ldots \otimes d\gamma_{u_k}\right\|_{\to Hom(\mathbb{R}^{d+1},\mathbb{R}^{d+1})} \\ &\leq \sum \alpha^k \left\|\int_{0<u_1<\ldots<u_k<1} d\gamma_{u_1} \otimes \ldots \otimes d\gamma_{u_k}\right\|_{\to Hom(\mathbb{R}^{d+1},\mathbb{R}^{d+1})} \\ &\leq e^{\alpha l}, \end{aligned}$$

where l is the length of γ. Letting

$$b_k = k!\left\|\int_{0<u_1<\ldots<u_k<1} d\gamma_{u_1} \otimes \ldots \otimes d\gamma_{u_k}\right\|_{\to Hom(\mathbb{R}^{d+1},\mathbb{R}^{d+1})}$$

one has for all α that

$$\begin{aligned} e^{d(o,\Gamma_\alpha o)-\alpha l} &\leq e^{-\alpha l}\sum_{k=0}^{\infty} \frac{\alpha^k}{k!} b_k \leq 1 \\ 0 &\leq b_k \leq l^k. \end{aligned}$$

Thus the expectation of b_n with respect to a Poisson measure with mean αl is close to one, while at the same time the b_n are all bounded above by one and positive. In particular

$$\begin{aligned}\sum_{k=0}^{\infty}\frac{\alpha^k}{k!}\left|l^k-b_k\right| &\leq e^{\alpha l}-e^{d(o,\Gamma_\alpha o)}\\ &\leq e^{\alpha l}\left(1-e^{-D_1\delta(D_2/\alpha)^2\alpha l}\right).\end{aligned}$$

and so

$$\left|l^k-b_k\right|\leq\inf_{\alpha>1}k!\alpha^{-k}e^{\alpha l}\left(1-e^{-D_1\delta(D_2/\alpha)^2\alpha l}\right).$$

Now applying Stirling's formula, that $k!=e^{k\log k-k+\frac{1}{2}\log k+C_k}$, where $C_k=o(1)$, and setting $\alpha=k/l$ gives

$$\begin{aligned}\left|l^k-b_k\right| &\leq e^{C_k}l^k\sqrt{k}\left(1-e^{-D_1\delta(D_2l/k)^2k}\right)\\ &\leq l^k\tilde{C}\delta\left(lD_2/k\right)^2k\sqrt{k}\end{aligned}$$

where $\tilde{C}=D_1e^{C_k}$. Thus we see that, if $\delta_\gamma\left(lD_2/k\right)^2k^{3/2}\to 0$ as $k\to\infty$, then $b_k/l^k\to 1$. We have shown the following result.

Theorem 9. *For any path of finite length with $\delta_\gamma(\varepsilon)=o\left(\varepsilon^{3/4}\right)$,*

$$l^{-k}k!\left\|\int_{0<u_1<\ldots<u_k<1}d\gamma_{u_1}\otimes\ldots\otimes d\gamma_{u_k}\right\|_{\to Hom(\mathbb{R}^{d+1},\mathbb{R}^{d+1})}\to 1,$$

as $k\to\infty$,

This is of course quite a strong result obtained by making strong assumptions. One could ask less and so we give a weaker but more widely applicable result.

Theorem 10. *Let γ be a path of finite length l, and suppose its derivative, when parameterised at unit speed, is continuous. Then the Poisson averages C_α of the b_k defined by*

$$C_\alpha=e^{-\alpha}\sum_{k=0}^{\infty}\frac{\alpha^k}{k!}b_k$$

satisfy

$$\lim_{\alpha\to\infty}\frac{1}{\alpha}\log C_\alpha=l-1.$$

Note that the C_α are averages of the b_k against Poisson measures; it is standard that these are close to Gaussian with mean α and variance α.

Proof. Note that

$$e^{d(o,\Gamma_\alpha o)-\alpha l}\leq e^{-\alpha l}\sum_{k=0}^{\infty}\frac{\alpha^k}{k!}b_k\leq 1,$$

and so

$$\frac{d\left(o,\Gamma_\alpha o\right)}{\alpha}-l\leq\frac{1}{\alpha}\log\left(e^{-\alpha}\sum_{k=0}^{\infty}\frac{\alpha^k}{k!}b_k\right)+1-l\leq 0. \tag{3.11}$$

Using (3.10) we have

$$\left|\frac{d\left(o,\Gamma_\alpha o\right)}{\alpha}-l\right|\leq D_1\delta_\gamma\left(D_2/\alpha\right)^2l$$

and hence the left hand side in (3.11) goes to zero. $\square$

In particular we see that the high order coefficients of the signature already determine the length of the path, and in fact we can obtain quantitative estimates in terms of the modulus of continuity for the derivative of γ.

As a consequence of the results in this paper one can show that among the paths of finite length with the same signature as γ there is a unique shortest one, $\tilde{\gamma}$, which we called the tree reduced path associated to γ. If $\tilde{\gamma}$ has a continuous derivative when parameterised at unit speed, then the asymptotic behaviour of the signature of γ gives the length of $\tilde{\gamma}$. In any case $\tilde{\gamma}$ cannot "double back" on itself, so it is at least reasonable to expect that the asymptotics of the signature of γ always give the length of $\tilde{\gamma}$.

Conjecture 3.14. *The length of $\tilde{\gamma}$ can be recovered from the asymptotic behaviour of averages of the b_k.*

It might be that $\lim_{\alpha\to\infty} 1+\frac{1}{\alpha}\log C_\alpha$ gives the length of $\tilde{\gamma}$ directly, although the Poisson averages may have to be replaced in some way.

We conclude with an analogous result to that proved for the lattice case in Theorem 5.

Theorem 11. *Let γ be a path of length l parameterised at unit speed, and let δ_γ be the modulus of continuity for γ'. Fix $C<1$ and $1\leq M\in\mathbb{N}$. Suppose that $\delta_\gamma(0)<\frac{1}{\sqrt{D_1(C,M)}}$, then there is an integer $N(l,\delta)$ such that at least one of the first $N(l,\delta)$ terms in the signature must be non-zero.*

Proof. In the case where the first $e\alpha l$ coefficients in the signature of the path γ are zero, by Lemma 2.4, we have some explicit constant C_1 such that

$$\begin{aligned}\|\Gamma_\alpha\| &\leq 1+\sum_{m>e\alpha l}\frac{(\alpha l)^m}{m!}\\ &\leq 1+C_1(\alpha l)^{-1/2}. \qquad (3.12)\end{aligned}$$

By Proposition 3.13, and letting α be sufficiently large so that we can apply (3.10),

$$\begin{aligned}\|\Gamma_\alpha\| &\geq e^{d(o,\Gamma_\alpha o)}\\ &\geq e^{l\alpha - D_1\delta_\gamma(D_2/\alpha)^2\alpha l}\\ &\geq 1+l\alpha - D_1\delta_\gamma\left(D_2/\alpha\right)^2\alpha l. \qquad (3.13)\end{aligned}$$

The inequalities (3.12) and 3.13) lead to a contradiction if, for large α, we have $l\alpha - D_1\delta_\gamma\left(D_2/\alpha\right)^2\alpha l > C_1(\alpha l)^{-1/2}$ or

$$\alpha^{3/2}\left(1-D_1\delta_\gamma\left(D_2/\alpha\right)^2\right) > C_1 l^{-3/2}.$$

Thus providing $1>D_1\delta_\gamma\left(D_2/\alpha\right)^2$ for some large α (continuity of the derivative is enough), then the left hand side goes to infinity as $\alpha\to\infty$. This always gives a contradiction and shows the existence of $N(l,\delta)$. $\square$

An explicit estimate is $N(l,\delta)=\lceil e\alpha l\rceil$, where one chooses the smallest $\alpha\geq\frac{MR_0}{l}$ large enough so that

$$\delta_\gamma\left(\frac{M+1}{M}\frac{R_0}{\alpha}\right) < \sqrt{2\left(\sqrt{2}-\sqrt{1+C^2}\right)4^{-\frac{M+1}{M}R_0}},$$

and so that $\alpha^{3/2}\left(1-D_1\delta_\gamma\left(D_2/\alpha\right)^2\right)>C_1 l^{-3/2}$. To give an idea of the numerical size of N, the number of iterated integrals required for this result, we note that if the path has $\delta(h)\leq h$, then the optimal value of α is around 15.2 at $C=0.8875$ (with $M=1$) and the number N is the integer greater than $41.38l$ (for large l). With a more careful optimization of the constants (varying M) our estimate can be reduced to the integer greater than $13.28l$ (for large l).

Remark 3.15. We note that this proof did not require that $\delta_\gamma(0)=0$ or that γ' is continuous.

Remark 3.16. An easy way to produce a path with each of the first N iterated integrals zero is to take two paths with the same signature up to the level of the N'th iterated integral and to take the first path concatenated with the second with time run backwards. Since these paths will, except at the point of joining, have the same smoothness as they did before, all focus goes to the point where they join. One could hope that a development of these ideas would prove that the two paths must be nearly tangential. If this were exactly true, then it would give a reconstruction theorem.

We have obtained quantitative lower bounds on the signature of γ when γ is parameterised at unit speed and γ' is close to continuously differentiable. In fact one could obtain estimates whenever the γ' is piecewise continuous and the jumps are less than π. However the main extra idea is already visible in the case where γ' is piecewise constant. We give an explicit estimate in Theorem 13 in Section 6.

4. Tree-Like paths

We now turn to our proof of the extension of Chen's theorem to the case of finite length paths. In this section we suppose that $X_{t\in[0,T]}$ is a path in a Banach or metric space E and we recall our definition 1.2 of tree-like paths in this more general setting.

Theorem 12. *If X is a tree-like path with height function h and, if X is of bounded variation, then there exists a new height function $\tilde{h}$ having bounded variation and hence X is a* Lipschitz *tree-like path; moreover, the variation of $\tilde{h}$ is bounded by the variation of X.*

Proof. The function h allows one to introduce a partial order and tree structure on $[0,T]$. Let $t\in[0,T]$. Define the continuous and monotone function $g_t(.)$ by

$$g_t(v)=\inf_{v\leq u\leq t}h(u),\ v\in[0,t].$$

The intermediate value theorem ensures that g_t maps $[0,t]$ onto $[0,h(t)]$. Let τ_t be a maximal inverse of h in that

$$\tau_t(x)=\sup\{u\in[0,t]\,|g_t(u)=x\},\quad x\in[0,h(t)]. \tag{4.1}$$

As g_t is monotone and continuous

$$\tau_t(x)=\inf\{u\in[0,t]\,|g_t(u)>x\} \tag{4.2}$$

for $x<h(t)$.

Now say $s\preceq t$ if and only if s is in the range of τ_t; that is to say if there is an $x\in[0,h(t)]$ so that $s=\tau_t(x)$. Since $\tau_t(h(t))=\sup\{u\in[0,t]\,|g_t(u)=h(t)\}$, it follows that $\tau_t(h(t)) = t$ and so $t \preceq t$. Since $h(\tau_t(x)) = x$ for $x \in [0, h(t)]$ we see there is an inequality-preserving bijection between the $\{s | s \preceq t\}$ and $[0, h(t)]$.

Suppose $t_1 \preceq t_0$ and that they are distinct; then $h(t_1) < h(t_0)$. We may choose $x_1 \in [0, h(t_0))$ so that $t_1 = \tau_{t_0}(x_1)$, and it follows that

$$\begin{aligned} t_1 &= \tau_{t_0}(x_1) \\ &= \inf\{u \in [0, t_0] \,|g_{t_0}(u) > x_1\}, \end{aligned}$$

and that

$$h(t_1) = x_1 < h(u), \ u \in (t_1, t_0].$$

Of course

$$\begin{aligned} g_{t_0}(t_1) &= \inf_{t_1 \leq u \leq t_0} h(u) \\ &= h(t_1) \\ &= g_{t_1}(t_1), \end{aligned}$$

and hence $g_{t_0}(u) = g_{t_1}(u)$ for all $u \in [0, t_1]$. Hence, $\tau_{t_0}(x) = \tau_{t_1}(x)$ for any $x < g_{t_1}(t_1) = h(t_1) = x_1$; we have already seen that $\tau_{t_1}(h(t_1)) = t_1 = \tau_{t_0}(x_1)$. It follows that the range $\tau_{t_1}([0, h(t_1)])$ is contained in the range of τ_{t_0}. In particular, we deduce that if $t_2 \preceq t_1$ and $t_1 \preceq t_0$ then $t_2 \preceq t_0$.

We have shown that $\preceq$ is a partial order, and that $\{t | t \preceq t_0\}$ is totally ordered under $\preceq$, and in one to one correspondence with $[0, h(t_0)]$.

Now, consider two generic times $s < t$. Let $x_0 = \inf_{s \leq u \leq t} h(u)$ and $I = \{v \in [s, t] \,| h(v) = x_0\}$. Since h is continuous and $[s, t]$ is compact the set I is non-empty and compact. By the construction of the function g_t it is obvious that $g_t \leq g_s$ on $[0, s]$ and that if $g_t(u) = g_s(u)$, then $g_t(v) = g_s(v)$ for $v \in [0, u]$. Thus, there will be a unique $r \in [0, s]$ so that $g_s = g_t$ on $[0, r]$ and $g_t < g_s$ on $(r, s]$. Observe that $g_t(r) = x_0$ and that $\tau_t(x_0) = \sup I$ and, essentially as above $\tau_s = \tau_t$ on $[0, h(r))$. Observe also that if $\tilde{t} \in [s, t]$ then $g_s = g_{\tilde{t}}$ on $[0, r]$ so that $\tau_s = \tau_{\tilde{t}}$ on $[0, h(r))$.

Having understood h and τ to the necessary level of detail, we return to the path X. For x, $y \in [0, h(t)]$ one has, for $x < y$,

$$\begin{aligned} \left\| X_{\tau_t(x)} - X_{\tau_t(y)} \right\| &\leq h(\tau_t(x)) + h(\tau_t(y)) - 2 \inf_{u \in [\tau(x), \tau(y)]} h(u) \\ &\leq x + y - 2 \inf_{z \in [x, y]} h(\tau_t(z)) \\ &= y - x, \end{aligned}$$

so we see that $X_{\tau_t(.)}$ is continuous and of bounded variation.

The intuition is that $X_{\tau_t(.)}$ is the branch of a tree corresponding to the time t. Consider two generic times $s < t$, then $X_{\tau_s(.)}$ and $X_{\tau_t(.)}$ agree on the initial segment $[0, h(r))$ but thereafter $\tau_s(.) \in [r, s]$ while $\tau_t(.) \in [\sup I, t]$. The restriction of $X_{\tau_t(.)}$ to the initial segment $[0, h(r))$ is the path $X_{\tau_{\sup I}(.)}$. As $h(r) = \inf [h(u | u \in [s, t])]$ they have independent trajectories after $h(r)$.

Let $\tilde{h}(t)$ be the total 1-variation of the path $X_{\tau_t(.)}$. The claim is that $\tilde{h}$ has total 1-variation bounded by that of X and is also a height function for X.

As the paths $X_{\tau_s(.)}$ and $X_{\tau_t(.)}$ share the common segment $X_{\tau_r(.)}$ we have

$$\|X_s - X_t\| = \left\| X_{\tau_s(s)} - X_{\tau_t(t)} \right\| \leq \tilde{h}(t) - \tilde{h}(r) + \tilde{h}(s) - \tilde{h}(r),$$

and in particular

$$\|X_s - X_t\| \leq \tilde{h}(s) + \tilde{h}(t) - 2\tilde{h}(r).$$

On the other hand $\tilde{h}(r) = \tilde{h}(\sup I) = \inf_{s\le u\le t}\left(\tilde{h}(u)\right)$ and so

$$\|X(s) - X(t)\| \le \tilde{h}(s) + \tilde{h}(t) - 2\inf_{s\le u\le t}\left(\tilde{h}(u)\right),$$

and $\tilde{h}$ is a height function for X.

Finally we control the total variation of $\tilde{h}$ by ω_X, the total variation of the path. In fact,

$$\begin{aligned} \left|\tilde{h}(s) - \tilde{h}(t)\right| &\le \tilde{h}(s) + \tilde{h}(t) - 2\inf_{s\le u\le t}\left(\tilde{h}(u)\right) \\ &\le \omega_X(s,t), \end{aligned}$$

where $\omega_X(s,t) = \sup_{D\in\mathcal{D}}\sum_D \left\|X_{t_{i+1}} - X_{t_i}\right\|$, with $\mathcal{D}$ denoting the set of all partitions of $[s,t]$ and for $D \in \mathcal{D}$, then $D = \{s \le \cdots < t_i < t_{i+1} < \cdots \le t\}$. The first of these inequalities is trivial, but the second needs explanation. As before, notice that the paths $X_{\tau_s(.)}$ and $X_{\tau_t(.)}$ share the common segment $X_{\tau_r(.)}$, and that $\inf_{s\le u\le t}\left(\tilde{h}(u)\right) = \tilde{h}(r)$. So $\tilde{h}(s) + \tilde{h}(t) - 2\inf_{s\le u\le t}\left(\tilde{h}(u)\right)$ is the total length of the two segments $X_{\tau_s(.)}|_{[h(r),h(s)]}$ and $X_{\tau_t(.)}|_{[h(r),h(t)]}$. Now the total variation of $X_{\tau_t(.)}|_{[h(r),h(t)]}$ is obviously bounded by $\omega_X(\sup I, t)$, as the path $X_{\tau_t(.)}|_{[h(r),h(t)]}$ is a time change of $X|_{[\sup I,t]}$.

It is enough to show that the total length of $X_{\tau_s(.)}|_{[h(r),h(s)]}$ is controlled by $\omega_X(s, \inf I)$ to conclude that

$$\tilde{h}(s) + \tilde{h}(t) - 2\inf_{s\le u\le t}\left(\tilde{h}(u)\right) \le \omega_X(s,t).$$

In order to do this we work backwards in time. Let

$$\begin{aligned} f_s(u) &= \inf_{s\le v\le u} h(v) \\ \rho_t(x) &= \inf\{u \in [s,T] \,|\, f_s(u) = x\} \end{aligned}$$

then, because X is tree-like,

$$X_{\tau_s(.)}|_{[0,h(s)]} = X_{\rho_s(.)}|_{[0,h(s)]},$$

and in particular, the path segment $X_{\rho_s(.)}|_{[h(r),h(s)]}$ is a time change (but backwards) of $X|_{[s,\inf I]}$. □

The property of being tree-like is re-parameterisation invariant. We see informally that a tree-like path X is the composition of a contraction on the R-tree defined by h and the based loop in this tree obtained by taking $t \in [0,T]$ to its equivalence class under the metric induced by h (for definitions and a proof see [4]).

Any path that can be factored through a based loop of finite length in an R -tree and a contraction of that tree to the space E is a Lipschitz tree-like path. If 0 is the root of the tree and ϕ is the based loop defined on $[0,T]$, then define $h(t) = d(0, \phi(t))$. This makes ϕ a tree-like path. Any Lipschitz image of a tree-like path is obviously a Lipschitz tree-like path.

We have the following trivial lemma.

Lemma 4.1. *A Lipschitz tree-like path X always has bounded variation less than that of any height function h for X.*

Proof. Let $\mathcal{D} = \{t_0 < \ldots < t_n\}$ be a partition of $[0,T]$. Choose $u_i \in [t_{i-1}, t_i]$ maximising $h(t_i) + h(t_{i-1}) - 2h(u_i)$ and let $\tilde{\mathcal{D}} = \{t_0 \le u_1 \ldots \le t_{n-1} \le u_n \le t_n\}$. Relabel the points of $\tilde{\mathcal{D}} = \{v_0 \le v_1 \ldots \le v_m\}$. Then

$$\sum_{\mathcal{D}} \left\| X_{t_i} - X_{t_{i-1}} \right\| \le \sum_{\tilde{\mathcal{D}}} \left| h(v_i) - h(v_{i-1}) \right| . \tag{4.3}$$

□

We now prove a compactness result.

Lemma 4.2. *Suppose that $\{h_n\}$ is a sequence of height functions on $[0,T]$ for a sequence of tree-like paths $\{X_n\}$. Suppose further that the h_n are parameterised at speeds of at most one and that the X_n take their values in a common compact set within E. Then we may find a subsequence $\left(X_{n(k)}, h_{n(k)}\right)$ converging uniformly to a Lipschitz tree-like path (Y,h). The speed of traversing h is at most one.*

Proof. The h_n are equi-continuous, and in view of (4.3) the X_n are as well. Our hypotheses are sufficient for us to apply the Arzela-Ascoli theorem to obtain a subsequence $\left(X_{n(k)}, h_{n(k)}\right)$ converging uniformly to some (Y,h). In view of the fact that the Lip norm is lower semi-continuous in the uniform topology, we see that h is a bounded variation function parameterised at speed at most one and that Y is of bounded variation; of course h takes the value 0 at both ends of the interval $[0,T]$.

Now $h_{n(k)}$ converge uniformly to h and hence $\inf_{u\in[s,t]} h_{n(k)}(u) \to \inf_{u\in[s,t]} h(u)$; meanwhile the h_n are height functions for the tree-like paths X_n and hence we can take limits through the definition to show that h is a height function for Y.

□

Corollary 4.3. *Every Lipschitz tree-like path X has a height function h of minimal total variation, and its total variation measure is boundedly absolutely continuous with respect to the total variation measure of any other height function.*

Proof. We see that this is an immediate corollary of Theorem 12 and Lemma 4.1 and 4.2. □

There can be more than one minimiser h for a given X.

5. Approximation of the path

5.1. **Representing the path as a line integral against a rank one 1-form.** Let γ be a path of finite variation in a finite dimensional Euclidean space V with total length T and parameterised at unit speed. Its parameter set is $[0,T]$. We note that the signature of γ is unaffected by this choice of parameterisation.

Definition 5.1. Let $\gamma([0,T])$ denote the range of γ in V and let the occupation measure μ on $(V, \mathcal{B}(V))$ be denoted

$$\mu(A) = \left| \{ s < T | \gamma(s) \in A \} \right| , \quad A \subset V.$$

Let $n(x)$ be the number of points on $[0,T]$ corresponding under γ to $x \in V$. By the area formulae [10] p125-126, one has the total variation, or length, of the path γ is given by

$$Var(\gamma) = \int n(x) \Lambda_1(dx) , \tag{5.1}$$

where Λ_1 is one dimensional Hausdorff measure. Moreover, for any continuous function f

$$\int f(\gamma(t))\,dt = \int f(x)\,n(x)\,\Lambda_1(dx).$$

Note that $\mu = n(x)\Lambda_1$ and that n is integrable.

Lemma 5.2. *The image under γ of a Lebesgue null set is null for μ. That is to say $\mu(\gamma(N)) = \left|\gamma^{-1}\gamma(N)\right| = 0$ if $|N| = 0$.*

Definition 5.3. We will say that $N \subset [0,T]$ is γ-stable if $\gamma^{-1}\gamma(N) = N$.

As a result of Lemma 5.2 we see that any null set can always be enlarged to a γ-stable null set.

The Lebesgue differentiation theorem tells us that γ is differentiable at almost every u in the classical sense, and with this parameterisation the derivative will be absolutely continuous and of modulus one.

Corollary 5.4. *There is a set G of full μ measure in V so that γ is differentiable with $|\gamma'(t)| = 1$ whenever $\gamma(t) \in G$. We set $M = \gamma^{-1}G$; M is γ-stable.*

Now it may well happen that the path visits the same point $x \in G$ more than once. A priori, there is no reason why the directions of the derivative on $\{t \in M | \gamma(t) = m\}$ should not vary. However this can only occur at a countable number of points.

Lemma 5.5. *The set of pairs (s,t) of distinct times in $M \times M$ for which*

$$\begin{array}{rcl}\gamma(s) & = & \gamma(t) \\ \gamma'(s) & \neq & \pm\gamma'(t)\end{array}$$

is countable.

Proof. If $\gamma(s_*) = \gamma(t_*)$ but $\gamma'(s_*) \neq \pm\gamma'(t_*)$ then, by a routine transversality argument, there is an open neighbourhood of (s_*, t_*) in which there are no solutions of $\gamma(s) = \gamma(t)$ except $s = s_*, t = t_*$. □

Up to sign and with countably many exceptions, the derivative of γ does not depend on the occasion of the visit to a point, only the location. Sometimes we will only be concerned with the unsigned or projective direction of γ and identify $v \in S$ with $-v$.

Definition 5.6. For clarity we introduce $\tilde{}_{\pm}$ as the equivalence relation that identifies v and $-v$ and let $[\gamma']_{\tilde{}_{\pm}} \in S/\tilde{}_{\pm}$ denote the unsigned direction of γ.

By Lemma 5.5 $[\gamma']_{\tilde{}_{\pm}}$ is defined on the full measure subset of $[0,T]$ where γ' is defined and in S.

Corollary 5.7. *There is a function ϕ defined on G with values in the projective sphere $S/\tilde{}_{\pm}$ so that $\phi(\gamma(t)) = [\gamma'(t)]_{\tilde{}_{\pm}}$.*

As a result we may define a useful vector valued 1-form μ-almost everywhere on G. If ξ is a vector in S, then $\langle \xi, u\rangle\,\xi$ is the linear projection of u onto the subspace spanned by ξ. As $\langle \xi, u\rangle\,\xi = \langle -\xi, u\rangle(-\xi)$ it defines a function from $S/\tilde{}_{\pm}$ to $\mathrm{Hom}(V,V)$.

Definition 5.8. Given a path γ there is a set G and a ϕ as defined by Corollary 5.7. Let ξ be a unit strength vector field on G with $[\xi]_{\sim_{\pm}} = \phi$. Then we define the tangential projection 1-form ω for a path γ by

$$\omega(g,u) = \langle \xi(g), u \rangle \xi(g), \quad \forall g \in G, \forall u \in V.$$

This is a vector 1-form on G which depends on ϕ, but is otherwise independent of the choice of ξ.

The tangential projection 1-form $\omega(g,u)$ is the projection of $u \in V$ onto the line determined by $\phi(g)$. By construction and the fundamental theorem of calculus for Lipschitz functions, and bounded measurable ω, we have the following.

Proposition 5.9. *The tangential projection 1-form ω, defined μ a.e. on G, is a linear map from $V \to V$ with rank one. For almost every t one has*

$$\gamma'(t) = \omega(\gamma(t), \gamma'(t)),$$

and as a result

$$\begin{aligned} \gamma(t) &= \int_{0<u<t} d\gamma_u + \gamma(0) \\ &= \int_{0<u<t} \omega(d\gamma_u) + \gamma(0), \end{aligned}$$

for every $t \leq T$.

By approximating ω by other rank one 1-forms we will be able to approximate γ by (weakly) piecewise linear paths that also have trivial signature. It will be easy to see that such paths are tree-like. The set of tree-like paths is closed. This will complete the argument.

5.2. **Iterated integrals of iterated integrals.** We now prove that if γ has a trivial signature $(1,0,0,\ldots)$, then it can always be approximated arbitrarily well by weakly piecewise linear paths with shorter length and trivial signature. Our approximations will all be line integrals of 1-forms against our basic path γ. Two key points we will need are that the integrals are continuous against varying the 1-form, and that a line integral of a path with trivial signature also has trivial signature. The Stone-Weierstrass theorem will allow us to reduce this second problem to one concerning line integrals against polynomial 1-forms, and in turn this will reduce to the study of certain iterated integrals. The application of the Stone-Weierstrass theorem requires a commutative algebra structure and this is provided by the coordinate iterated integrals and the shuffle product. For completeness we set this out below.

Suppose that we define

$$Z_u := \underset{0<u_1<\ldots<u_r<u}{\int\cdots\int} d\gamma_{u_1}\ldots d\gamma_{u_r} \in V^{\otimes r}$$

and

$$\tilde{Z}_u := \underset{0<u_1<\ldots<u_{\tilde{r}}<u}{\int\cdots\int} d\gamma_{u_1}\ldots d\gamma_{u_{\tilde{r}}} \in V^{\otimes \tilde{r}},$$

then it is interesting as a general point, and necessary here, to consider iterated integrals of Z and $\tilde{Z}$

$$\int_{0<u_1<u_2<T}\cdots\int d\tilde{Z}_{u_1}dZ_{u_2} \in V^{\otimes\tilde{r}} \otimes V^{\otimes r}.$$

It will be technically important to us to observe that such integrals can also be expressed as linear combinations of iterated integrals of γ so we do this with some care. Some of the results stated below follow from the well known shuffle product and its relationship with multiplication of coordinate iterated integrals.

Definition 5.10. The truncated or n-signature $\mathbf{X}_{s,t}^{(n)}=\left(1, X_{s,t}^1, X_{s,t}^2, \ldots, X_{s,t}^n\right)$ is the projection of the signature $\mathbf{X}_{s,t}$ to the (quotient) algebra $T^{(n)}(V) := \bigoplus_{r=0}^n V^{\otimes r}$ of tensors with degree at most n.

Definition 5.11. If e is an element of the dual space V^* to V, then $\gamma_u^e = \langle e, \gamma_u\rangle$ is a scalar path and $d\gamma_u^e = \langle e, d\gamma_u\rangle$. If $\mathbf{e} = (e_1, \ldots, e_r)$ is a list of elements of the dual space to V, then we define the coordinate iterated integral

$$X_{s,t}^{\mathbf{e}} := \int_{s<u_1<\ldots<u_r<t}\cdots\int d\gamma_{u_1}^{e_1}\ldots d\gamma_{u_r}^{e_r} = \left\langle \mathbf{e}, X_{s,t}^r\right\rangle.$$

Lemma 5.12. *The map $\mathbf{e} \to X_{s,t}^{\mathbf{e}}$ defined above extends uniquely as a linear map from $T^{(n)}(V^*)$ to the space of real valued functions on paths of bounded variation.*

Proof. Let $\mathbf{e} \in T^{(n)}(V^*)$. Since $T^{(n)}(V^*)$ is dual to $T^{(n)}(V)$ the pairing $\mathbf{e} \to \left\langle \mathbf{e}, \mathbf{X}_{s,t}^{(n)}\right\rangle$ defines a real number for each path γ. If $\mathbf{e} = (e_1, \ldots, e_r)$ then this coincides with $X_{s,t}^{\mathbf{e}}$; since such vectors span $T^{(n)}(V^*)$ the result is immediate. □

We therefore extend Definition 5.11.

Definition 5.13. For any n, $\mathbf{e} \in T^{(n)}(V^*)$ we call $X_{s,t}^{\mathbf{e}}$ the $\mathbf{e}$-coordinate iterated integral of γ over the interval $[s, t]$.

These functions on path space are important because they form an algebra under pointwise multiplication and because they are like polynomials and so it is easy to define a differentiation operator on this space. Given two tensors $\mathbf{e}$, $\mathbf{f}$ there is a natural product $\mathbf{e} \sqcup\!\sqcup \mathbf{f}$, called the shuffle product, derived from the above. For basic tensors

$$\begin{aligned} \mathbf{e} &= e_1 \otimes \ldots \otimes e_r \in V^{\otimes r} \\ \mathbf{f} &= f_1 \otimes \ldots \otimes f_s \in V^{\otimes s} \end{aligned}$$

and a shuffle (π_1, π_2) (a pair of increasing injective functions from $(1, .., r), (1, .., s)$ to $(1, .., r+s)$ with disjoint range) one can define a tensor of degree $r+s$:

$$\omega_{(\pi_1,\pi_2)} = \omega_1 \otimes \ldots \otimes \omega_{r+s},$$

where $\omega_{\pi_1(j)} = e_j$ for $j = 1, \ldots r$ and $\omega_{\pi_2(j)} = f_j$ for $j = 1, \ldots s$. Since the ranges of π_1 and π_2 are disjoint a counting argument shows that the union of the ranges is $1, \ldots, r+s$, and that ω_k is well defined for all k in $1, \ldots r+s$ and hence $\omega_{(\pi_1,\pi_2)}$ is defined. By summing over all shuffles

$$\mathbf{e} \sqcup\!\sqcup \mathbf{f} = \sum_{(\pi_1,\pi_2)} \omega_{(\pi_1,\pi_2)}$$

one defines a multilinear map of $V^{\otimes r}\times V^{\otimes s}\to V^{\otimes(r+s)}$.

Definition 5.14. The unique extension of $\sqcup\!\sqcup$ to a map from $T(V)\times T(V)\to T(V)$ is called the shuffle product.

The following is standard.

Lemma 5.15. *The class of coordinate iterated integrals is closed under pointwise multiplication. For each γ the point-wise product of the $\mathbf{e}$-coordinate iterated integral and the $\mathbf{f}$-coordinate iterated integral is the $(\mathbf{e}\sqcup\!\sqcup\mathbf{f})$-coordinate iterated integral:*

$$X^{\mathbf{e}}_{s,t}X^{\mathbf{f}}_{s,t}=X^{\mathbf{e}\sqcup\!\sqcup\mathbf{f}}_{s,t}.$$

Corollary 5.16. *Any polynomial in coordinate iterated integrals is a coordinate iterated integral.*

Remark 5.17. It is at first sight surprising that any polynomial in the linear functionals on $\mathrm{T}(V)$ coincides with a unique linear functional on $T(V)$ when restricted to signatures of paths and reflects the fact that the signature of a path is far from being a generic element of the tensor algebra.

A slightly more demanding remark relates to iterated integrals of coordinate iterated integrals.

Proposition 5.18. *The iterated integral*

$$\int\cdots\int_{s<u_1<\ldots<u_r<t} dX^{\mathbf{e}_1}_{s,u_1}\ldots dX^{\mathbf{e}_r}_{s,u_r} \tag{5.2}$$

is itself a coordinate iterated integral.

Proof. A simple induction ensures that it suffices to consider the case

$$\iint_{s<u_1<u_r<t} dX^{\mathbf{e}}_{s,u_1}dX^{\mathbf{f}}_{s,u_2},$$

where

$$\begin{aligned}\mathbf{e} &= e_1\otimes\ldots\otimes e_r\in(V^*)^{\otimes r}\\ \mathbf{f} &= f_1\otimes\ldots\otimes f_s\in(V^*)^{\otimes s}\end{aligned}$$

and in this case

$$\iint_{s<u_1<u_2<t} dX^{\mathbf{e}}_{s,u_1}dX^{\mathbf{f}}_{s,u_2} = \int\cdots\int_{\substack{s<v_1<\ldots<v_r<t\\ s<w_1<\ldots<w_s<t\\ v_r<w_s}} d\gamma^{e_1}_{v_1}\ldots d\gamma^{e_r}_{v_r}d\gamma^{f_1}_{w_1}\ldots d\gamma^{f_s}_{w_s}.$$

Expressing the integral as a sum of integrals over the regions where the relative orderings of the v_i and w_j are preserved (i.e. all shuffles for which the last card comes from the right hand pack) we have

$$\begin{aligned}\iint_{s<u_1<u_2<t} dX^{\mathbf{e}}_{s,u_1}dX^{\mathbf{f}}_{s,u_2} &= X^{(\mathbf{e}\sqcup\!\sqcup\tilde{\mathbf{f}})\otimes f_s}_{s,t}\\ \tilde{\mathbf{f}} &= f_1\otimes\ldots\otimes f_{s-1}.\end{aligned}$$

$\square$

From this it is, of course, clear that

Lemma 5.19. *If a path has trivial signature, then all non-trivial iterated integrals of its iterated integrals are zero.*

5.3. **Bounded, measurable, and integrable forms.** Recall that γ is a path of finite length in V, and that it is parameterised at unit speed. The occupation measure is μ and has total mass equal to the length T of the path γ. Let $(W, \|\|)$ be a normed space with a countable base (usually V itself). If ω is a μ-integrable 1-form with values in W then we write

$$\|\omega\|_{L^1(V,\mathcal{B}(V))} = \int_V \|\omega(y)\|_{Hom(V,W)}\, \mu(dy) = \int_0^T \|\omega(\gamma_t)\|_{Hom(V,W)}\, dt.$$

Proposition 5.20. *Let $\omega \in L^1(V, \mathcal{B}(V), \mu)$ be a μ-integrable 1-form with values in W. Then the indefinite line integral $y_t := \int_0^t \omega(d\gamma_t)$ is well defined, linear in ω, and a path in W with 1-variation at most $\|\omega\|_{L^1(V,\mathcal{B}(V),\mu)}$.*

Proof. Since ω is a 1-form defined μ-almost surely, $\omega(\gamma_t) \in Hom(V, W)$ (where $Hom(V, W)$ is equipped with the operator norm $\|\cdot\|$) is defined dt almost everywhere. Since ω is integrable, it is measurable, and hence $\omega(\gamma_t)$ is measurable on $[0, T]$. Since γ has finite variation and is parameterised at unit speed, it is differentiable almost everywhere and its derivative is measurable with unit length dt almost surely. Hence $\omega(\gamma_t)(\gamma_t')$ is measurable and dominated by $\|\omega(\gamma_t)\|$, which is an integrable function, and hence $\omega(\gamma_t)(\gamma_t')$ is integrable. Thus the line integral can be defined to be

$$\begin{aligned} y_t &= \int_0^t \omega(\gamma_u)(\gamma_u')\, du \\ \|y_t - y_s\| &\leq \int_s^t \|\omega(\gamma_u)\|_{Hom(V,W)} \|\gamma_u'\|\, du \\ &= \int_s^t \|\omega(\gamma_u)\|_{Hom(V,W)}\, du \end{aligned}$$

and so has 1-variation bounded by $\|\omega\|_{L^1(V,\mathcal{B}(V))}$. □

Proposition 5.21. *Let $\omega_n \in L^1(V, \mathcal{B}(V), \mu)$ be a uniformly bounded sequence of integrable 1-forms with values in a vector space W. Suppose that they converge in $L^1(V, \mathcal{B}(V), \mu)$ to ω, then the signatures of the line integrals $\int \omega_n(d\gamma_t)$ converge to the signature of $\int \omega(d\gamma_t)$.*

Proof. The r'th term in the iterated integral of the line integral $\int \omega_n(d\gamma_t)$ can be expressed as

$$\int\cdots\int_{0<u_1<\ldots<u_r<T} \omega_n(\gamma_{u_1}) \otimes \ldots \otimes \omega_n(\gamma_{u_r})\ (\gamma_{u_1}') \ldots (\gamma_{u_r}')\ du_1 \ldots du_r$$

and since the ω_n converge in $L^1(V, \mathcal{B}(V), \mu)$, it follows from the definition of μ that the $\omega_n(\gamma_u)$ converge to $\omega(\gamma_u)$ in $L^1([0, T], \mathcal{B}(\mathbb{R}), du)$ almost everywhere. Thus $\omega_n(\gamma_{u_1}) \otimes \ldots \otimes \omega_n(\gamma_{u_r})$ converges in $L^1([0, T]^r, \mathcal{B}(\mathbb{R}), du_1 \ldots du_r)$. Since $\|\gamma_u'\| = 1$ for almost every u, Fubini's theorem implies that

$$\omega_n(\gamma_{u_1}) \otimes \ldots \otimes \omega_n(\gamma_{u_r})\ (\gamma_{u_1}') \ldots (\gamma_{u_r}')$$

converges in $L^1\left([0,T]^r, \mathcal{B}(\mathbb{R}), du_1 \dots du_r\right)$ to

$$\omega\left(\gamma_{u_1}\right) \otimes \dots \otimes \omega\left(\gamma_{u_r}\right)\ \left(\gamma'_{u_1}\right) \dots \left(\gamma'_{u_r}\right).$$

Thus, integrating over $0 < u_1 < \dots < u_r < T$, the proposition follows. $\square$

Corollary 5.22. *Let $\omega \in L^1\left(V, \mathcal{B}(V), \mu\right)$. If γ has trivial signature, then so does $\int \omega\left(d\gamma_t\right)$. That is to say, for each r,*

$$\underset{0<u_1,\dots,u_r<T}{\int \cdots \int} \omega\left(d\gamma_{u_1}\right) \dots \omega\left(d\gamma_{u_r}\right) = 0 \in W^{\otimes r}.$$

Proof. It is a consequence of Proposition 5.21 that the set of $L^1\left(V, \mathcal{B}(V), \mu\right)$ forms producing line integrals having trivial signature is closed. By Lusin's theorem, one may approximate, in the $L^1\left(V, \mathcal{B}(V), \mu\right)$ norm, any integrable form by bounded continuous forms. If the initial form is uniformly bounded then the approximations can be chosen to satisfy the same uniform bound.

The support of μ is compact, so by the Stone Weierstrass theorem, we can uniformly approximate these continuous forms by polynomial forms $\omega = \sum_i p_i e_i$, where the p_i are polynomials and e_i are a basis for V^*. Using the fact that

$$\frac{(\gamma_T^{e_1})^r}{r!} = \underset{0<u_1,\dots,u_r<T}{\int \cdots \int} d\gamma_{u_1}^{e_1} \dots d\gamma_{u_r}^{e_1},$$

with Corollary 5.16 and Proposition 5.18, we have that the line integrals against these polynomial forms and their iterated integrals can be expressed as linear combinations of coordinate iterated integrals. If γ has trivial signature over $[0,T]$, then by Lemma 5.19, these integrals will also be zero. It follows from the $L^1\left(V, \mathcal{B}(V), \mu\right)$ continuity of the truncated signature that the signature of the path formed by taking the line integral against a form ω in $L^1\left(V, \mathcal{B}(V), \mu\right)$ will always be trivial. $\square$

5.4. **Approximating rank one 1-forms.**

Definition 5.23. A vector valued 1-form ω is (at each point of V) a linear map between vector spaces. We say the 1-form ω is of rank $k \in \mathbb{N}$ on the support of μ if $\dim\left(\omega(V)\right) \leq k$ at μ-almost every point in V.

A linear multiple of a form has the same rank as the original form, but in general the sum of two forms has any rank less than or equal to the sum of the ranks of the individual components. However, we will now explain how one can approximate any rank one 1-form by piecewise constant rank one 1-forms ω. Additionally we will choose the approximations so that, for some $\varepsilon > 0$, if $\omega(x) \neq \omega(y)$ and $|x-y| \leq \varepsilon$, then either $\omega(x)$ or $\omega(y)$ is zero.

In other words ω is rank one and constant on patches which are separated by thin barrier regions on which it is zero. The patches can be chosen to be compact and such that the μ-measure of the compliment is arbitrarily small.

We will use the following easy consequence of Lusin's theorem for 1-forms defined on a μ-measurable set K. :

Lemma 5.24. *Let ω be a measurable 1-form in $L^1\left(V, \mathcal{B}(V), \mu\right)$. For each $\varepsilon > 0$ there is a compact subset L of $\gamma[0,T]$ so that ω restricted to L is continuous, while $\int_{K \setminus L} \|\omega\|_{Hom(V,W)} \mu(dx) < \varepsilon$.*

Lemma 5.25. *If ω is a measurable 1-form in $L^1(V, \mathcal{B}(V), \mu)$, then for each $\varepsilon > 0$ there are finitely many disjoint compact subsets K_i of K and a 1-form $\tilde{\omega}$, that is zero off $\cup_i K_i$ and constant on each K_i, such that*

$$\int_K \|\omega - \tilde{\omega}\|_{Hom(V,W)} \mu(dx) \leq 4\varepsilon$$

and with the property that $\tilde{\omega}$ is rank one if ω is.

Proof. Let L be the compact subset introduced in Lemma 5.24. Now $\omega(L)$ is compact. Fix $\varepsilon > 0$ and choose $l_1, \ldots, l_n$ so that

$$\omega(L) \subset \cup_{i=1}^n B\left(\omega(l_i), \frac{\varepsilon}{\mu(L)}\right)$$

and put

$$F_j = \omega^{-1}\left(\cup_{i=1}^j B\left(\omega(l_i), \frac{\varepsilon}{\mu(L)}\right)\right).$$

Now choose a compact set $K_j \subset F_j \setminus F_{j-1}$ so that

$$\mu((F_j \setminus F_{j-1}) \setminus K_j) \leq \frac{\varepsilon 2^{-j}}{\|\omega\|_{L^\infty(L,\mathcal{B}(L),\mu)}}.$$

Then the K_j are disjoint and of diameter at most $\frac{2\varepsilon}{\mu(L)}$. Moreover

$$\begin{aligned} L &= F_n \\ \mu(L \setminus \cup_{i=1}^n K_j) &\leq \frac{\varepsilon}{\|\omega\|_{L^\infty(L,\mathcal{B}(L),\mu)}} \end{aligned}$$

and

$$\int_{L\setminus\cup_{i=1}^n K_j} \|\omega\|_{Hom(V,W)} \mu(dx) < \varepsilon.$$

For each non-empty K_j choose $k_j \in K_j$. Define $\tilde{\omega}$ as follows:

$$\begin{aligned} \tilde{\omega}(k) &= \omega(k_j), \quad k \in K_j \\ \tilde{\omega}(k) &= 0, \qquad k \in K \setminus \cup_{i=1}^n K_j. \end{aligned}$$

Then

$$\begin{aligned} \int_{L\setminus\cup_{i=1}^n K_j} \|\omega - \tilde{\omega}\|_{Hom(V,W)} \mu(dx) &< \varepsilon \\ \int_{\cup_{i=1}^n K_j} \|\omega - \tilde{\omega}\|_{Hom(V,W)} \mu(dx) &< \frac{2\varepsilon}{\mu(L)}\mu(L), \end{aligned}$$

and using Lemma 5.24 one has

$$\int_{K\setminus L} \|\omega - \tilde{\omega}\|_{Hom(V,W)} \mu(dx) < \varepsilon.$$

Finally we have

$$\int_K \|\omega - \tilde{\omega}\|_{Hom(V,W)} \mu(dx) \leq 4\varepsilon.$$

If ω had rank one at almost every point of K, then it will have rank one everywhere on L since ω is continuous. As either $\tilde{\omega}(k) = \omega(k_j)$ for some k_j in L or is zero, the form $\tilde{\omega}$ has rank one also. $\square$

The following is an easy consequence.

Proposition 5.26. *Consider the set $\mathcal{P}$ of 1-forms on a set K with the property that there exists finitely many disjoint compact subsets K_i of K so that the 1-form is zero off the K_i and constant on each K_i. The set of rank one 1-forms in $\mathcal{P}$ is a dense subset in the $L^1(V, \mathcal{B}(V), \mu)$ topology of the set of rank one 1-forms in $L^1(V, \mathcal{B}(V), \mu)$.*

6. Piecewise linear paths with no repeated edges

We call a path γ piecewise linear if it is continuous, and if there is a finite partition

$$0 = t_0 < t_1 < t_2 < \ldots < t_r = T$$

such that γ is linear (or more generally, geodesic) on each segment $[t_i, t_{i+1}]$.

Definition 6.1. We say the path is non-degenerate if we can choose the partition so that $\left[\gamma_{t_{i-1}}, \gamma_{t_i}\right]$ and $\left[\gamma_{t_i}, \gamma_{t_{i+1}}\right]$ are not collinear for any $0 < i < r$ and if the $\left[\gamma_{t_{i-1}}, \gamma_{t_i}\right]$ are non-zero for every $0 < i \leq r$.

The positive length condition is automatic if the path is parameterised at unit speed and $0 < T$. If θ_i is the angle $\measuredangle\gamma_{t_{i-1}}\gamma_{t_i}\gamma_{t_{i+1}}$, then γ is non-degenerate if we can find a partition so that for each $0 < i < r$ one has

$$|\theta_i| \neq 0 \bmod \pi.$$

This partition is unique, and we refer to the $\left[\gamma_{t_{i-1}}, \gamma_{t_i}\right]$ as the i-th linear segment in γ. We see, from the quantitative estimate in Lemma 3.7 (1), that if we choose $\theta = \frac{1}{2}\min|\theta_i|$ and scale it so that the length of the minimal segment is at least $K(\theta) = \log\left(\frac{2}{1-\cos|\theta|}\right)$, then its development into hyperbolic space is non-trivial and so its signature is not zero. That is to say, Lemma 3.7 contains all the information needed to give a quantitative form of Chen's uniqueness result in the context of piecewise linear paths:

Theorem 13. *If γ is a non-degenerate piecewise linear path, 2θ is the smallest angle between adjacent edges, and $D > 0$ is the length of the shortest edge, then there is at least one n for which*

$$\left(\frac{2}{1-\cos|\theta|}\right)^{\left(1-\frac{1}{D}\right)} \leq n! \left\| \int_{0<u_1<\ldots<u_n<T} \cdots \int d\gamma_{u_1} \ldots d\gamma_{u_{\tilde{r}}} \right\|$$

and in particular γ has non-trivial signature.

Proof. Choose $\alpha = K(\theta)/D$. Isometrically embed V into $SO(I_d)$ and let Γ_α be the development of $\alpha\gamma$. Then $\Gamma_{\alpha,t}o$ is a piecewise geodesic path in hyperbolic space satisfying the hypotheses in Lemma 3.7. Thus we can deduce that the distance $d(o, \Gamma_\alpha o)$ is at least $K(\theta) > 0$. As in the discussion before Theorem 9 in Section

3.4 we have

$$\begin{aligned} e^{K(\theta)} &\leq \|\Gamma_\alpha\| \\ &\leq \sum_{n=0}^{\infty} \alpha^n \left\| \int \cdots \int_{0<u_1<\ldots<u_n<T} d\gamma_{u_1} \ldots d\gamma_{u_{\tilde{r}}} \right\| \\ &= \sum_{n=0}^{\infty} \frac{1}{n!} \left(\frac{K(\theta)}{D} \right)^n n! \left\| \int \cdots \int_{0<u_1<\ldots<u_n<T} d\gamma_{u_1} \ldots d\gamma_{u_{\tilde{r}}} \right\|. \end{aligned}$$

Now multiplying both side by $e^{-\frac{K(\theta)}{D}}$ we have

$$e^{-\frac{K(\theta)}{D}} e^{K(\theta)} \leq e^{-\frac{K(\theta)}{D}} \sum_{n=0}^{\infty} \frac{1}{n!} \left(\frac{K(\theta)}{D} \right)^n n! \left\| \int \cdots \int_{0<u_1<\ldots<u_n<T} d\gamma_{u_1} \ldots d\gamma_{u_{\tilde{r}}} \right\|.$$

Since any integrable function has at least one point where its value equals or exceeds its average, and since

$$1 = e^{-\frac{K(\theta)}{D}} \sum_{n=0}^{\infty} \frac{1}{n!} \left(\frac{K(\theta)}{D} \right)^n,$$

we can conclude an absolute lower bound on the L^1 norm of the signature, against the Poisson measure. Thus there is an n for which

$$e^{K(\theta)\left(1-\frac{1}{D}\right)} \leq n! \left\| \int \cdots \int_{0<u_1<\ldots<u_n<T} d\gamma_{u_1} \ldots d\gamma_{u_{\tilde{r}}} \right\|.$$

Recalling the form of $K(\theta)$ we have the result. □

Corollary 6.2. *Any piecewise linear path γ that has trivial signature is tree-like with a height function h having the same total variation as γ.*

Proof. We will proceed by induction on the number r of edges in the minimal partition

$$0 = t_0 < t_1 < t_2 < \ldots < t_r = T$$

of γ. We assume that γ is linear on each segment $[t_i, t_{i+1}]$ and that γ is always parameterised at unit speed.

We assume that γ has trivial signature. Our goal is to find a continuous real valued function h with $h \geq 0$, $h(0) = h(T) = 0$, and so that for every s, $t \in [0, T]$ one has

$$\begin{aligned} |h(s) - h(t)| &\leq |t - s| \\ |\gamma_s - \gamma_t| &\leq h(s) + h(t) - 2 \inf_{u \in [s,t]} h(u). \end{aligned}$$

If $r = 0$ the result is obvious; in this case $T = 0$ and the function $h = 0$ does the job.

Now suppose that the minimal partition into linear pieces has $r > 0$ pieces. By Theorem 13, it must be a degenerate partition. In other words one of the $\theta_i = \measuredangle \gamma_{t_{i-1}} \gamma_{t_i} \gamma_{t_{i+1}}$ must have

$$|\theta_i| = 0 \bmod \pi.$$

If $\theta_i = \pi$, the point t_i could be dropped from the partition and the path would still be linear. As we have chosen the partition to be minimal this case cannot occur and we conclude that $\theta_i = 0$ and that the path retraces its trajectory for an interval of length

$$s = \min\left(|t_i - t_{i-1}|, |t_{i+1} - t_i|\right) > 0.$$

Now $\gamma(t_i - u) = \gamma(t_i + u)$ for $u \in [0, s]$ and either $t_i - s = t_{i-1}$ or $t_i + s = t_{i+1}$. Suppose that the former holds. Consider the path segments obtained by restricting the path to the disjoint intervals

$$\begin{aligned} \gamma_- &= \gamma|_{[0,t_{i-1}]} \\ \gamma_+ &= \gamma|_{[t_i+s,T]} \\ \tau &= \gamma|_{[t_i-s,t_i+s]}, \end{aligned}$$

then $\gamma = \gamma_- * \tau * \gamma_+$ where $*$ denotes concatenation.

As the signature map $\gamma \to S(\gamma)$ is a homomorphism, the product of the signatures associated to the segments is the signature of the concatenation of the paths and hence is trivial,

$$\begin{aligned} S\left(\gamma_-\right) \otimes S(\tau) \otimes S\left(\gamma_+\right) &= S(\gamma) \\ &= 1 \oplus 0 \oplus 0 \oplus \ldots \in T(V). \end{aligned}$$

On the other hand the path τ is a linear trajectory followed by its reverse and as reversal produces the inverse signature,

$$S(\tau) = 1 \oplus 0 \oplus 0 \oplus \ldots \in T(V).$$

Thus

$$S\left(\gamma_-\right) \otimes S\left(\gamma_+\right) = 1 \oplus 0 \oplus 0 \oplus \ldots \in T(V)$$

and so the concatenation of γ_- and γ_+ (γ with τ excised) also has a trivial signature. As it is piecewise linear with at least one less edge we may apply the induction hypothesis to conclude that this reduced path is tree-like. Let $\tilde{h}$ be the height function for the reduced path. Then define

$$\begin{aligned} h(u) &= \tilde{h}(u), u \in [0, t_{i-1}] \\ h(u) &= \tilde{h}(u - 2s), u \in [t_i + s, T] \\ h(u) &= s - |t_i - u| + \tilde{h}(t_{i-1}), u \in [t_i - s, t_i + s]. \end{aligned}$$

It is easy to check that h is a height function for γ with the required properties. □

The reader should note that the main result of the paper Theorem 4 linking the signature to tree-like equivalence relies on Chen's result only through the above Corollary and hence only requires a version for piecewise linear paths with no repeated edges. Our quantitative Theorem 13 provides an independent proof of this result but, in this context, is stronger than is necessary; Chen's non-quantitative result could equally well have been used.

We end this section with two straightforward results which will establish half of our main theorem.

Lemma 6.3. *If γ is a Lipschitz tree-like path with height function h, then one can find piecewise linear Lipschitz tree-like paths converging in total variation to a re-parameterisation of γ.*

Proof. Without loss of generality we may re-parameterise time to be the arc length of h. Since h is of bounded variation, so is γ and using the area formula (5.1), for a sequence $\eta_n \downarrow 0$ we can find a nested sequence of finite partitions $u^n = \{u_i^n\}_{i=1}^{N_n}$, $u^n \subset u^{n+1}$, which are increasing in $[0,T]$, with $u_{i+1}^n - u_i^n < \eta_n$, and so that h takes the value $h(u_i^n)$ only finitely many times and only at the times u_i^n. Consider the path γ_n that is linear on the intervals $\left(u_i^n, u_{i+1}^n\right)$ and agrees with γ at the times u_i^n. Define h_n similarly. Then by construction h_n is a height function for γ_n and hence γ_n is a tree. The paths γ_n converge to γ uniformly, and in p-variation for all $p > 1$. However, as we have parameterised h by arc length, it follows that the total variation of γ is absolutely continuous with respect to arc length. As the time partitions u^n are nested it is clear that γ_n is a martingale with respect to the filtration determined by the successive time partitions. Thus, applying the martingale convergence theorem, it follows that γ_n converges to γ in L_1. □

Corollary 6.4. *Any Lipschitz tree-like path has all iterated integrals equal to zero.*

Proof. For piecewise linear tree-like paths it is obvious by induction on the number of segments that all the iterated integrals are 0. Since the process of taking iterated integrals is continuous in p-variation norm for $p < 2$, and Lemma 6.3 proves that any Lipschitz tree-like path can be approximated by piecewise linear tree-like paths in 1-variation, the result follows. □

In the next section we introduce the concept of a weakly piecewise linear path. After reading the definition, the reader should satisfy themselves that the arguments of this section apply equally to weakly piecewise linear paths.

7. Weakly piecewise linear paths

Paths that lie in lines are special.

Definition 7.1. A continuous path γ_t is weakly linear (geodesic) on $[0,T]$ if there is a line l (or geodesic l) so that $\gamma_t \in l$ for all $t \in [0,T]$.

Suppose that γ is smooth enough that one can form its iterated integrals.

Lemma 7.2. *If γ is weakly linear, then the n-signature of the path $\gamma(t)_{t\in[0,T]}$ is*

$$\sum_{n=0}^{\infty} \frac{(\gamma_T - \gamma_0)^{\otimes n}}{n!}.$$

In particular the signature of a weakly linear path is trivial if and only if the path has $\gamma_T = \gamma_0$ or, equivalently, that it is a loop.

Lemma 7.3. *A weakly geodesic, and in particular a weakly linear, path with $\gamma_0 = \gamma_T$ is always tree-like.*

Proof. By definition, γ lies in a single geodesic. Define $h(t) = d(\gamma_0, \gamma_t)$. Clearly

$$\begin{aligned} h(0) &= h(T) = 0 \\ h &\geq 0. \end{aligned}$$

If $h(u) = 0$ at some point $u \in (s,t)$ then

$$\begin{aligned} d(\gamma_s, \gamma_t) &\leq d(\gamma_0, \gamma_s) + d(\gamma_0, \gamma_t) \\ &= h(s) + h(t) - 2 \inf_{u\in[s,t]} h(u) \end{aligned}$$

while if $h(u) > 0$ at all points $u \in (s,t)$ then γ_s and γ_t are both on the same side of γ_0 in the geodesic. Assume that $d(\gamma_0, \gamma_s) \geq d(\gamma_0, \gamma_t)$, then

$$\begin{aligned} d(\gamma_s, \gamma_t) &= d(\gamma_0, \gamma_s) - d(\gamma_0, \gamma_t) \\ &= h(s) - h(t) \\ &\leq h(s) + h(t) - 2 \inf_{u \in [s,t]} h(u). \end{aligned}$$

as required. □

There are two key operations, splicing and excising, which preserve the triviality of the signature and (because we will prove it is the same thing) the tree-like property. However, the fact that excision of tree-like pieces preserves the tree-like property will be a consequence of our work.

Definition 7.4. If $\gamma \in V$ is a path taking $[0,T]$ to the vector space V, $t \in [0,T]$ and τ is a second path in V, then the insertion of τ into γ at the time point t is the concatenation of paths

$$\gamma|_{[0,t]} * \tau * \gamma|_{[t,T]}.$$

Definition 7.5. If $\gamma \in V$ is a path on $[0,T]$,with values in a vector space V, and $[s,t] \subset [0,T]$, then γ with the segment $[s,t]$ excised is

$$\gamma|_{[0,s]} * \gamma|_{[t,T]}.$$

Remark 7.6. Note that these definitions make sense for paths in manifolds as well as in the linear case, but in this case concatenation requires the first path to finish where the second starts. We will use these operations for paths on manifolds, but it will always be a requirement for insertion that τ is a loop based at γ_t, while for excision we require that $\gamma|_{[s,t]}$ is a loop.

We have the following two easy lemmas:

Lemma 7.7. *Suppose that $\gamma \in M$ is a tree-like path in a manifold M, and that τ is a tree-like path in M that starts at γ_t, then the insertion of τ into γ at the point t is also tree-like. Moreover, the insertion at the time point t of any height function for τ into any height function coding γ is a height function for $\gamma|_{[0,t]} * \tau * \gamma|_{[t,T]}$.*

Proof. Assume $\gamma \in M$ is a tree-like path on a domain $[0,T]$, then by definition there is a positive and continuous function h so that for every s, $\tilde{s}$ in the domain $[0,T]$

$$\begin{aligned} d(\gamma_s, \gamma_{\tilde{s}}) &\leq h(s) + h(\tilde{s}) - 2 \inf_{u \in [s,\tilde{s}]} h(u), \\ h(0) &= h(T) = 0. \end{aligned}$$

In a similar way, let the domain of τ be $[0,R]$ and let g be the height function for τ

$$\begin{aligned} d(\tau_s, \tau_{\tilde{s}}) &\leq g(s) + g(\tilde{s}) - 2 \inf_{u \in [s,\tilde{s}]} g(u), \\ g(0) &= g(R) = 0. \end{aligned}$$

Now insert g in h at t and τ in γ at t. Let $\tilde{h}$, $\tilde{\gamma}$ be the resulting functions defined on $[0,T+R]$. Then

$$\begin{aligned} \tilde{\gamma}(s) &= \gamma(s), & 0 \leq s \leq t \\ \tilde{\gamma}(s) &= \tau(s-t), & t \leq s \leq t+R \\ \tilde{\gamma}(s) &= \gamma(s-R), & t+R \leq s \leq T+R, \end{aligned}$$

and

$$\begin{aligned}
\tilde{h}(s) &= h(s), & 0 \leq s \leq t \\
\tilde{h}(s) &= g(s-t), & t \leq s \leq t+R \\
\tilde{h}(s) &= h(s-R), & t+R \leq s \leq T+R,
\end{aligned}$$

where the definition of these functions for $s \in [t+R, T+R]$ uses the fact that τ and g are both loops.

Now it is quite obvious that if $s, \tilde{s} \in [0, T+R] \setminus [t, t+R]$, then

$$\begin{aligned}
d(\tilde{\gamma}_s, \tilde{\gamma}_{\tilde{s}}) &\leq \tilde{h}(s) + \tilde{h}(\tilde{s}) - 2 \inf_{u \in [s,\tilde{s}] \setminus [t,t+R]} \tilde{h}(u) \\
&\leq \tilde{h}(s) + \tilde{h}(\tilde{s}) - 2 \inf_{u \in [s,\tilde{s}]} \tilde{h}(u) \\
\tilde{h}(0) &= \tilde{h}(T+R) = 0,
\end{aligned}$$

and that for $s, \tilde{s} \in [t, t+R]$,

$$\begin{aligned}
d(\tilde{\gamma}_s, \tilde{\gamma}_{\tilde{s}}) &= d(\tau_{s-t}, \tau_{\tilde{s}-t}) \\
&\leq g(s-t) + g(\tilde{s}-t) - 2 \inf_{u \in [s-t,\tilde{s}-t]} g(u) \\
&= \tilde{h}(s) + \tilde{h}(\tilde{s}) - 2 \inf_{u \in [s,\tilde{s}]} \tilde{h}(u).
\end{aligned}$$

To finish the proof we must consider the case where $0 \leq s \leq t \leq \tilde{s} \leq t+R$ and the case where $0 \leq t \leq s \leq t+R \leq \tilde{s} \leq T+R$. As both cases are essentially identical we only deal with the first. In this case

$$\begin{aligned}
d(\tilde{\gamma}_s, \tilde{\gamma}_{\tilde{s}}) &= d(\gamma_s, \tau_{\tilde{s}-t}) \\
&\leq d(\gamma_s, \gamma_t) + d(\tau_0, \tau_{\tilde{s}-t}) \\
&\leq h(s) + h(t) - 2 \inf_{u \in [s,t]} h(u) + g(\tilde{s}-t) - g(0) \\
&= \tilde{h}(s) + \tilde{h}(\tilde{s}) - 2 \inf_{u \in [s,t]} \tilde{h}(u) \\
&\leq \tilde{h}(s) + \tilde{h}(\tilde{s}) - 2 \inf_{u \in [s,\tilde{s}]} \tilde{h}(u).
\end{aligned}$$

□

Remark 7.8. The argument above is straightforward and could have been left to the reader. However, we draw attention to the converse result, which also seems very reasonable: that a tree-like path with a tree-like piece excised is still tree-like. This result seems very much more difficult to prove as the height function one has initially, as a consequence of γ being tree-like, may well not certify that τ is tree-like even though there is a second height function defined on $[s, t]$ that certifies that it is. A direct proof that there is a new height function simultaneously attesting to the tree-like nature of γ and τ seems difficult. Using the full power of the results in the paper, we can do this for paths of bounded variation.

Lemma 7.9. *Let be γ a path defined on $[0, T]$ with values in V and suppose that $\gamma|_{[s,t]}$ has trivial signature where $[s, t] \subset [0, T]$. Then γ has trivial signature if and only if γ with the segment $[s, t]$ excised has trivial signature.*

Proof. This is also easy. Since the signature map is a homomorphism we see that

$$\begin{aligned}
\gamma &= \gamma|_{[0,s]} * \gamma|_{[s,t]} * \gamma|_{[t,T]} \\
S(\gamma) &= S\left(\gamma|_{[0,s]}\right) \otimes S\left(\gamma|_{[s,t]}\right) \otimes S\left(\gamma|_{[t,T]}\right)
\end{aligned}$$

and by hypothesis $S\left(\gamma|_{[s,t]}\right)$ is the identity in the tensor algebra. Therefore

$$\begin{aligned} S(\gamma) &= S\left(\gamma|_{[0,s]}\right) \otimes S\left(\gamma|_{[t,T]}\right) \\ &= S\left(\gamma|_{[0,s]} * \gamma|_{[t,T]}\right). \end{aligned}$$

□

Definition 7.10. A continuous path γ, defined on $[0, T]$ is weakly piecewise linear (or more generally, weakly geodesic) if there are finitely many times

$$0 = t_0 < t_1 < t_2 < \ldots < t_r = T$$

such that for each $0 < i \leq r$, the path segment $\gamma_{[t_{i-1},t_i]}$ is weakly linear (geodesic).[5]

Our goal in this section is to prove, through an induction, that a weakly linear path with trivial signature is tree-like and to construct its height function. As before, every such path admits a unique partition.

Lemma 7.11. *If γ is a weakly piecewise linear path, then there exists a unique partition $0 = t_0 < t_1 < t_2 < \ldots < t_r = T$ so that the lines defined by the path segments $\{\gamma_u : u \in [t_{i-1}, t_i]\}$ and $\{\gamma_u : u \in [t_i, t_{i+1}]\}$ are not collinear for any $0 < i < r$.*

We will henceforth only use this partition and refer to r as the number of segments in γ.

Lemma 7.12. *If γ is a weakly linear path with trivial signature and at least one segment, then there exists $0 < i \leq r$ so that $\gamma_{t_{i-1}} = \gamma_{t_i}$.*

Proof. The arguments in the previous section on piecewise linear paths apply equally to weakly piecewise linear and weakly piecewise geodesic paths. In particular Lemma 3.7 only refers to the location of γ at the times t_i at which the path changes direction (by an angle different from π). □

Proposition 7.13. *Any weakly piecewise linear path γ with trivial signature is tree-like with a height function whose total variation is the same as that of γ.*

Proof. The argument is a simple induction using the lemmas above. If it has no segments we are clearly finished with $h \equiv 0$. We now assume that any weakly piecewise linear path $\gamma^{(r-1)}$, consisting of at most $r-1$ segments, with trivial signature is tree-like with a height function whose total variation is the same as that of $\gamma^{(r-1)}$. Suppose that $\gamma^{(r)}$ is chosen so that it is a weakly piecewise linear path of r segments with trivial signature but there was no height function coding it as a tree-like path with total variation controlled by that of $\gamma^{(r)}$. Then, by Lemma 7.12, in the standard partition there must be $0 < i \leq r$ so that $\gamma^{(r)}_{t_{i-1}} = \gamma^{(r)}_{t_i}$, and by assumption $t_{i-1} < t_i$. In other words, the segment $\gamma^{(r)}|_{[t_{i-1},t_i]}$ is a weakly linear segment and a loop. It therefore has trivial signature, is tree-like and the height function we constructed for it in the proof of Lemma 7.3 was indeed controlled by the variation of the loop.

Let $\hat{\gamma}$ be the result of excising the segment $\gamma|_{[t_{i-1},t_i]}$ from $\gamma^{(r)}$. As $\gamma^{(r)}|_{[t_{i-1},t_i]}$ has trivial signature, by Lemma 7.9, $\hat{\gamma}$ also has trivial signature. On the other hand, $\hat{\gamma}$ is weakly piecewise linear with fewer edges than γ (it is possible that γ restricted

[5] The geodesic will always be unique since the path has unit speed and $t_i < t_{i+1}$ so contains at least two distinct points.

to $[t_{i-2}, t_{i-1}]$ and $[t_i, t_{i+1}]$ is collinear and so the number of edges drops by more than one in the canonical partition - but it will always drop!). So by induction, $\hat{\gamma}$ is tree-like and is controlled by some height function $\hat{h}$ that has total variation controlled by the variation of $\hat{\gamma}$.

Now insert the tree-like path $\gamma^{(r)}|_{[t_{i-1},t_i]}$ into $\hat{\gamma}$. By Lemma 7.7 this will be tree-like and the height function is simply the insertion of the height function for $\gamma^{(r)}|_{[t_{i-1},t_i]}$ into that for $\hat{\gamma}$ and by construction is indeed controlled by the variation of $\gamma^{(r)}$ as required. Thus we have completed our induction. □

8. Proof of the main theorem

We can now combine the results of the last sections to conclude the proof of our main theorem and its corollaries.

Proof of Theorem 4. Corollary 6.4 establishes that tree-like paths have trivial signature.

Thus we only need to establish that if the path of bounded variation has trivial signature, then it is tree-like. By Lemma 5.9 we can write the path as an integral against a rank one 1-form. By Corollary 5.26 we can approximate any rank one 1-form by a sequence of rank one 1-forms with the property that each 1-form is piecewise constant on finitely many disjoint compact sets and 0 elsewhere. By integrating γ against the sequence of 1-forms we can construct a sequence of weakly piecewise linear paths approximating γ in bounded variation. By Corollary 5.22, these approximations have trivial signature. By Proposition 7.13 this means that these weakly piecewise linear paths must be tree-like. Hence we have a sequence of tree-like paths which approximate γ. By re-parameterising the paths at unit speed and using Lemma 4.2 γ must be tree-like, completing the proof. □

Proof of Corollary 1.5. Recall that we defined $X \sim Y$, by the relation that X then Y run backwards is tree-like. The transitivity is the part that is not obvious. However, we can now say $X \sim Y$ if and only if the signature of $\mathbf{X}\mathbf{Y}^{-1}$ is trivial. As multiplication in the tensor algebra is associative, it is now simple to check the conditions for an equivalence relation. Denoting the signature of X by $\mathbf{X}$ etc. one sees that

1. The path run backward has signature $\mathbf{Y}\mathbf{X}^{-1} = -\mathbf{X}\mathbf{Y}^{-1} = \mathbf{0}$.
2. $\mathbf{X}\mathbf{X}^{-1} = \mathbf{0}$ by definition.
3. If $X \sim Y$ and $Y \sim Z$, then $\mathbf{X}\mathbf{Y}^{-1} = \mathbf{0}$ and $\mathbf{Y}\mathbf{Z}^{-1} = \mathbf{0}$. Thus

$$\mathbf{0} = \left(\mathbf{X}\mathbf{Y}^{-1}\right)\left(\mathbf{Y}\mathbf{Z}^{-1}\right) = \mathbf{X}\left(\mathbf{Y}^{-1}\mathbf{Y}\right)\mathbf{Z}^{-1} = \mathbf{X}\left(\mathbf{0}\right)\mathbf{Z}^{-1} = \mathbf{X}\mathbf{Z}^{-1}$$

and hence $X \sim Z$ as required.

It is straightforward to see that the equivalence classes form a group. □

Proof of Corollary 1.6. In order to deduce the existence and uniqueness of minimisers for the length within each equivalence class we observe that;

1. We can re-parameterise the paths to have unit speed and thereafter to be constant. The equivalence classes of paths with the same signature and uniformly bounded length is compact. By the extension theorem of [9] and the continuity of the iterated integrals in the path we see that the signatures of a convergent sequence of paths will converge in p-variation for all $p > 1$. Thus any sequence of paths will have a subsequential uniform limit with the same signature. As length is lower-semicontinuous in the uniform topology, the limit of a sequence of paths

with length decreasing to the minimum will have length less than or equal to the minimum. We have seen, through a subsubsequence where the height functions also converge, that it will also be in the same equivalence class as far as the signature is concerned, so it is a minimiser.

2. Within the class of paths with given signature and finite length there will always be *at least one* minimal element. Let X and Y be two minimisers parameterised at unit speed, and let h be a height function for XY^{-1}. Let the time interval on which h is defined be $[0, T]$ and let τ denote the time at which the switch from X to Y occurs. The function h is monotone on $[0, \tau]$ and on $[\tau, T]$ for otherwise there would be an interval $[s, t] \subset [0, \tau]$ with $h(s) = h(t)$. Then the function $u \to h(u) - h(s)$ is a height function confirming that the restriction of X to $[s, t]$ is tree-like. Now we know from the associativity of the product in the tensor algebra that the signature is not changed by excision of a tree-like piece. Therefore, X with the interval $[s, t]$ excised is in the same equivalence class as X but has strictly shorter length. Thus X could not have been a minimiser - as it is, we deduce the function h is strictly monotone. A similar argument works on $[\tau, T]$.

Let $\sigma : [0, \tau] \to [\tau, T]$ be the unique function with $h(t) = h(\sigma(t))$. Then σ is continuous decreasing and $\sigma(0) = T$ and $\sigma(\tau) = \tau$. Moreover, $X_u = Y_{T-\sigma(u)}$ and so we see that (up to reparameterisations), the two paths are the same.

Hence we have a unique minimal element! □

References

[1] Cannon, J. W.; Floyd, W. J.; Kenyon, R.; Parry, W. R. *Hyperbolic geometry.* In Flavors of geometry, 59–115, Math. Sci. Res. Inst. Publ., 31, Cambridge Univ. Press, Cambridge, 1997.
[2] Chen, K. T. *Integration of paths—a faithful representation of paths by non-commutative formal power series.* Trans. Amer. Math. Soc. 89 (1958), 395–407
[3] Fawcett, T. A. *DPhil Thesis*, Mathematical Institute, University of Oxford 2002.
[4] Hambly, B.M. and Lyons, T.J. *Some notes on trees and paths* available at http://arxiv.org/abs/0809.1365
[5] Le Gall, J. F. *Brownian excursions, trees and measure-valued branching processes* Ann. Probab. 19 (1991), 1399–1439.
[6] Li, X. D., Lyons, T. J., *Smoothness of Itô maps and diffusion processes on path spaces (I)* Ann. Sci. Ecole Norm. Sup. 39 (2006), 649–677.
[7] Lyons, T. J., Qian, Z. *Rough paths and system control*, Oxford University Press, 2003.
[8] Lyons, T. J., Sidorova, Nadia *On the radius of convergence of the logarithmic signature*, Illinois J. Math. 50 (2006), 763–790.
[9] Lyons, T. J., Caruana, M., Lévy, T., *Differential equations driven by rough paths* Ecole d'Eté de Probabilités de Saint-Flour XXXIV-2004, 1–116, Lecture Notes in Mathematics 1908, Springer Berlin, (2006).
[10] Ohtsuka, M. *Dirichlet problems on Riemann surfaces and conformal mappings.* Nagoya Math J., 3 (1951), 91–137.
[11] Pisier, G. *Factorization of linear operators and geometry of Banach spaces.* CBMS Regional Conference Series in Mathematics, 60. American Mathematical Society, Providence, RI, 1986.
[12] Shorack, G. R. *Convergence of Reduced Empirical and Quantile Processes with Application to Functions of Order Statistics in the Non-I.I.D. Case.* Ann. Statist. 1, (1973), 146–152.

(B.M. Hambly and Terry J. Lyons) Mathematical Institute, Oxford University, 24-29 St. Giles, Oxford OX1 3LB, England

E-mail address, B.M. Hambly: `hambly@maths.ox.ac.uk`

E-mail address, Terry J. Lyons: `tlyons@maths.ox.ac.uk`