Uniqueness for the Signature of a Path of Bounded Variation and Continuous Analogues for the Free Group

Ben Hambly & Terry J. Lyons*

December 2003

Abstract

This paper is at an interface between analysis, topology and control theory. In it we prove a non-commutative analogue of the result that functions on the circle are determined, up to Lebesgue null sets, by their Fourier coefficients. The equivalence relation determined by the "null sets" for this new problem enable us to construct a continuous analogue of the free group as a quotient subspace of the space of paths of bounded variation. In this development our main theorem extends the work of K.T. Chen, in particular the paper [1], where he considers a similar theorem under a greater smoothness assumption.

1 Introduction

Definition 1.1 *Let $X_t|_{t\in[0,T]}$ be a piecewise smooth, or more generally bounded variation, path in $\mathbb{R}^d$. Then its signature is the sequence of definite iterated integrals*

$$\begin{aligned}\mathbf{X} &= \left(1 + X^1 + \ldots + X^k + \ldots\right)\\ &= \left(1 + \int_{0<u<T} dX + \ldots + \int_{0<u_1<\ldots<u_k<T} dX_{u_1} \otimes \ldots \otimes dX_{u_k} + \ldots\right)\\ &\in T\left((V)\right)\end{aligned}$$

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

This signature provides a fundamental description of the path X and is a strongly non-commutative analogue of the Fourier Series. The goal of this paper is to determine the precise geometric equivalence relation ˜ on paths so that

$$X \,\tilde{}\, Y \iff \mathbf{X} = \mathbf{Y}$$

and hence identify the sense in which the signature of a path determines the path.

We will prove that $\mathbf{X}$ completely determines X as a control in the sense of [5]. That is if X acts on a system through the differential equation $dY_u = f(Y_u)dX_u$, where Y_u represents the state of the system at time u, then the state of the system after the application of X is completely determined by the signature of X. If two paths have different signatures, then they

---

*The Mathematical Institute, University of Oxford, 24 - 29 St. Giles, Oxford OX1 3LB.

will always be different as controls because the signature $\mathbf{X}(=\mathbf{X}_T)$ can itself be computed by solving a differential equation:

$$\begin{aligned} d\mathbf{X}_u &= \mathbf{X}_u \otimes dX_u \\ \mathbf{X}_0 &= (1,0,0,\ldots) \end{aligned}$$

The first detailed studies of the iterated integrals of paths that we consider are due to K. T. Chen. In [1] Chen proves the following theorems:

Chen Theorem 1: Let $dx_1, \cdots, dx_d$ be the canonical 1-forms on $\mathbb{R}^d$. If $\alpha, \beta \in [a,b] \to R^d$ are sufficiently smooth paths, then the iterated integrals of the vector valued paths $\int_{\alpha(0)}^{\alpha(t)} dx$ and $\int_{\beta(0)}^{\beta(t)} dx$ 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 sufficiently smooth 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 $\mathbb{R}^d$, and a continuous increasing change of parameter $\lambda : [a,b] \to [a,b]$ such that $\alpha = T\beta\lambda$.[1]

Our goal is to extend these theorems and establish in a precise sense, that the signature of a path of bounded variation determines that path up to tree-like extensions and determines it completely as a control. As a corollary, we will see that each equivalence class has a canonical element, the (*tree*) *reduced path* and that there is a natural group structure making this space of paths into a continuous analogue of the free group.

The methods in [1] exploit the piecewise smooth nature of the paths he considers and we cannot see how they might be used directly to prove our result. Our proof has commonality with [1] but also relies on various analytic tools (the Lebesgue differentiation theorem, the area theorem), a mollification of paths that retain certain deeply nonlinear properties of these paths, as well as basic properties of hyperbolic space. It also involves continuous trees. (These are called R-trees in parts of the mathematical literature, but occur in many other settings). We will particularly need the continuous trees coded by positive continuous functions on the line developed for instance in [4].

1.1 The discrete analogue.

Consider an alphabet $A = \{a, b, \ldots\}$ and new letters $A^{-1} = \left\{a^{-1}, b^{-1}, \ldots\right\}$. The set Ω of words in $A \cup A^{-1}$ have a natural multiplication (concatenation) and an equivalence relation that respects this multiplication.

Definition 1.2 *A word $w \in \Omega$ is said to cancel 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 $(e \ldots f)$

$$(a \ldots b) \,\tilde{}\, (e \ldots f)$$

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

---

[1] We borrow these formulations from the Math Review of the paper.

It is well known that the free group F_A can be identified as $\Omega/\tilde{}$.

Notice that there is an obvious bijection between words in Ω of length K, and the piecewise linear paths x_u defined for $u \in [0,K]$ which satisfy $x_0 = 0$ and $\|x_k - x_{k+1}\| = 1$, are linear on each interval $[k, k+1] \subset [0,K]$, and have $\mathrm{x}_k \in \mathbb{Z}^{|A|}$ for each $k \in \mathbb{N} \cap [0,K]$.

The equivalence relation between words can be re-articulated in this language: Consider two such paths x and y, 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 "contracted" step by step to a point (while keeping the deformations inside the graph of z). The converse is also true.

This converse is instructive and quite easy so we sketch a proof: The paths x_t are the unit speed trajectories in the 'jungle gym' Γ that only change direction at points of the integer lattice. The simply connected covering surface of the 'jungle gym' is the regular tree T of valence $2|A|$. Fix some pre-image in T of $0 \in \Gamma$ and call it the root of the tree. We may lift any path $x_t \in \Gamma$ starting at 0 to a path $\tilde{x}_t$ in T starting at the root. The path $\tilde{x}$ is unique. Now suppose that x is contractible to a point in Γ relative to its end points, then the monodromy theorem allows us to lift the homotopy as well and hence deduce that $\tilde{x}(0) = \tilde{x}(K)$. Thus $\tilde{x}$ is a rooted loop in the tree, traversed at unit speed and only changing direction at vertices of T. It is easy to see that its range in T is also a rooted tree, and that it can be contracted, edge by edge, to the root of T. These contractions systematically reduce $\tilde{x}$, and looking at the projections of these contractions into Γ we see that the word associated to x is reducible to the empty word.

Remark 1.3 *The words equivalent to the empty word correspond with loops x_t that can be factored into a loop in T and a projection π of T.*

$$\begin{tikzcd}
{[0,K]} \arrow[r, dashed, "?"] \arrow[dr, "x"] & T \arrow[d, "\pi"] \\
 & \Gamma
\end{tikzcd}$$

We call such factorisable paths tree-like; a loop in a tree can be thought of as representing an exploration and mapping out of a sub-tree, in which each point is visited at least twice. With a slight metric refinement, and allowing continuous trees in place of T, this approach will be used in Section 2 to provide a robust and tractable general definition of *tree-like path.*

1.2 Results

Our goal in this paper is to prove that the equivalence relation defined by tree-like paths extends to the case of paths with finite length and is a good equivalence relationship; one can quotient by it to get a group and also give a nice algebraic representation. Unfortunately, the proof is not a simple induction following the algebraic example.

Our goal in this paper is to make a precise definition of Lipschitz tree-like path and to prove the following.

Theorem 1.4 *Let X be a bounded variation path of finite length in $\mathbb{R}^d$. The path X is tree-like if and only if the signature of X is $\mathbf{0} = (1, 0, 0, \dots)$.*

This ensures that the signatures $\mathbf{X}$, $\mathbf{Y}$ of two paths of finite length are equal if and only if the concatenation of X and 'Y run backwards' is a Lipschitz tree-like path. The 'if' part is essentially trivial given current understanding. The harder and more interesting part is the proof that if a path has all the algebraic coefficients in their signature 0, then it is a tree-like path. An immediate consequence of this result is the following.

Corollary 1.5 *The relation* $\mathbf{X} \sim \mathbf{Y}$ *defined so that two paths of finite length are related if and only if the concatenation of* X *and 'Y run backwards' is a Lipschitz tree-like path is an equivalence relation.*

Another important consequence is that there is an analogue of the reduced word in the discrete setting.

Corollary 1.6 *Given any bounded variation path there exists a unique path of minimal length, called the reduced path, with the same signature.*

1.2.1 Continuous analogues of the free group

The main insights characterising the free group F_A arise out of either

(a) its universal property that every function f taking A into a group G extends uniquely to a group homomorphism of F into G, or

(b) its construction out of the monoid of words through the systematic reduction of words to their reduced representatives as set out in definition 1.2. A particular point is that the reduction process (which is not unique) always produces as its terminal result the same word - justifying the phrase *reduced word*.

There are a number of possible spaces of paths that could be used to replace the words with letters in A. If we are going to construct an analogue of the free group, then one must provide a relation ˜ on the space of paths. A key test in the continuous case will be to establish the transitivity of the relation one defines as it is not (quite) trivial to show this even for words. Another crucial point is to make sure the quotient space has a good topology.

If one can introduce a natural equivalence relation, then one should be able to prove the analogue of the universal property (a). The group of equivalence classes of paths should at least have the property that any linear map from $\mathbb{R}^d$ into the Lie algebra of a Lie group should induce a canonical (i.e. in some sense unique) homomorphism from path space with concatenation into the group. Moreover it would be very satisfactory if one could, as in the discrete case, identify a canonical representative for each class - the *reduced path* and also identify the equivalence class of paths associated to the identity element.

It will follow from our work that we can then define a continuous analogue of the 'free' group with all these properties. The bounded variation paths, with the operation of concatenation, when quotiented by the equivalence relation of Corollary 1.5 form a 'free' group. The group is faithfully represented by the signature of the path. The 'kernel' of the map is the space of tree-like paths of 1-variation. The reduced path is a canonical representative of each equivalence class (a non-trivial fact that is essentially equivalent to the full content of this paper). The universal property then follows from the ability to solve differential equations, and from the uniqueness of their solutions:

Theorem 1.7 (Cartan development) *Let* θ *be a linear map of* $\mathbb{R}^d$ *to the Lie algebra* $\mathfrak{g}$ *of a Lie group* G *(the infinitesimal version of a function taking letters to elements of the group) 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* $\tilde{\theta} : X \rightarrow Y_T$. *This map* $\tilde{\theta}$ *is a homomorphism from the space of paths with concatenation to* G.

It will be obvious from our arguments that tree-like paths are mapped by Cartan development to the identity element and so the map is indeed defined on the quotient space of reduced paths and is a homomorphism of that group.

One might instead consider the space of *continuous* paths with the uniform topology as a natural generalisation of words - certainly concatenation makes them a monoid. However, despite their popularity in homotopy theory, there seems little hope that a 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. For paths that fail to have finite 2-variation, and certainly for continuous functions, we know that there is no canonical Cartan development into the Heisenberg group

As an alternative, one might instead consider the space of *geometric p-rough* paths following the definition of Lyons [5]. These "paths" also form a monoid under concatenation and 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 so they certainly have the analogous universal property to the one mentioned in (a). The homomorphism taking a path to its signature also extends in a canonical way to this space. However, to date, we have not been able to characterise the "universal" kernel of negligible paths; the paths whose development will always produce a null effect. Our theorem precisely identifies this class in the context of $p = 1$ or bounded variation but our proof uses the 1-dimensionality of the path in an essential way. An extension to p-rough paths with $p > 1$ would be interesting but requires new ideas to account for the fact that these rougher paths are of higher "dimension". We have some hope that this can be done - and in this case it would suggest that the union over p of the spaces of geometric p-rough paths (ie the space of geometric rough paths) would be a useful space that could replace the continuous paths for some purposes.

1.3 Three open questions

As indicated above a natural open question is:

Problem 1.8 *Given a path X of finite p-variation for some $p > 1$, is the triviality of the signature of X equivalent to the path being tree-like?*

By Corollary 1.6 in each ˜-equivalence class of bounded variation paths, there is a unique shortest one - the reduced path.

Problem 1.9 *How does one reconstruct the reduced path from its signature?*

The question is interesting even for the paths in Γ associated to the discrete free group.

A related question is to:

Problem 1.10 *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.*

Some interesting progress in this direction can be found in [2].

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 also seems reasonable 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.
3. We note that the multiplication operation acting on reduced paths is not continuous in the bounded variation topology.

1.4 Outline

We set out the steps required to prove our core theorem.

The overall idea is to treat piecewise linear paths as a special case. We begin by introducing tree-like paths in Section 2 and establishing the key properties that we need. In Section 3 we prove that any path of bounded variation and trivial signature can, after reparameterisation, be approximated by weakly piecewise linear paths with trivial signature. Then, in Sections 4 and 5, a quite separate argument shows that any weakly piecewise linear path with trivial signature is tree-like. As it is clear from the definitions, that uniform limits of tree-like paths with uniformly bounded length are themselves tree-like, the argument is complete. We draw together all the parts to give the proofs of our main Theorem and Corollaries in Section 6.

We point out that the proof hides some subtleties - for example, one could avoid hyperbolic spaces in the second part, and recover the length of a piecewise linear path directly from the asymptotic form of the signature. However, the hyperbolic argument and the closed form special functions allow one to easily perform cancellations that appear in more bare handed approaches as highly oscillatory sums whose convergence properties deteriorate as one goes to the limit.

2 Tree-Like paths

In this section we work in a more general setting. Suppose that $X_{t\in[0,T]}$ is a path in a Banach or metric space E.

Definition 2.1 *$X_t,\ t\in[0,T]$ is a* tree-like *path if there exists a positive real valued continuous function h defined on $[0,T]$ such that $h(0)=h(T)=0$ and such that*

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

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 (i.e. rectifiable).*

Proposition 2.2 *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

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

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

$$\tau_t(x)=\sup\{u\in[0,t]\,|g_t(u)=x\},\quad x\in[0,h(t)]. \tag{2.1}$$

As g_t is monotone and continuous

$$\tau_t(x)=\inf\{u\in[0,t]\,|g_t(u)>x\} \tag{2.2}$$

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 $t_1 = \tau_{t_0}(x_1)$, 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. Now consider

$$\begin{aligned} g_t(v) &= \inf_{v \leq u \leq t} h(u), \ v \in [0, t], \\ g_s(v) &= \inf_{v \leq u \leq s} h(u), \ v \in [0, s]. \end{aligned}$$

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 controlled 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

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

and in particular

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

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

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

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,

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

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 $D = \{s \leq \cdots < t_i < t_{i+1} < \cdots \leq t\}$. 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 $\inf_{s\leq u\leq t}\left(\tilde{h}\left(u\right)\right) = \tilde{h}\left(r\right)$. So $\tilde{h}\left(s\right) + \tilde{h}\left(t\right) - 2\inf_{s\leq u\leq t}\left(\tilde{h}\left(u\right)\right)$ 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 controlled by $\omega_X\left(\sup I, t\right)$, as the path $X_{\tau_t(.)}|_{[h(r),h(t)]}$ is a subordinator 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 $\omega_X\left(s, \inf I\right)$ to conclude that

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

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

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

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 subordinator (but backwards) of $X|_{[s,\inf I]}$. ∎

The property of being tree-like is reparameterisation 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 the Appendix).

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 $h\left(t\right) = d\left(0, \phi\left(t\right)\right)$. 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 2.3 *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 $\tilde{\mathcal{D}} = \{t_0 \leq u_1 \ldots \leq t_{n-1} \leq u_n \leq t_n\}$. Relabel the points of $\tilde{\mathcal{D}} = \{v_0 \leq v_1 \ldots \leq v_m\}$ Then

$$\sum_{\mathcal{D}} \left\| X_{t_i} - X_{t_{i-1}} \right\| \leq \sum_{\tilde{\mathcal{D}}} \left| h(v_i) - h(v_{i-1}) \right|. \tag{2.3}$$

∎

One can always re-parameterise a continuous path of bounded variation in a finite dimensional Euclidean space so that it is continuous and traversed at speed one.[2]

We now prove a compactness result.

Lemma 2.4 *Suppose that $\{h_n\}$ are 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[3]. 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 equicontinuous, and in view of (2.3) the X_n are as well. Our hypotheses ensure they are "compactly" bounded and so we may 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 semicontinuous 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 $[0,T]$. To finish the proof we need to demonstrate that Y is a Lipschitz tree-like path with height function h.

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 so

$$\left\| X_{n(k),t} - X_{n(k),s} \right\| \leq h_{n(k)}(s) + h_{n(k)}(t) - 2 \inf_{u\in[s,t]} h_{n(k)}(u)$$

holds for each k. Taking the limit $k \to \infty$, one has

$$\left\| Y_t - Y_s \right\| \leq h(s) + h(t) - 2 \inf_{u\in[s,t]} h(u)$$

and h is a height function for Y confirming its tree-like structure. ∎

Corollary 2.5 *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 Proposition 2.2 and Lemma 2.3. ∎

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

---

[2] However the reparameterisation involves traversing intervals where the path is constant at "infinite" speed. In our context it is clear that one can re-parameterise so that h has constant speed while X is still kept continuous.

[3] This would be automatic if E were finite dimensional.

3 Approximation of the path

3.1 Representing the path as a line integral against a rank one form

Let γ be a path of finite variation in a finite dimensional Euclidean space E. Amongst the continuous arcs Fréchet equivalent to γ there is a unique one, "the representation in terms of arc length", parameterised at unit speed. We work with this parameterisation, so that

$$\omega_\gamma(s,t) = \sup_{\mathcal{D}\subset[s,t]} \sum_{\mathcal{D}} \left\| \gamma_{t_i} - \gamma_{t_{i-1}} \right\| = t - s.$$

We assume that γ has total length T and its parameter set is $[0,T]$. We note that the signature of γ is unaffected by this choice of parameterisation.

Definition 3.1 *Let $\gamma\left([0,T]\right)$ denote the range of γ in V and let the occupation measure μ on $(V,\ \mathcal{B}(V))$ be denoted*

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

Let $n\left(x\right)$ be the number of points on $[0,T]$ corresponding under γ to $x \in E$. By the area formulae [7] p125-126, one has the total variation, or length, of the path γ is given by

$$Var\left(\gamma\right) = \int n\left(x\right) \Lambda_1\left(dx\right), \tag{3.1}$$

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

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

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

Lemma 3.2 *The image under γ of a Lesbegue null set is null for μ. That is to say $\mu(\gamma(N)) = \left|\gamma^{-1}\gamma\left(N\right)\right| = 0$ if $|N| = 0$.*

Definition 3.3 *We will say that $N \subset [0,T]$ is γ-stable if $\gamma^{-1}\gamma\left(N\right) = N$.*

As a result of Lemma 3.2 we see that any null set can always be enlarged to a γ-stable null set.

The Lesbegue 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 3.4 *There is a set G of full μ measure in E so that γ is differentiable with $\left|\gamma'\left(t\right)\right| = 1$ whenever $\gamma\left(t\right) \in G$. We set $M = \gamma^{-1}G$. M is γ-stable.*

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

Theorem 3.5 *The set of pairs (s,t) of distinct times in $M \times M$ for which*

$$\begin{aligned} \gamma\left(s\right) &= \gamma\left(t\right) \\ \gamma'\left(s\right) &\neq \gamma'\left(t\right) \end{aligned}$$

is countable.

Proof. If (s,t) is such a pair, then $\gamma'(s) \neq \gamma'(t)$. So there is an $\varepsilon > 0$ such that the cones

$$\begin{aligned} |y-\gamma(s)| &< (1+\varepsilon)(y-\gamma(s)).\gamma'(s) \\ |y-\gamma(t)| &< (1+\varepsilon)(y-\gamma(t)).\gamma'(t) \end{aligned}$$

intersect only at $\gamma(s) = \gamma(t)$.

As γ is differentiable with non-zero derivative there is a $\delta_1 > 0$ so that, if $0<|s-\tilde{s}| < \delta_1$, then $\gamma(\tilde{s})$ is in the cone

$$|y-\gamma(s)| < (1+\varepsilon)(y-\gamma(s)).\gamma'(s)$$

and $\gamma(\tilde{s}) \neq \gamma(s)$. Similarly there is a δ_2 so that, if $0<\left|t-\tilde{t}\right| < \delta_2$, then $\gamma\left(\tilde{t}\right)$ is in the cone

$$|y-\gamma(t)| < (1+\varepsilon)(y-\gamma(t)).\gamma'(t),$$

and $\gamma\left(\tilde{t}\right) \neq \gamma(t)$. As a result, $\gamma(\tilde{s}) \neq \gamma\left(\tilde{t}\right)$ or $\gamma(t)$ and $\gamma\left(\tilde{t}\right) \neq \gamma(\tilde{s})$ or $\gamma(s)$.

In other words, if $\gamma(s) = \gamma(t)$ and $\gamma'(s) \neq \gamma'(t)$, then there exist $\delta_1 > 0$ and $\delta_2 > 0$ so that, if

$$\begin{aligned} \gamma(\tilde{s}) &= \gamma\left(\tilde{t}\right), \\ |s-\tilde{s}| &< \delta_1, \\ \left|t-\tilde{t}\right| &< \delta_2, \end{aligned}$$

then $s = \tilde{s}$ and $t = \tilde{t}$. In particular the pairs (s,t) are isolated in $[0,T] \times [0,T]$ and so are countable in number. ▮

Up to sign and with countably many exceptions, the derivative of γ does not depend 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 3.6 *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*.*

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

Corollary 3.7 *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 very interesting 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 $Hom(V,V)$.

Definition 3.8 *We define the tangential projection* 1*-form* ω*. Let* ξ *be a unit strength vector field on* G *with* $[\xi]_{\tilde{\ }_{\pm}} = \phi$*. Then*

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

defines a vector 1-form. The 1-form depends on ϕ*, but is otherwise independent of the choice of* ξ*.*

The 1-form ω is the projection of u onto the line determined by $\phi(g)$.

Lemma 3.9 *The tangential projection* 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, using the fundamental theorem of calculus for Lipschitz functions,

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

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.

3.2 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 co-ordinate 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 $\tilde{Z}$

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

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 3.10 *The r'th iterated integral of a bounded variation path* γ *over the interval* $[s,t]$ *is defined to be*

$$\gamma_{s,t}^{r} := \underset{s<u_1<\ldots<u_r<t}{\int\cdots\int} d\gamma_{u_1}\ldots d\gamma_{u_r} \in V^{\otimes r}.$$

The collection of these integrals, $\boldsymbol{\gamma}_{s,t}= \left(1, \gamma_{s,t}^{1}, \gamma_{s,t}^{2}, \ldots, \gamma_{s,t}^{r}, \ldots\right)$, *is known as the signature of* γ *over the interval* $[s,t]$.

Definition 3.11 *The truncated or n-signature* $\boldsymbol{\gamma}_{s,t}^{(n)}= \left(1, \gamma_{s,t}^{1}, \gamma_{s,t}^{2}, \ldots, \gamma_{s,t}^{n}\right)$ *is the projection of* $\boldsymbol{\gamma}_{s,t}$ *to the algebra* $T^{(n)}\left(V\right) := \bigoplus_{r=0}^{r=n} V^{\otimes r}$ *of tensors with degree at most n.*

Definition 3.12 *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*

$$\gamma_{s,t}^{\mathbf{e}} := \underset{s<u_1<\ldots<u_r<t}{\int\cdots\int} d\gamma_{u_1}^{e_1}\ldots d\gamma_{u_r}^{e_r} = \left\langle \mathbf{e}, \boldsymbol{\gamma}_{s,t} \right\rangle.$$

Lemma 3.13 *The map* $\mathbf{e} \to \langle \mathbf{e}, \cdot \rangle$ *extends uniquely as a linear map from* $T^{(n)}\left(V^*\right)$ *to the space of real valued functions on paths of bounded variation.*

Proof. For each γ the map $\mathbf{e} \to \left\langle \mathbf{e}, \boldsymbol{\gamma}_{s,t} \right\rangle$ is obviously multilinear; hence, for every n it extends in a unique way to a linear function on $T^{(n)}\left(V^*\right) \overset{\boldsymbol{\gamma}_{s,t}}{\to} \mathbb{R}$ and for each $\tilde{\mathbf{e}} \in T^{(n)}\left(V^*\right)$ defines a function on path space

$$\boldsymbol{\gamma}_{s,t} \to \boldsymbol{\gamma}_{s,t}^{\mathbf{e}} := \left\langle \mathbf{e}, \boldsymbol{\gamma}_{s,t} \right\rangle.$$

∎

Definition 3.14 *We call these functionals coordinate iterated integrals.*

The justification for this definition is that they are scalar quantities and because, in the case of paths $\gamma = \left(\gamma_u\left(1\right), \ldots, \gamma_u\left(d\right)\right)$ in $\mathbb{R}^d$, any iterated integral of coordinates

$$\gamma_{s,t}^{(i_1,\ldots,i_r)} := \underset{s<u_1<\ldots<u_r<t}{\int\cdots\int} d\gamma_{u_1}\left(i_1\right)\ldots d\gamma_{u_r}\left(i_r\right), \qquad i_j \in 1, \ldots, d$$

is of this form. One simply identifies i with projection to the i'th coordinate. Moreover any other functional $\boldsymbol{\gamma} \to \boldsymbol{\gamma}_{s,t}^{\mathbf{e}}$ is a linear combination of these basic ones.

Definition 3.15 *A shuffle of a pack of n cards and a pack of m cards is a pair of* increasing and injective *functions* $\left(\pi_1, \pi_2\right)$,*with domains* $(1, \ldots, r)$ *and* $(1, \ldots, s)$ *respectively, with common co-domain* $(1, \ldots, r+s)$ *and* disjoint ranges.

It is clear that there is a correspondence between shuffles and colourings of the set $(1, \ldots, r+s)$ into r of one and s of a second specified colors. In particular, there are as many shuffles as there are ways of choosing r from $r+s$.

Given two tensors $\mathbf{e}$, $\mathbf{f}$ there is a natural product $\mathbf{e} \sqcup \mathbf{f}$, 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) 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

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

one defines a multilinear map of $V^{\otimes r} \times V^{\otimes s} \rightarrow V^{\otimes r+s}$ defined for all choices of r and s.

Definition 3.16 *The unique extension of* $\sqcup$ *to a map from* $T(V) \times T(V) \rightarrow T(V)$ *is called the shuffle product.*

The shuffle product is well known to algebraists and we only define it here for the convenience of analysts for whom it is probably less familiar.

Let us return to iterated integrals. Recall that $\boldsymbol{\gamma}^{\mathbf{e}}_{s,t}$ is the linear functional $\mathbf{e}$ on the tensor algebra, contracted with the signature of a path. As a functional on paths it has an important property:

Lemma 3.17 *The point-wise product of coordinate iterated integrals is itself such a functional:*

$$\boldsymbol{\gamma}^{\mathbf{e}}_{s,t} \boldsymbol{\gamma}^{\mathbf{f}}_{s,t} = \boldsymbol{\gamma}^{\mathbf{e} \sqcup \mathbf{f}}_{s,t},$$

where $\mathbf{e} \sqcup \mathbf{f}$ *is the shuffle product.*

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

$$\boldsymbol{\gamma}^{\mathbf{e}}_{s,t} \boldsymbol{\gamma}^{\mathbf{f}}_{s,t} = \int\limits_{s<u_1<t} d\boldsymbol{\gamma}^{\mathbf{e}}_{s,u_1} \int\limits_{s<u_2<t} d\boldsymbol{\gamma}^{\mathbf{f}}_{s,u_2},$$

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

$$\begin{aligned} & \int\limits_{s<u_1<t} d\boldsymbol{\gamma}^{\mathbf{e}}_{s,u_1} \int\limits_{s<u_2<t} d\boldsymbol{\gamma}^{\mathbf{f}}_{s,u_2} \\ = & \iint\limits_{\substack{s < u_1 < t \\ s < u_2 < t}} d\boldsymbol{\gamma}^{\mathbf{e}}_{s,u_1} d\boldsymbol{\gamma}^{\mathbf{f}}_{s,u_2} \\ = & \int\limits_{\substack{s < v_1 < \ldots < v_r < t \\ s < w_1 < \ldots < w_s < t}} \cdots \int d\gamma^{e_1}_{v_1} \ldots d\gamma^{e_r}_{v_r} d\gamma^{f_1}_{w_1} \ldots d\gamma^{f_s}_{w_s}. \end{aligned}$$

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) we have

$$\iint\limits_{s<u_1,u_2<t} d\boldsymbol{\gamma}^{\mathbf{e}}_{s,u_1} d\boldsymbol{\gamma}^{\mathbf{f}}_{s,u_2} = \boldsymbol{\gamma}^{\mathbf{e}\sqcup\mathbf{f}}_{s,t}.$$

∎

Corollary 3.18 *Any polynomial in coordinate iterated integrals is a coordinate iterated integral.*

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

Proposition 3.19 *The iterated integral*

$$\int\cdots\int\limits_{s<u_1<\ldots<u_r<t} d\boldsymbol{\gamma}^{\mathbf{e}_1}_{s,u_1} \ldots d\boldsymbol{\gamma}^{\mathbf{e}_r}_{s,u_r} \tag{3.2}$$

is itself a coordinate iterated integral.

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

$$\iint\limits_{s<u_1<u_r<t} d\boldsymbol{\gamma}^{\mathbf{e}}_{s,u_1} d\boldsymbol{\gamma}^{\mathbf{f}}_{s,u_2},$$

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

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

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

$$\begin{aligned} \iint\limits_{s<u_1<u_2<t} d\boldsymbol{\gamma}^{\mathbf{e}}_{s,u_1} d\boldsymbol{\gamma}^{\mathbf{f}}_{s,u_2} &= \boldsymbol{\gamma}^{\left(\mathbf{e}\sqcup\tilde{\mathbf{f}}\right)\otimes f_s}_{s,t} \\ \tilde{\mathbf{f}} &= f_1 \otimes \ldots \otimes f_{s-1}. \end{aligned}$$

∎

From this it is, of course, clear that

Lemma 3.20 *If a path segment has trivial signature, then it has all iterated integrals of iterated integrals zero.*

3.3 Bounded measurable and integrable forms

Recall that γ is a path of finite length, and that it is parameterized at unit speed. The occupation measure is μ and has total mass equal to the length T of the path γ.

Proposition 3.21 *Let $\omega \in L^1(V, \mathcal{B}(V), \mu)$ be a μ-integrable 1-form with values in a vector space W. Then the indefinite line integral $y_t := \int_0^t \omega(d\gamma_t)$ is well defined, 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 equiped with the operator norm) is defined dt almost everywhere. Since ω is integrable, it is measurable, and hence $\omega(\gamma_t)$ is measurable on $[0, T]$. In addition

$$\int_V \|\omega(y)\| \mu(dy) = \int_0^T \|\omega(\gamma_t)\| dt.$$

Since γ has finite variation and is parameterized at unit speed, it is differentiable almost everywhere and its derivative is measurable with unit length dt almost surely. Hence $\omega(\gamma_t)(\dot{\gamma}_t)$ is measurable and dominated by $\|\omega(\gamma_t)\|$, which is an integrable function, and hence $\omega(\gamma_t)(\dot{\gamma}_t)$ is integrable (at least in finite dimensions). Thus the line integral can be defined to be

$$y_t = \int_0^t \omega(\gamma_t)(\dot{\gamma}_t) dt.$$

∎

If ω is bounded, then the path y is parameterized at finite speed. However, in general, it is not parameterized at unit speed. The length of y is at most $\|\omega\|_{L^1(V,\mathcal{B}(V),\mu)}$:

Corollary 3.22 *The map taking a μ-integrable one form ω to the line integral $y_t = \int_0^t \omega(\gamma_t)(\dot{\gamma}_t) dt$, for $t \leq T$, is continuous in 1-variation and if*

$$\begin{aligned} y_t &= \int_0^t \omega(\gamma_t)(\dot{\gamma}_t) dt \\ \tilde{y}_t &= \int_0^t \tilde{\omega}(\gamma_t)(\dot{\gamma}_t) dt, \end{aligned}$$

then

$$\|y_t - \tilde{y}_t\|_{1\text{-}Var} \leq \|\omega - \tilde{\omega}\|_{L^1(V,\mathcal{B}(V),\mu)}. \tag{3.3}$$

Proof. It is enough to prove (3.3) in the form

$$\|y_t\|_{\text{1-Var}} \leq \|\omega\|_{L^1(V,\mathcal{B}(V),\mu)}.$$

Let $\mathcal{D} = \{0 = t_0 \leq \ldots \leq t_k \leq \ldots \leq t_r\}$ be a partition of $[0, T]$. Then

$$\begin{aligned} \sum_{\mathcal{D}} \left\| y_{t_{k+1}} - y_{t_{k+1}} \right\|_W &\leq \sum_{\mathcal{D}} \left\| \int_{t_k}^{t_{k+1}} \omega(\gamma_t)(\dot{\gamma}_t) dt \right\| \\ &\leq \sum_{\mathcal{D}} \int_{t_k}^{t_{k+1}} \|\omega(\gamma_t)\| dt \\ &\leq \int_0^T \|\omega(\gamma_t)\| dt \end{aligned}$$

∎

Proposition 3.23 *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 line integral $\int \omega_n(d\gamma_t)$ can be expressed as

$$\underset{0<u_1<\ldots<u_r<T}{\int \cdots \int} \omega_n(\gamma_{u_1}) \otimes \ldots \otimes \omega_n(\gamma_{u_r}) \ (\dot{\gamma}_{u_1}) \ldots (\dot{\gamma}_{u_r}) \ du_1 \ldots du_r$$

and since ω_n converge in $L^1(V, \mathcal{B}(V), \mu)$, it follows that from the definition of μ that $\omega_n(\gamma_u)$ 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 $\|\dot{\gamma}_u\| = 1$ for almost every u, Fubini's theorem implies that

$$\omega_n(\gamma_{u_1}) \otimes \ldots \otimes \omega_n(\gamma_{u_r}) \ (\dot{\gamma}_{u_1}) \ldots (\dot{\gamma}_{u_r})$$

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

$$\omega(\gamma_{u_1}) \otimes \ldots \otimes \omega(\gamma_{u_r}) \ (\dot{\gamma}_{u_1}) \ldots (\dot{\gamma}_{u_r}).$$

Thus, integrating over $0 < u_1 < \ldots < u_r < T$, the proposition follows. ∎

Corollary 3.24 *If γ has trivial signature, then so does $\int \omega(d\gamma_t)$. That is to say, for each r,*

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

Proof. It is a consequence of Proposition 3.23 that the set of $L^1(V, \mathcal{B}(V), \mu)$ forms producing line integrals having trivial signature is closed. By Lusin's theorem, one may approximate, in the $L^1(V, \mathcal{B}(V), \mu)$ norm, any bounded and measurable form by bounded continuous forms with 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 ω. 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, then these integrals will all be zero. It follows from the $L^1(V, \mathcal{B}(V), \mu)$ continuity of the truncated signature, that the signature of the path formed by taking the line integral against any form ω in $L^1(V, \mathcal{B}(V), \mu)$ will always be trivial. ∎

3.4 Approximating rank one 1-forms

Definition 3.25 *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 $\dim(\omega(V)) \leq k$ 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 so that the μ measure of the compliment is arbitrarily small.

The following easy consequence of Lusin's theorem will be essential:

Lemma 3.26 *Let ω be a measurable 1-form ω in $L^1(V, \mathcal{B}(V), \mu)$. For each $\varepsilon > 0$ there is a compact subset L of $\gamma[0,T]$ so that ω restricted to L is continuous, while $\int_{K \backslash L} \|\omega\| \mu(dx) < \varepsilon$.*

Lemma 3.27 *If ω is a measurable 1-form in $L^1(V, \mathcal{B}(V), \mu)$, then there are finitely many disjoint compact subsets K_i of K and a 1-form $\tilde{\omega}$, that is zero off the K_i and constant on each K_i, such that*

$$\int_K \|\omega - \tilde{\omega}\| \mu(dx) \leq 4\varepsilon$$

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

Proof. Let L be the compact subset introduced in Lemma 3.26. Now $\omega(L)$ is compact. Fix $\varepsilon > 0$ and choose $l_1, \ldots, l_n$ so that

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

and put

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

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

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

Then the K_j are disjoint and of diameter $2\varepsilon/\mu(L)$. Moreover

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

and

$$\int_{L \backslash \cup_{i=1}^n K_j} \|\omega\| \mu(dx) < \varepsilon.$$

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

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

Then

$$\begin{aligned} \int_{L \backslash \cup_{i=1}^n K_j} \|\omega - \tilde{\omega}\| \mu(dx) &< \varepsilon \\ \int_{\cup_{i=1}^n K_j} \|\omega - \tilde{\omega}\| \mu(dx) &< (2\varepsilon/\mu(L))\, \mu(L), \end{aligned}$$

using Lemma 3.26 one has

$$\int_{K \backslash L} \|\omega - \tilde{\omega}\| \mu(dx) < \varepsilon$$

and finally

$$\int_K \|\omega - \tilde{\omega}\| \mu(dx) \leq 4\varepsilon.$$

If ω had rank 1 at almost every point of K, then it will have rank 1 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. ∎

Corollary 3.28 *Every* 1*-form in* $L^1(V, \mathcal{B}(V), \mu)$ *with rank one is the* $L^1(V, \mathcal{B}(V), \mu)$ *limit of rank one* 1*-forms. For each approximating 1-form there are finitely many disjoint compact subsets* K_i *of* K *so that the 1-form* $\tilde{\omega}$ *is zero off the* K_i *and constant on each* K_i*.*

4 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 4.1 *We say the path is nondegenerate 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 colinear 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.

In this section, our goal is a quantitative proof that such a nondegenerate piecewise linear path has a non-trivial signature. We use some simple hyperbolic geometry. Fix 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[4]:

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

and note the following simple lemmas:

Lemma 4.2 *If the distance* c *from* A *to* B *is at least* $\ln\left(\frac{\cos|\theta_A|+1}{1-\cos|\theta_A|}\right)$*, 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:

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

∎

---

[4] Our source for this was http://www.maths.gla.ac.uk/˜wws/cabripages/hyperbolic/hypertrig.html

Lemma 4.3 *If* $\max(b,c) \geq \log \frac{2}{1-\cos\theta_A}$, *then* $a > \min(b,c)$.

Proof. Suppose that θ_A is fixed and the triangle has sides $a(\lambda)$, λb, λc. 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

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

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

Corollary 4.4 *If the distance* c *from* A *to* B *is at least* $\ln\left(\frac{2}{1-\cos|\theta_A|}\right)$, *then*

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

and $a \geq b$.

Corollary 4.5 *If* X_t *is a continuous piecewise geodesic path of finite length in hyperbolic space with at least one non-trivial geodesic section, and suppose that,*

1. *at each change in direction* t *the angle between the two geodesic segments:* $\angle X_{t-}X_tX_{t+}$ *is at least* $2\theta_A$ *and*
2. *that each geodesic segment has length at least* $R(\theta_A)$ $=\ln\left(\frac{2}{1-\cos|\theta_A|}\right)$,

then $d(X_0, X_T) \geq R(\theta_A)$ *and the angle between* $\overrightarrow{X_{T-}X_T}$ *and* $\overrightarrow{X_0X_T}$ *is at most* θ_A.

Proof. As the path has finite length and each segment is of length at least $R(\theta_A) > 0$ there can be at most a finite number of distinct piecewise linear segments in the path. We proceed by induction on the number of these geodesic segments in the path.

There is only ever one geodesic through two points in hyperbolic space and so the distance between the ends of a geodesic segment is always the length of the connecting segment. So if there is only one segment we can conclude from 1) that $d(X_0, X_T) \geq R(A)$. The angle between $\overrightarrow{X_{T-}X_T}$ and $\overrightarrow{X_0X_T}$ is zero.

Suppose there are N segments and the penultimate one ends and the last one begins at a time $S < T$. By the induction hypothesis we can assume that the distance $d(X_0, X_S) \geq R(\theta_A)$ and by 2) that $d(X_S, X_T) \geq R(\theta_A)$. Moreover the angle between $\overrightarrow{X_{S-}X_S}$ and $\overrightarrow{X_0X_S}$ is at most

θ_A while the angle $X_{S-}X_SX_T$ is at least $2\theta_A$ so that the angle $X_0X_SX_T$ is at least θ_A. We can apply Lemma 4.3 and deduce that

$$d(X_0, X_T) \geq R(\theta_A)$$

and that the angle $X_SX_TX_0$ is at most θ_A. ∎

Corollary 4.6 *Any non-degenerate piecewise linear path γ has non-trivial signature.*

Proof. Suppose γ is a non-degenerate piecewise linear path in V and let 2δ be the smallest angle between adjacent edges, and let $D > 0$ denote the length of the shortest edge. Choose $\lambda > R(\delta)/D$.

Now isometrically embed V into the tangent space to a fixed point in hyperbolic space. Then one can consider the development Γ of $\lambda\gamma$ to hyperbolic space. It is a piecewise geodesic path in hyperbolic space with edge lengths greater than $R(\delta)$ and with the angles between any two edges at least 2δ. Thus we can deduce that the distance $|\Gamma(0) - \Gamma(T)|$ is at least $R(\delta) > 0$.

On the other hand, we may recover the same path Γ through solving a linear differential equation in a matrix group. As a consequence the solution can, at any time, be expanded into a convergent series of iterated integrals. As a result we can conclude that, if the signature of $\lambda\gamma$ were trivial, then the development Γ must have $\Gamma(0) = \Gamma(T)$. This is a contradiction. Hence $\lambda\gamma$ and γ must have non-degenerate signatures. ∎

Corollary 4.7 *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 Corollary 4.6, 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 the path retraces its trajectory for an interval of length

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

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.

The signature $\gamma \to S(\gamma)$ is a function taking path segments with the operation $*$ to sequences of tensors with the operation $\otimes$. It is quite easy to see that it is a homomorphism (c.f. Chen's Identity [5]). As a consequence one sees that the product of the signatures associated to the segments is the signature of the concatenation of the paths and hence is trivial,

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

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

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

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. ∎

We end this section with a straightforward result which will establish half of our main theorem.

Lemma 4.8 *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 reparameterisation 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, using the area formula (3.1), we can find finitely many points u_n within δ of one another and increasing in $[0, T]$ so that h takes the value $h(u_n)$ only finitely many times and only at the times u_n. Consider the path γ_n that is linear on the intervals (u_n, u_{n+1}) and agrees with γ at the times u_n. Define h_n similarly. Then h_n is a height function for γ_n and so 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 γ_n is a martingale with respect to the filtration determined by the successive time partitions, applying the martingale convergence theorem, it follows that γ_n converges to γ in L_1. ∎

Corollary 4.9 *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 4.8 proves that any Lipschitz tree-like path can be approximated by piecewise linear tree-like paths, 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.

5 Weakly piecewise linear paths

Paths that lie in lines are special.

Definition 5.1 *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 5.2 *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 5.3 *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 5.4 *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 5.5 *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 5.6 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, for excision we require that $\gamma|_{[s,t]}$ is a loop.

We have the following two easy lemmas:

Lemma 5.7 *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 $[0,T]$, 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

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

and

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

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 5.8 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. The point is that 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 5.9 *Suppose $\gamma \in V$ is a path taking $[0, T]$ to V and that $\tau = \gamma|_{[s,t]}$ is a path with trivial signature. Then γ has trivial signature if and only if γ with the segment $[s, t]$ excised has trivial signature.*

Proof. This is also easy. If $S(\gamma)$ is the signature of a path γ in the (truncated) tensor algebra, then S is a homomorphism (see the remarks about Chen's identity in the proof of Corollary 4.7). We see that

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

and by hypothesis $S(\tau)$ is the identity in the tensor algebra. Therefore

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

∎

Definition 5.10 *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 construct the height function. As before, every such path admits a unique partition so that

Lemma 5.11 *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 linear segments associated to $\left[\gamma_{t_{i-1}}, \gamma_{t_i}\right]$ and $\left[\gamma_{t_i}, \gamma_{t_{i+1}}\right]$ are not colinear for any $0 < i < r$.*

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

Definition 5.12 *We say γ is fully non-degenerate if, in addition, $\gamma_{t_{i-1}} \neq \gamma_{t_i}$ for every $0 < i \leq r$.*

Lemma 5.13 *If γ is a weakly linear path with trivial signature and at least one segment, then there exist $0 < i \leq r$ 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 Corollary 4.5 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 5.14 *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 5.13, 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 5.3 was indeed controlled by the variation of the loop.

---

[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.

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 5.9, $\hat{\gamma}$ 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 to $[t_{i-2},t_{i-1}]$ and $[t_i,t_{i+1}]$ are colinear 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 5.7 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. ∎

6 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 1.4. Corollary 4.9 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 3.9 we can write the path as an integral against a rank one 1-form. By Corollary 3.28 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 3.24, these approximations have trivial signature. By Proposition 5.14 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 parameterising the paths at unit speed and using Lemma 2.4 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 X times the inverse of the signature of Y 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. If $X \sim Y$, then $\mathbf{XY}^{-1} = \mathbf{0}$. The path run backward has signature $\mathbf{YX}^{-1} = -\mathbf{XY}^{-1} = \mathbf{0}$.
2. $\mathbf{XX}^{-1} = \mathbf{0}$ by definition.
3. If $X \sim Y$ and $Y \sim Z$, then XY^{-1} and YZ^{-1} are tree-like and so have trivial signature. Thus

$$\begin{aligned}
\mathbf{XY}^{-1} &= \mathbf{0}\\
\mathbf{YZ}^{-1} &= \mathbf{0}\\
\left(\mathbf{XY}^{-1}\right)\left(\mathbf{YZ}^{-1}\right) &= \mathbf{0}\\
\left(\mathbf{XY}^{-1}\right)\left(\mathbf{YZ}^{-1}\right) &= \mathbf{X}\left(\mathbf{Y}^{-1}\mathbf{Y}\right)\mathbf{Z}^{-1}\\
&= \mathbf{X}\left(\mathbf{0}\right)\mathbf{Z}^{-1}\\
&= \mathbf{XZ}^{-1}
\end{aligned}$$

and hence $XZ^{-1}$is tree-like as required. ∎

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

1. we can reparameterise the paths to have unit speed and thereafter to be constant. Then by the compactness of the equivalence classes of paths with the same signature and bounded length, 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 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 treelike. Now we know from the associativity of the product in the tensor algebra that the signature is not changed by excision of a treelike 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! ∎

A Appendix

A.1 Trees and paths - background information

We have shown in this paper that trees have an important role as the negligible sets of control theory, quite analogous to the null sets of Lebesgue integration. The trees we need to consider are *analytic* objects in flavour, and not the finite combinatorial objects of undergraduate courses. In this appendix we collect together a few related ways of looking at them, and prove a basic characterisation generalising the concept of height function.

We first recall that

- Graphs (E,V) that are acyclic and connected are generally called *trees*. If such a tree is non-empty and has a distinguished vertex $\mathbf{v}$ it is called a *rooted tree*.
- A rooted tree induces and is characterised by a *partial order on V with least element* $\mathbf{v}$. The partial order is defined as follows

  $$a\preceq b \quad \text{if the circuit free path from the root } \mathbf{v}\to b \text{ goes through } a.$$

  This order has the property that for each fixed b *the set* $\{a\preceq b\}$ *is totally ordered* by $\preceq$.

  Conversely any partial order on a finite set V with a least element v and the property that for each b the set $\{a\preceq b\}$ is totally ordered defines a unique rooted tree on V. One of the simplest ways to construct a tree is to consider a (finite) collection Ω of paths in a graph with all paths starting at a fixed vertex, and with the partial order that $\omega\preceq\omega'$ iff ω is an initial segment of ω'.

- Alternatively, let (E, V) be a graph extended into a continuum by assigning a length to each edge. Let $d(a, b)$ be the infimum of the lengths of paths[6] between the two vertices a, b in the graph. Then g is a geodesic metric on V. Trees are exactly *the graphs that give rise to 0-hyperbolic metrics* in the sense of Gromov (see for example [3]).

- There are many ways to enumerate the edges and nodes of a finite rooted tree. One way is to think of a family tree recording the male descendants of a single male individual (the root). Start with the root. At the root, if all children have been visited stop, at any other node, if all the children have been visited, move up to the parent. If there are children who have not been visited, then visit the oldest unvisited child. At each time n the enumeration either moves up an edge or down an edge - each edge is visited exactly twice. Let $h(n)$ denote the distance from the top of the family tree after n steps in this enumeration with the convention that $h(0) = 0$, then h is similar to the path of a random walk, moving up or down one unit at each step, except that it is positive and returns to zero exactly as many times as there are edges coming from the root. Hence $h(2|E|) = 0$.

  The function h completely describes the rooted tree. The function h directly yields the nearest neighbour metric on the tree. If h is a function such that $h(0) = 0$, it moves up or down one unit at each step, is positive and $h(2|E|) = 0$, then d defined by

  $$d(m, n) = h(m) + h(n) - 2 \inf_{u \in [m,n]} h(u),$$

  is a pseudometric on $[0, 2|V|]$. If we identify points in $[0, 2|V|]$ that are zero distance apart and join by edges the equivalence classes of points that are distance one apart, then one recovers an equivalent rooted tree.

  Put less pedantically, let the enumeration be a at step n and b at step m and define

  $$d(a, b) = h(m) + h(n) - 2 \inf_{u \in [m,n]} h(u),$$

  then it is simple to check that d is well defined and is a metric on vertices making the set of vertices a tree.

  Thus excursions of simple (random) walks are a convenient (and well studied) way to describe abstract graphical trees. This particular choice for *coding a tree with a positive function on the interval* can be extended to describe continuous trees. To the authors knowledge this was first done by Le Gall [4] in his development of the Brownian snake associated to the measure valued Dawson-Watanabe process.

A.2 R-trees are coded by continuous functions

One of the early examples of a continuous tree is the evolution of a continuous time stochastic process, where, as is customary in probability, one identifies the evolution of two trajectories until the first time they separate. (This idea dates back at least to Kolmogorov and his introduction of filtrations). Another popular and equivalent approach to continuous trees is through $\mathbf{R}$-trees ([6] p425 and the references there).

Interestingly, analysts and probabilists have generally rejected the abstract tree as too wild an object, and usually add extra structure, essentially a second topology or Borel structure on the tree that comes from thinking of the tree as a family of paths in a space which also has some topology. This approach is critical to the arguments used here, as we prove our tree-like paths

---

[6] the sum of the lengths of the edges

are tree-like by approximating them with simpler tree-like paths. (They would never converge in the 'hyperbolic' metric). In contrast, group theorists and low dimensional topologists have made a great deal of progress by studying specific symmetry groups of these trees and do not seem to find their hugeness too problematic.

Our goal in this subsection of the appendix is to prove the simple representation: that the general **R**-tree arises from identifying the contours of a continuous function on a locally connected and connected space. The height functions we considered on $[0,T]$ are a special case.

Definition A.1 *An* **R***-tree is a uniquely arcwise connected metric space, in which the arc between two points is isometric to an interval.*

Such a space is locally connected, for let B_x be the set of points a distance at most $1/n$ from x. If $z \in B_x$, then the arc connecting x with z is isometrically embedded, and hence is contained in B_x. Hence B_x is the union of connected sets with non-empty common intersection (they contain x) and is connected. The sets B_x form a basis for the topology induced by the metric. Observe that if two arcs meet at two points, then the uniqueness assertion ensures that they coincide on the interval in between.

Fix some point v as the 'root' and let x and y be two points in the **R**-tree. The arcs from x and y to v have a maximal interval in common starting at v and terminating at some v_1, after that time they never meet again. One arc between them is the join of the arcs from x to v_1 to y (and hence it is the arc and a geodesic between them). Hence

$$d(x,y) = d(x,v) + d(y,v) - 2d(v,v_1).$$

Example A.2 *Consider the space Ω of continuous paths $X_t \in E$ where each path is defined on an interval $[0,\xi(\omega))$ and has a left limit at $[0,\xi(\omega))$. Suppose that if $X \in \Omega$ is defined on $[0,\xi)$, then $X|_{[0,s)} \in \Omega$ for every s less than ξ. Define*

$$d(\omega,\omega') = \xi(\omega) + \xi(\omega') - 2\sup\{t < \min(\xi(\omega),\xi(\omega')) \mid \omega(s) = \omega'(s)\ \forall s \le t\}.$$

Then (Ω, d) is an R-tree.

We now give a way of constructing **R**-trees. The basic idea for this is quite easy, but the core of the argument lies in the detail so we proceed carefully in stages.

Let I be a connected and locally connected topological space, and $h : I \to \mathbb{R}$ be a positive continuous function that attains its lower bound at a point $v \in I$.

Definition A.3 *For each $x \in I$ and $\lambda \le h(x)$ define $C_{x,\lambda}$ to be the maximal connected subset of $\{y \mid h(y) \ge \lambda\}$ containing x.*

Lemma A.4 *The sets $C_{x,\lambda}$ exist, and are closed. Moreover, if $C_{x,\lambda} \cap C_{x',\lambda'} \ne \phi$ and $\lambda \le \lambda'$, then*

$$C_{x',\lambda'} \subset C_{x,\lambda}.$$

Proof. An arbitrary union of connected sets with non-empty intersection is connected, taking the union of all connected subsets of $\{y \mid h(y) \ge \lambda\}$ containing x constructs the unique maximal connected subset. Since h is continuous the closure $D_{x,\lambda}$ of $C_{x,\lambda}$ is also a subset of $\{y \mid h(y) \ge \lambda\}$. The closure of a connected set is always connected hence $D_{x,\lambda}$ is also connected. It follows from the fact that $C_{x,\lambda}$ is maximal that $C_{x,\lambda} = D_{x,\lambda}$ and so is a closed set.

If $C_{x,\lambda} \cap C_{x',\lambda'} \ne \phi$ and $\lambda \le \lambda'$, then

$$x \in C_{x,\lambda} \cup C_{x',\lambda'} \subset \{y \mid h(y) \ge \lambda\},$$

and since $C_{x,\lambda} \cap C_{x',\lambda'} \neq \phi$, the set $C_{x,\lambda} \cup C_{x',\lambda'}$ is connected. Hence maximality ensures $C_{x,\lambda} = C_{x,\lambda} \cup C_{x',\lambda'}$ and hence $C_{x',\lambda'} \subset C_{x,\lambda}$. ▮

Corollary A.5 *Either $C_{x,\lambda}$ equals $C_{x',\lambda}$ or it is disjoint from it.*

Proof. If they are not disjoint, then the previous Lemma can be applied twice to prove that $C_{x',\lambda} \subset C_{x,\lambda}$ and $C_{x,\lambda} \subset C_{x',\lambda}$. ▮

Corollary A.6 *If $C_{x,\lambda} = C_{x',\lambda}$, then $C_{x,\lambda''} = C_{x',\lambda''}$ for all $\lambda'' < \lambda$.*

Proof. The set $C_{x,\lambda}, C_{x',\lambda}$ are nonempty and have nontrivial intersection. $C_{x,\lambda} \subset C_{x,\lambda''}$ and $C_{x',\lambda} \subset C_{x',\lambda''}$ hence $C_{x,\lambda''}$ and $C_{x',\lambda''}$ have nontrivial intersection. Hence they are equal. ▮

Corollary A.7 *$y \in C_{x,\lambda}$ if and only if $C_{y,h(y)} \subset C_{x,\lambda}$.*

Proof. Suppose that $y \in C_{x,\lambda}$, then $C_{y,h(y)}$ and $C_{x,\lambda}$ are not disjoint. It follows from the definition of $C_{x,\lambda}$ and $y \in C_{x,\lambda}$ that $h(y) \geq \lambda$. By Lemma A.4 $C_{y,h(y)} \subset C_{x,\lambda}$. Suppose that $C_{y,h(y)} \subset C_{x,\lambda}$, since $y \in C_{y,h(y)}$ it is obvious that $y \in C_{x,\lambda}$. ▮

Definition A.8 *The set $C_x := C_{x,h(x)}$ is commonly referred to as the* contour *of h through x.*

The map $x \to C_x$ induces a partial order on I with $x \preceq y$ if $C_x \supseteq C_y$. If h attains its lower bound at x, then $C_x = I$ since $\{y \mid h(y) \geq h(x)\} = I$ and I is connected by hypothesis. Hence the root $v \preceq y$ for all $y \in I$.

Lemma A.9 *Suppose that $\lambda \in [h(v), h(x)]$, then there is a y in $C_{x,\lambda}$ such that $h(y) = \lambda$ and, in particular, there is always a contour ($C_{x,\lambda}$) at height λ through y that contains x.*

Proof. By the definition of $C_{x,\lambda}$ it is the maximal connected subset of $h \geq \lambda$ containing x; assume the hypothesis that there is no y in $C_{x,\lambda}$ with $h(y) = \lambda$ so that it is contained in $h > \lambda$, hence $\mathrm{C}_{x,\lambda}$ is a maximal connected subset of $h > \lambda$. Now $h > \lambda$ is open and locally connected, hence its maximal connected subsets of $h > \lambda$ are open and $C_{x,\lambda}$ is open. However it is also closed, which contradicts the connectedness of the I. Thus we have established the existence of the point y. ▮

The contour is obviously unique, although y is in general not. If we consider the equivalence classes $x\tilde{}y$ if $x \preceq y$ and $y \preceq x$, then we see that the equivalence classes $[y]_{\tilde{}}$ of $y \preceq x$ are totally ordered and in one to one correspondence with points in the interval $[h(v), h(x)]$.

Lemma A.10 *If $z \in C_{y,\lambda}$ and $h(z) > \lambda$, then z is in the interior of $C_{y,\lambda}$. If $C_{x',\lambda'} \subset C_{x,\lambda}$ with $\lambda' > \lambda$, then $C_{x,\lambda}$ is a neighbourhood of $C_{x',\lambda'}$.*

Proof. I is locally connected, and h is continuous, hence there is a connected neighbourhood U of z such that $h(z) \geq \lambda$. By maximality $U \subset C_{z,\lambda}$. Since $C_{z,\lambda} \cap C_{y,\lambda} \neq \phi$ we have $C_{z,\lambda} = C_{y,\lambda}$ and thus $U \subset C_{y,\lambda}$. Hence $C_{y,\lambda}$ is a neighbourhood of z. The last part follows trivially once by noting that for all $z \in C_{x',\lambda'}$ we have $h(z) \geq \lambda' > \lambda$ and hence $C_{y,\lambda}$ is a neighbourhood of z. ▮

We now define a pseudo-metric on I. Lemma A.10 (the only place we will use local connectedness) is critical to showing that the map from I to the resulting quotient space is continuous.

Definition A.11 *If* y *and* z *are points in* I*, define* $\lambda(y,z) \leq \min(h(y), h(z))$ *such that* $C_{y,\lambda} = C_{z,\lambda}$

$$\lambda(y,z) = \sup\{\lambda \mid C_{y,\lambda} = C_{z,\lambda},\ \lambda \leq h(y),\ \lambda \leq h(z)\}.$$

The set

$$\{\lambda \mid C_{y,\lambda} = C_{z,\lambda},\ \lambda \leq h(y),\ \lambda \leq h(z)\}$$

is a non-empty interval $[h(v), \lambda(y,z)]$ or $[h(v), \lambda(y,z))$ where $\lambda(y,z)$ satisfies

$$h(v) \leq \lambda(y,z) \leq \min(h(y), h(z)).$$

Clearly $\lambda(x,x) = h(x)$.

Lemma A.12 *The function* λ *is lower semicontinuous*

$$\liminf_{z \to z_0} \lambda(y,z) \geq \lambda(y,z_0).$$

Proof. Fix y, z_0 and choose some $\lambda' < \lambda(y,z_0)$. By the definition of $\lambda(y,z_0)$ we have that $C_{y,\lambda'} = C_{z_0,\lambda'}$. Since $h(z_0) \geq \lambda'$ there is a neighbourhood U of z_0 so that $U \subset C_{z_0,\lambda'}$. For any $z \in U$ one has $z \in C_{z,\lambda'} \cap C_{z_0,\lambda'}$. Hence $C_{z_0,\lambda'} = C_{z,\lambda'}$ and $C_{y,\lambda'} = C_{z,\lambda'}$. Thus $\lambda(y,z) \geq \lambda'$ for $z \in U$ and hence

$$\liminf_{z \to z_0} \lambda(y,z) \geq \lambda'.$$

Since $\lambda' < \lambda(y,z_0)$ was arbitrary

$$\liminf_{z \to z_0} \lambda(y,z) \geq \lambda(y,z_0)$$

and the result is proved. ▮

Lemma A.13 *The following inequality holds*

$$\min\{\lambda(x,z), \lambda(y,z)\} \leq \lambda(x,y).$$

Proof. If $\min\{\lambda(x,z), \lambda(y,z)\} = h(v)$, then there is nothing to prove. Recall that

$$\{\lambda \mid C_{y,\lambda} = C_{z,\lambda},\ \lambda \leq h(y),\ \lambda \leq h(x)\}$$

is connected and contains $h(v)$. Suppose $h(v) \leq \lambda < \min\{\lambda(x,z), \lambda(y,z)\}$, then it follows that the identity $C_{x,\lambda} = C_{z,\lambda}$ holds for λ. Similarly $C_{y,\lambda} = C_{z,\lambda}$. As a result $C_{x,\lambda} = C_{y,\lambda}$ and $\lambda(x,y) \geq \lambda$. ▮

Definition A.14 *Define* d *on* $I \times I$ *by*

$$d(x,y) = h(x) + h(y) - 2\lambda(x,y).$$

Lemma A.15 *The function* d *is a pseudometric on* I*. If* $\left(\tilde{I}, d\right)$ *is the resulting quotient metric space, then the projection* $I \to \tilde{I}$ *from the topological space* I *to the metric space is continuous.*

Proof. Clearly d is positive, symmetric and we have remarked that for all x, $\lambda(x,x) = h(x)$ hence it is zero on the diagonal. To see the triangle inequality, assume

$$\lambda(x,z) = \min\{\lambda(x,z), \lambda(y,z)\}$$

and then observe

$$\begin{aligned} d(x,y) &= h(x) + h(y) - 2\lambda(x,y) \\ &\leq h(x) + h(y) - 2\lambda(x,z) \\ &= h(x) + h(z) - 2\lambda(x,z) + h(y) - h(z) \\ &\leq d(x,z) + |h(y) - h(z)| \end{aligned}$$

but $\lambda(y,z) \leq \min(h(y), h(z))$ and hence

$$\begin{aligned} |h(y) - h(z)| &= h(y) + h(z) - 2\min(h(y), h(z)) \\ &\leq h(y) + h(z) - 2\lambda(y,z) \\ &= d(y,z) \end{aligned}$$

hence

$$d(x,y) \leq d(x,z) + d(y,z)$$

as required. ∎

We can now introduce the equivalence relation $x\tilde{}y$ if $d(x,y) = 0$ and the quotient space $I/\tilde{}$. We write $I/\tilde{} = \tilde{I}$ and $i : I \to \tilde{I}$ for the canonical projection. The function d projects onto $\tilde{I} \times \tilde{I}$ and is a metric there.

It is tempting to think that $x\tilde{}y$ if and only if $C_x = C_y$ and this is true if I is compact Hausdorff. However the definitions imply a slightly different criteria: $x\tilde{}y$ iff

$$h(x) = h(y) = \lambda \text{ and } C_{x,\lambda''} = C_{y,\lambda''} \text{ for all } \lambda'' < \lambda.$$

The stronger statement $x\tilde{}y$ if and only if $C_x = C_y$ is not true for all continuous functions h on $\mathbb{R}^2$ as it is easy to find a decreasing family of closed connected sets there whose limit is a closed set that is not connected.

Consider again the new metric space $\tilde{I}$ that has as its points the equivalence classes of points indistinguishable under d. We now prove that the projection i taking I to $\tilde{I}$ is continuous. Fix $y \in I$ and $\varepsilon > 0$. Since $\lambda(y,.)$ is lower semicontinuous and h is (upper semi)continuous there is a neighbourhood U of y so that for $z \in U$ one has $\lambda(y,z) > \lambda(y,y) - \varepsilon/4$ and $h(z) < h(y) + \varepsilon/2$. Thus $d(y,z) < \varepsilon$ for $z \in U$. Hence $\tilde{d}(i(y), i(z)) < \varepsilon$ if $z \in U$. The function i is continuous and as continuous images of compact sets are compact we have the following.

Corollary A.16 *If I is compact, then $\tilde{I}$ is a compact metric space.*

To complete this section we will show $\tilde{I}$ is a uniquely arcwise connected metric space, in which the arc between two points is isometric to an interval and give a characterisation of compact trees.

Theorem A.17 *If I is a connected and locally connected topological space, and $h : I \to \mathbb{R}$ is a positive continuous function that attains its lower bound, then its "contour tree" the metric space $\left(\tilde{I}, \tilde{d}\right)$ is an* **R***-tree. Every* **R** *-tree can be constructed in this way.*

Proof. It is enough to prove that the metric space $\tilde{I}$ we have constructed is really an **R**-tree and that every **R**-tree can be constructed in this way. Let $\tilde{x}$ any point in $\tilde{I}$ and $x \in I$ satisfy $i(x) = \tilde{x}$. Then $h(x)$ does not depend on the choice of x. Fix $h(v) < \lambda < h(x)$. We have seen that there is a y such that $h(y)=\lambda$ and $y \prec x$ moreover any two choices have the same contour through them and hence the same $\tilde{y}(\lambda)$. In this way we see that there is a map from $[h(v), h(x)]$ into $\tilde{I}$ that is injective. Moreover, it is immediate from the definition of d that it is an isometry and that $\tilde{I}$ is uniquely arc connected.

Suppose that Ω is an **R**-tree, then we may fix a base point, and for each point in the tree consider the distance from V it is clear that this continuous function is just appropriate to ensure that the contour tree is the original tree. ∎

Remark A.18 1. In the case where I is compact, obviously $\tilde{I}$ is both complete and totally bounded as it is compact.

2. An R-tree is a metric space; it is therefore possible to complete it. Indeed the completion consists of those paths, all of whose initial segments are in the tree[7]; we have not identified a simple sufficient condition on the continuous function and topological space Ω to ensure this. An R-tree is totally bounded if it is bounded and for each $\varepsilon > 0$ there is an N so that for each t the paths that extend a distance t from the root have at most N ancestral paths between them at time $t-\varepsilon$. In this way we see that the R-tree that comes out of studying the historical process for the Fleming-Viot or the Dawson Watanabe measure-valued processes is, with probability one, a compact R-tree for each finite time.

Lemma A.19 *Given a compact R-tree, there is always a height function on a closed interval that yields the same tree as its quotient.*

Proof. As the tree is compact, path connected and locally path connected, there is always as based loop mapping $[0, 1]$ onto the tree. Let h denote the distance from the root. Its pullback onto the interval $[0, 1]$ is a height function and the natural quotient is the original tree. In this way we see that there is always a version of Le Gall's snake [4] traversing a compact tree. ∎

References

[1] Chen, Kuo-sai T. *Integration of paths—a faithful representation of paths by non-commutative formal power series.* Trans. Amer. Math. Soc. 89 (1958), 395–407

[2] Fawcett, T. A. *DPhil Thesis, Mathematical Institute, University of Oxford 2002.*

[3] Kapovich, I. *A Non-quasiconvex Subgroup of a Hyperbolic Group with an Exotic Limit Set*, New York J. Math. 1 (1995), 184-195.

[4] Le Gall, J.F. *Brownian excursions, trees and measure-valued branching processes* Ann. Probab. 19 (1991), 1399–1439

[5] Lyons, T.J. and Qian, Z. *Rough paths and system control*, Oxford University Press, 2003.

[6] Morgan, John W., Shalen, Peter B., *Valuations, trees, and degenerations of hyperbolic structures. I.* Ann. of Math. (2) 120 (1984), 401–476.

---

[7] We fix a root and identify the tree with the geodesic arc from the root to the point in the tree.

[7] Ohtsuka, M. *Dirichlet problems on Riemann surfaces and conformal mappings.* Nagoya Math J., 3 (1951), 91–137.