IMPERIAL

Imperial College London

Department of Mathematics

An Invertible Lie Group Embedding for Piecewise Linear Paths

Author: Rhys KENNY(CID: 02481413)

A thesis submitted for the degree of

MSc in Mathematics and Finance, 2023-2024

Declaration

The work contained in this thesis is my own work unless otherwise stated.

Acknowledgements

I would like firstly like to express my gratitude to my supervisor Dr Cristopher Salvi for providing me with direction, perspective and for introducing me to such a fascinating branch of mathematics. I would also like to thank Barbora Barancikova for her contribution to my discussions with Dr Salvi and for aiding in my understanding of these topics.

To my family, whatever I can write here will be entirely inadequate to express how lucky I have been to have your support throughout my years at school and university. I have taken somewhat of a scenic route through third level education, but for now at least, we are done.

Abstract

The Cartan development of a path onto a Lie group has emerged as a useful tool in rough path theory, containing considerable information about the underlying path. We examine its potential for use in synthesising time-series data, specifically a method for reclaiming a d-dimensional path from its development onto the Lie group $SO(d, 1)$. We provide numerical implementation of existing theory and propose a refinement of existing methods to aid practical viability for certain important classes of paths.

Contents

List of Figures

List of Tables

Introduction

The notion of a path evolving in continuous time is pivotal in mathematical finance, be it the study of long term price dynamics on a macro scale, or the piecewise constant paths seen in high-frequency tick data. In particular, we frequently find ourselves dealing with financial time-series data, manifesting as a linear interpolation of discretely sampled data from an underlying continuous process. It is natural then that statistical models related to time series data, as both a model input and output, are a prominent area of research in the field.

An issue which often appears in the context of financial modelling concerns the availability of input data, when backtesting trading strategies for instance. Often for a strategy to be profitable, it needs to be inherently optimised to the current market regime, or with respect to other constraints which require that the dynamics of our training data closely match those of the realised data. In other applications such as image recognition, we can often increase the size of our training sample by including various transformations of the existing data, such as adding noise or certain isometries. For financial data however, these transformations alter the inherent structure that we want our models to infer. With that in mind, the ability to generate synthetic time-series data which are statistically indistinguishable from the realised data is an incredibly useful one. Models such as *Generative Adversarial Networks* (GANs) [5] have been used for this purpose, although can struggle with dimensionality when dealing with particularly long or high-frequency data [11, Section 1]. It has recently been proposed in [1] to consider the problem in continuous time, by instead generating samples of the *Signature* (or more precisely the log-signature) of a continuous path using *Score-Based Diffusion Models* [14]. Summarised in Section 1.3, the signature of a path is a formal power series of its iterated integrals, which has been shown to uniquely characterise certain classes of paths [7]. The characteristic properties of the signature mean that if we can train a generative model on the signatures of existing paths, producing synthetic samples in the signature space, we can then invert the signature to synthesise continuous paths with similar characteristics to the training data. We note that signature inversion remains an area of research although various methods have been proposed such as in [1, 12].

The signature is not without issues however, particularly related to dimensionality which we discuss further below. As a result, an object strongly related to the signature known as the *Cartan development* has been proposed as an alternative for certain applications, for example [10] where it is shown to be a useful feature representation in deep learning-based models where time-series data is an input. The Cartan development can also be seen as a series of iterated integrals, though has the advantage of constant dimension across its terms. We show below that many of the desirable properties enjoyed by the signature are also present in the Cartan development, making it an attractive option for generative modelling in the context of continuous paths. We highlight in particular a version of the Cartan development known as the *hyperbolic development*, used in [7] to prove the uniqueness of the signature and more recently in [12] as part of a procedure for signature inversion. This particular development allows us to use certain geometric properties of hyperbolic space in order to extract information about the underlying path.

The structure of this paper is as follows. In Chapter 1 we introduce the necessary mathematical background for the study of the signature and Cartan development. Chapter 2 formally introduces the Cartan development, going through some useful properties as well as detailing the specific case of the hyperbolic development. In Chapter 3, we examine why the hyperbolic development is particularly useful in the context of inversion, and propose an alternative step in the procedure of [12], which leverages geometric properties of the development to dramatically improve convergence in certain cases, making the procedure much more feasible in practice. Finally in Chapter 4 we provide a practical implementation of the theory in [12] to reclaim a path from both its hyperbolic development and its signature.

Chapter 1

Preliminaries

1.1 Lie Groups and Lie Algebras

Definition 1.1.1 (Lie Group)**.** *A group G is a Lie Group if it is also a differentiable manifold, such that group multiplication and inversion are differentiable.*

An important concept in the study of Lie groups is the following:

Definition 1.1.2 (Lie Algebra)**.** *A Lie Algebra over a field $\mathbb{F}$ is an $\mathbb{F}$-vector space $\mathfrak{g}$ together with a binary operation $[\cdot,\cdot] : \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}$ called the Lie bracket, satisfying:*

1. $[\cdot,\cdot]$ *is bilinear*
2. $[X, X] = 0 \; \forall \; X \in \mathfrak{g}$
3. $[X, [Y, Z]] + [Y, [Z, X]] + [Z, [X, Y]] = 0 \; \forall \; X, Y, Z \in \mathfrak{g}$

Ado's Theorem [15, Section 3.17] states that if a field $\mathbb{F}$ has characteristic 0 and $\mathfrak{g}$ is a finite dimensional Lie algebra over $\mathbb{F}$, then $\mathfrak{g}$ is isomorphic to a matrix Lie algebra. For this reason, we refer only to matrix Lie algebra in what follows, and for simplicity will also consider only matrix Lie groups G (although no statement equivalent to that of Ado's Theorem exists for Lie groups).

With every Lie group, we can associate a corresponding Lie algebra. In fact, in the case of matrix Lie groups we can say the following:

Definition 1.1.3. *Let G be a matrix Lie group. The Lie algebra $\mathfrak{g}$ of G is*

$$\mathfrak{g} := \{X \in M_n(\mathbb{C}) : e^{tX} \in G \; \forall \; t \in \mathbb{R}\},$$

where e^X is the matrix exponential. The Lie bracket of $\mathfrak{g}$ is given by

$$[X, Y] = XY - YX,$$

and is known as the commutator.

Geometrically, $\mathfrak{g}$ is the tangent space to G at the identity [6, Corollary 3.46, page 71], and the exponential maps vectors in this tangent space to corresponding points on the differentiable manifold. Here, the matrix exponential restricted to $\mathfrak{g}$ is known as the *Lie group exponential* of G (we can also define an exponential mapping for general Lie groups), though we note that this mapping is neither injective nor surjective in general.

1.2 Paths of Bounded Variation

Definition 1.2.1. *(p-variation) Let V be a finite dimensional Banach space and $p \geq 1$ a real number. A path $X : [a, b] \to V$ had finite p-variation if*

$$\|X\|_p := \left(\sup_{\mathcal{D} \subset [a,b]} \sum_{t_i \in \mathcal{D}} \|X_{t_{i+1}} - X_{t_i}\|^p \right)^{1/p} < \infty,$$

where $\|\cdot\|$ is any norm on V and $\mathcal{D}$ is a partition of the interval $[a, b]$. We denote the set of continuous paths from $[a, b]$ to V with finite p-variation by $C_p([a, b], V)$. We can also refer to finite p-variation on some sub-interval $[s, t] \subset [a, b]$, and will sometimes write $X_{s,t}$ to denote the path X restricted to the interval $[s, t]$.

We focus in particular on paths $X \in C_1([a, b], V)$. We say such paths have *bounded variation*, and refer to $\|X\|_1$ as its length.

We now introduce an important binary operation on paths known as concatenation. Namely, if X and Y are paths in $C_p([a, b], V)$ and $C_p([b, c], V)$ respectively, their concatenation is the path $X * Y \in C_p([a, c], V)$ given by

$$(X * Y)_t = \begin{cases} X_t & \text{if } t \in [a, b] \\ Y_t - Y_b + X_b & \text{if } t \in [b, c] \end{cases}$$

Definition 1.2.2 (Piecewise Linear Paths)**.** *A path $X \in C_1([a, b], \mathbb{R}^d)$ is piecewise linear if it is a concatenation of linear paths in $C_1([a, b], \mathbb{R}^d)$. That is, we can write*

$$X_t = (x_1 * x_2 * \cdots * x_n)_t$$

where each x_i is a linear path from $[t_{i-1}, t_i]$ to $\mathbb{R}^d$ and $a = t_0 < t_1 < \cdots < t_n = b$.

Example 1.2.1 (Axis Path)**.** *An* axis path *is a piecewise linear path where each linear segment moves parallel to the Euclidean axes. Formally, if X is an axis path in $C_1([a, b], \mathbb{R}^d)$, we can write*

$$X_t = (\alpha_1 e_{i_1} * \cdots * \alpha_n e_{i_n})_t$$

where e_{i_k} a path of length 1 in the direction of the i_k^{th} standard Euclidean basis vector and the coefficients $\{\alpha_k\}$ are finite. Without loss of generality we assume that $e_{i_k} \neq e_{i_{k+1}}$ for any k, as otherwise we can combine the segments:

$$\alpha_k e_{i_k} * \alpha_{k+1} e_{i_{k+1}} = (\alpha_k + \alpha_{k+1}) e_{i_k}.$$

We also note that in this formulation, each e_{i_k} is positive and so direction is given by the sign of α_k.

Definition 1.2.3. *The unit-speed parameterisation of a path X of length L is a reparameterisation onto $C_1([0, L], \mathbb{R}^d)$ such that it has constant speed $\|\dot{X}\| = 1$ for all $t \in [0, L]$. We will often use this parameterisation so that $\dot{X}$ is on the unit sphere $\mathbb{S}^{d-1}$, the reasons for which will become apparent in later sections.*

Definition 1.2.4. *For a path $X \in C_1([a, b], \mathbb{R}^d)$, we can define the time reversal of X as the path*

$$\overleftarrow{X}_t := X_{a+b-t},$$

namely the path which retraces X starting at X_b and ending at X_a [2, Section 1.3.2].

1.3 The Signature Transform

The *Signature transform* is a mathematical object which represents a path X by a series of its iterated integrals. The signature contains considerable geometric and analytic information about a path and has been shown to be a one-to-one map for certain classes of paths, namely two paths of bounded variation have the same signature if and only if they are *tree-like equivalent*.

Remark 1.3.1. *The above notion of tree-like equivalence is owed a more in depth explanation. The term 'tree-like path' was introduced in [7, Definition 1.2] and is more rigorous than what is required here. It suffices to see that in the case of bounded variation paths, a path being tree-like amounts to it retracing all of its segments exactly, in a sense cancelling out its effects. It was proved in [7] that a path has signature equal to the identity if and only if it is tree-like. Tree-like equivalence builds on this concept, defining an equivalence class of paths where $X \sim Y$ when the path X concatenated with the time reversal of Y is tree-like. This equivalence occurs if and only if the signature of $X * \overleftarrow{Y}$ is the identity and moreover that $X \sim Y$ if and only if they have identical signatures.*

Example 1.3.1. *The easiest example of the above is that $X * \overleftarrow{X}$ is tree-like for any bounded variation path X. Likewise, when Y is also a bounded variation path, we have that $(X * Y * \overleftarrow{Y}) \sim X$.*

As a result of the above, the signature has proved to be a useful feature in machine learning where paths are concerned (for example [13]). It has numerous other desirable properties, though we refer the reader to [2] for a more thorough introduction than what is required here.

In order to formalise concept of the signature transform, we begin with some necessary mathematical notions.

Definition 1.3.2 (Tensor Product)**.** *For vector spaces S and T, the tensor product $S \otimes T$ is a vector space with an associated bilinear map*

$$\begin{aligned} S \times T &\to S \otimes T \\ (s,t) &\mapsto s \otimes t. \end{aligned}$$

If the bases of S and T are $\mathcal{B}(S)$ and $\mathcal{B}(T)$ respectively, then the basis of $S \otimes T$ is the set

$$\left\{ s_i \otimes t_j \middle| \; s_i \in \mathcal{B}(S), \; t_j \in \mathcal{B}(T) \right\}$$

We denote by $S^{\otimes k}$ the tensor product $\overbrace{S \otimes \cdots \otimes S}^{k \text{ times}}$ which likewise has basis elements of the form $s_{i_1} \otimes \cdots \otimes s_{i_k}$.

Definition 1.3.3 (Extended Tensor Algebra [2, Definition 1.1.9])**.** *For a vector space V, elements $v = (v_0, v_1, \dots)$ of the product space $\prod_{k=0}^{\infty} V^{\otimes k}$ can be interpreted as the formal tensor series $\sum_{k=0}^{\infty} v_k$. This space forms an algebra with product $v \cdot w$ given by*

$$\begin{aligned} v \cdot w &= (z_0, z_1, \dots) \\ z_k &= \sum_{i=0}^{k} v_i w_{k-i} \in V^{\otimes k}. \end{aligned}$$

This is called the extended tensor algebra of V and is denoted $T((V))$. The identity of $T((V))$ is given by the tensor $(1, \mathbf{0}, \mathbf{0}, \dots)$.

Definition 1.3.4 (Signature Transform [2, Definition 1.1.6]). *Let V be a Banach space and X a path in $C_p([a,b],V)$ for $p \in [1,2)$. The signature transform of X on the interval $[s,t] \subseteq [a,b]$ is given by the series*

$$S(X)_{[s,t]} := \left(1,\, S(X)^{(1)}_{[s,t]},\, S(X)^{(2)}_{[s,t]}, \ldots,\, S(X)^{(k)}_{[s,t]}, \ldots\right) \in T((V))$$

where

$$S(X)^{(k)}_{[s,t]} = \int_{s<t_1<\ldots<t_k<t} dX_{t_1} \otimes dX_{t_2} \otimes \cdots \otimes dX_{t_k} \;\in\; V^{\otimes k}.$$

We generally drop the subscript from this notation when the interval is equal to the full domain of X.

Remark 1.3.5. *Let X be a path taking values in $\mathbb{R}^d$. $S(X)^{(k)}$ is a tensor in $(\mathbb{R}^d)^{\otimes k}$ and it will often be useful to refer to its individual coefficients. Each of these coefficients is a real number which we can write as $S(X)^{i_1,i_2,\ldots,i_k}$ where $i_j \in \{1,\ldots,d\}$. Writing the coordinates of X_t as $(x_t^{(1)},\ldots,x_t^{(d)})$, we see that*

$$S(X)^{i_1,i_2,\ldots,i_k} = \int_{a<t_1<\ldots<t_k<b} dx_{t_1}^{(i_1)} dx_{t_2}^{(i_2)} \cdots dx_{t_k}^{(i_k)} \in \mathbb{R}.$$

This is known as a coordinate iterated integral.

Two crucial lemmas in the computation of the signature in practice are the following.

Lemma 1.3.6 (Signature of a Linear Path). *If $X \in C_1([a,b],V)$ is linear, then its signature is given by*

$$S(X) = \exp\left(i_1(X_b - X_a)\right)$$

where $i_1 : V \to T((V))$ is the canonical inclusion and exp *is the tensor exponential given by*

$$\exp(v) := \sum_{k=0}^{\infty} \frac{v^{\otimes k}}{k!}.$$

Lemma 1.3.7 (Chen's Relation). *The signature of the concatenated path $X * Y$ is given by*

$$S(X * Y) = S(X) \cdot S(Y)$$

where $\cdot$ is the product given in 1.3.3.

These lemmas allow us to easily compute the signature of a piecewise linear path, a class of paths which can be used to approximate bounded variation paths in practice.

Chapter 2

Path Development

2.1 Cartan Path Development

A particular drawback of the signature transform is that the dimension of $S(X)^{(k)}$ increases exponentially with respect to k. This makes it necessary to truncate the series to relatively few terms in order to compute it in practice. While the factorial decay in the magnitude of its coefficients [2, Proposition 1.2.3] means that truncating the signature retains a considerable amount of information, it remains a computationally intensive task. This is a primary reason for why the following has emerged as an alternative for certain applications.

Definition 2.1.1 (Cartan Development)**.** *Let G be a matrix Lie group, $\mathfrak{g}$ its Lie algebra and $M : \mathbb{R}^d \to \mathfrak{g}$ a linear map. For a Euclidean path $X \in C_1([a,b],\mathbb{R}^d)$, the* Cartan Development *of X onto G under M is the solution Z_t of*

$$\mathrm{d}Z_t = Z_t \cdot M(\mathrm{d}X_t), \quad Z_a = I_n. \tag{2.1.1}$$

or equivalently

$$Z_t = I + \sum_{k=1}^{\infty} Z_t^{(k)} \tag{2.1.2}$$

where I is the identity element of $\mathfrak{g}$ and

$$Z_t^{(k)} = \int_{a<t_1<\ldots<t_k<t} M(dX_{t_1})M(dX_{t_2})\cdots M(dX_{t_k}). \tag{2.1.3}$$

The solution to (2.1.1) is a path $Z : [a,b] \to G$ and we focus particularly on its endpoint Z_b, which we denote by $D_M(X)$. We note the similarity of the expression in (2.1.2) with the definition of the signature transform of X. In order to formalise this relationship, we can say the following:

Proposition 2.1.2. *(Link with the Signature [10, Lemma 2.6]) Let G be a Lie group and $M : \mathbb{R}^d \to \mathfrak{g}$ a linear map onto its Lie algebra. For each k we can define the canonical extension of M as the linear map $\widetilde{M}_k : (\mathbb{R}^d)^{\otimes k} \to \mathfrak{g}$ given by*

$$\widetilde{M}_k(e_{i_1} \otimes \cdots \otimes e_{i_k}) = M(e_{i_1}) \cdots M(e_{i_k}). \tag{2.1.4}$$

where the e_{i_j} are basis elements of $\mathbb{R}^d$. This map then extends linearly to general tensors in $(\mathbb{R}^d)^{\otimes k}$. If X is a path in $C_1([a,b],\mathbb{R}^d)$, we have for every k that

$$\begin{aligned} D_M^{(k)}(X) &= \widetilde{M}_k\big(S^{(k)}(X)\big) \\ \Longrightarrow D_M(X) &= \sum_{k=0}^{\infty} \widetilde{M}_k\big(S^{(k)}(X)\big) =: \widetilde{M}\big(S(X)\big) \end{aligned}$$

Proof. This follows quickly from the linearity of M:

$$\begin{aligned} D_M^{(k)}(X) &= \int_{a<t_1<\ldots<t_k<b} M(dX_{t_1})\cdots M(dX_{t_k}) \\ &= \int_{a<t_1<\ldots<t_k<b} \widetilde{M}_k(dX_{t_1}\otimes\cdots\otimes dX_{t_k}) \\ &= \widetilde{M}_k\big(S^{(k)}(X)\big) \end{aligned}$$

□

In contrast to the signature, the terms of the infinite series defining the Cartan development are matrices of constant dimension, leading to a substantial dimensionality reduction when both are truncated at the k^{th} term. The Cartan development also retains a number of desirable properties possessed by the signature, such as invariance under time reparameterisation [10, Lemma 2.7], as well as the following two lemmas:

Lemma 2.1.3 (Concatenation)**.** *Consider paths $X \in C_1([a,b],\mathbb{R}^d)$ and $Y \in C_1([b,c],\mathbb{R}^d)$. Then*

$$D_M(X*Y) = D_M(X)\cdot D_M(Y)$$

for all linear maps $M:\mathbb{R}^d\to\mathfrak{g}$, where $\cdot$ is the group multiplication defined in G. [10, Lemma 2.4]

Proof. We can proceed by adapting the proof of Chen's relation for the signature given in [2, Lemma 1.3.1]. Let $W = X*Y \in C_1([a,c],\mathbb{R}^d)$. We know that

$$D_M(W) = I + \sum_{k=1}^{\infty} D_M^{(k)}(W)$$

where

$$D_M^{(k)}(W) = \int_{a<t_1<\ldots<t_k<c} M(dW_{t_1})\cdots M(dW_{t_k}).$$

For each k, we can write the corresponding integral as

$$\begin{aligned} D_M^{(k)}(W) &= \sum_{i=1}^{k}\int_{a<t_1<\ldots<t_i<b<t_{i+1}<\ldots<t_k<c} M(dW_{t_1})\cdots M(dW_{t_k}) \\ &= \sum_{i=1}^{k}\int_{a<t_1<\ldots<t_i<b<t_{i+1}<\ldots<t_k<c} M(dX_{t_1})\cdots M(dX_{t_i})M(dY_{t_{i+1}})\cdots M(dY_{t_k}) \\ &= \sum_{i=1}^{k}\left(\int_{a<t_1<\ldots<t_i<b} M(dX_{t_1})\cdots M(dX_{t_i})\right)\left(\int_{b<t_{i+1}<\ldots<t_k<c} M(dY_{t_{i+1}})\cdots M(dY_{t_k})\right) \\ &= \sum_{i=1}^{k} D_M^{(i)}(X)\cdot D_M^{(k-i)}(Y) \end{aligned}$$

where the third line uses Fubini's Theorem. It now follows that

$$D_M(X*Y) = \sum_{k=0}^{\infty}\sum_{i=0}^{k} D_M^{(i)}(X)\cdot D_M^{(k-i)}(Y) = D_M(X)\cdot D_M(Y). \tag{2.1.5}$$

□

Lemma 2.1.4 (Development of a Linear Path)**.** *When $X \in C_1([a,b],\mathbb{R})$ is a linear path and G is a matrix Lie group, we have that*

$$D_M(X) = \exp\big(M(X_b - X_a)\big).$$

Proof. We can again adapt the equivalent proof for the signature in [2, Section 1.3.1]. Since X is linear, $dX_t = \frac{X_b - X_a}{b-a}dt$ for all $t \in [a,b]$, so for each k we have:

$$\begin{aligned}
D_M^{(k)}(X) &= \int_{a<t_1<\ldots<t_k<b} M\left(\frac{X_b - X_a}{b-a}dt_1\right)\cdots M\left(\frac{X_b - X_a}{b-a}dt_k\right) \\
&= M\left(\frac{X_b - X_a}{b-a}\right)^k \int_{a<t_1<\ldots<t_k<b} dt_1 dt_2 \cdots dt_k \qquad \text{(linearity of } M\text{)} \\
&= M\left(\frac{X_b - X_a}{b-a}\right)^k \left(\frac{(b-a)^k}{k!}\right)
\end{aligned}$$

Finally,

$$D_M(X) = I + \sum_{k=1}^{\infty} \frac{M(X_b - X_a)^k}{k!} = \exp\big(M(X_b - X_a)\big).$$

□

As with the signature, lemmas 2.1.3 and 2.1.4 provide a straightforward method for calculating the development of piecewise linear paths. In particular, if $X : [0,T] \to \mathbb{R}^d$ is piecewise linear such that $X_t = (x_1 * x_2 * \cdots * x_n)_t$ for linear segments $x_i \in C_1([t_{i-1}, t_i], \mathbb{R}^d)$, it is clear that we can write

$$\begin{aligned}
D_M(X) &= D_M(x_1) \cdot D_M(x_2) \cdots D_M(x_n) \\
&= \exp\Big(M\big((x_1)_{t_1} - (x_1)_{t_0}\big)\Big) \cdots \exp\Big(M\big((x_n)_{t_n} - (x_n)_{t_{n-1}}\big)\Big).
\end{aligned} \qquad (2.1.6)$$

Example 2.1.1. *As before, a path of the form $X * \overleftarrow{X}$ is tree-like, and we can easily see from Proposition 2.1.2 and Remark 1.3.1 that*

$$\begin{aligned}
D_M(X * \overleftarrow{X}) &= \widetilde{M}\big(S(X * \overleftarrow{X})\big) \\
&= \widetilde{M}(\mathbf{1}) \\
&= I
\end{aligned}$$

where $\mathbf{1}$ is the identity of $T((\mathbb{R}^d))$. Furthermore, any piecewise linear path which can be written as the concatenation $(X * Y * \overleftarrow{Y} * Z)$ for paths X, Y, Z has development equal to $D_M(X * Z)$. As a result, we consider only what is known as the 'tree-reduced' [7] path for any path X, namely the unique shortest path that is tree-like equivalent to X.

In most data science and financial applications, we deal with discrete time series data that can be either linearly interpolated or piecewise constant (high-frequency tick data, for instance). Both of these cases can be treated as piecewise linear paths, with a piecewise constant time series being represented as an axis path. We can therefore use the method described above to calculate the Cartan developments of discrete time series in these applications.

2.1.1 Characteristic Properties of the Cartan Development

Lemma 2.1.5 (First term of the development)**.** *Let X be a path in $C_1([a,b],\mathbb{R}^d)$, G a Lie group and $M:\mathbb{R}^d\to\mathfrak{g}$ an injective linear map onto the Lie algebra of G. The first term $D_M^{(1)}(X)$ of the development $D_M(X)$ is the zero matrix if and only if X is a loop, i.e. $X_b = X_a$.*

Proof.

$$\begin{aligned} D_M^{(1)}(X) &= \int_{a<t_1<b} M(dX_{t_1}) \\ &= M(X_b - X_a) \\ &= \mathbf{0} \iff X_b = X_a \end{aligned}$$

by the injectivity and linearity of M. □

We note that although M need not be injective in general for the Cartan development, the map used in below sections does have this property and injective maps onto commonly used Lie algebras are easily defined.

Remark 2.1.6. *With regard to Lemma 2.1.5 and* (2.1.4)*, we can consider the first term of the Cartan development of X as a linear function of displacement $X_b - X_a$ between its start and endpoints and note that the first term of the signature is indeed this displacement.*

Remark 2.1.7. *Using the same logic as Lemma 2.1.5 we see that if M is injective and $X, Y \in C_1([a,b],\mathbb{R}^d)$, we have that $D_M^{(1)}(X) = D_M^{(1)}(Y)$ if and only if X and Y have identical displacement. That is, up to translation, X and Y have the same start and end points.*

Example 2.1.2. *Consider the unitary group $U(d)$, a commonly used Lie group for this application (see for example [9], [4]). Its Lie algebra is $\mathfrak{su}(d)$, the space of $d\times d$ matrices A with $A + A^* = 0$. Define the injective, linear map $M:\mathbb{R}^d \to \mathfrak{su}(d+1)$ as*

$$M(x) = \begin{pmatrix} 0 & x \\ -x^\top & 0 \end{pmatrix}.$$

The terms $D_M^{(k)}(X)$ of a path $X:[a,b]\to\mathbb{R}^2$ onto $U(3)$ can be given by

$$D_M^{(k)}(X) = \int_{a<t_1<\ldots<t_k<b} M(dX_{t_1})\cdots M(dX_{t_k}).$$

In particular, writing $X_t = (x_t^{(1)}, x_t^{(2)})$ we see that

$$\begin{aligned} D_M^{(1)}(X) &= \begin{pmatrix} 0 & 0 & \int_{a<t<b} dx_t^{(1)} \\ 0 & 0 & \int_{a<t<b} dx_t^{(2)} \\ -\int_{a<t<b} dx_t^{(1)} & -\int_{a<t<b} dx_t^{(2)} & 0 \end{pmatrix} \\ &= \begin{pmatrix} 0 & 0 & S(X)^1 \\ 0 & 0 & S(X)^2 \\ -S(X)^1 & -S(X)^2 & 0 \end{pmatrix} \end{aligned}$$

as in 2.1.5, an injective function of the depth-one signature coefficients. Here $S(X)^1$ represents displacement in the $x^{(1)}$ direction and $S(X)^2$ the $x^{(2)}$ direction. Furthermore, the second term is

$$D_M^{(2)}(X) = \begin{pmatrix} -S(X)^{1,1} & -S(X)^{1,2} & 0 \\ -S(X)^{2,1} & -S(X)^{2,2} & 0 \\ 0 & 0 & -S(X)^{1,1} - S(X)^{2,2} \end{pmatrix}$$

where

$$S(X)^{i,j} = \int_{a<t_1,t_2<b} dx_t^{(i)} dx_t^{(j)}.$$

Interestingly, we note that $\frac{1}{2}(S(X)^{1,2} - S(X)^{2,1})$ *defines the signed Lévy area of* X *[3, Equation 1.35]. This is the net area enclosed between* X *and the chord connecting* X_a *with* X_b*, for example the sum of signed areas* A^+ *and* A^- *in Figure 2.1.*

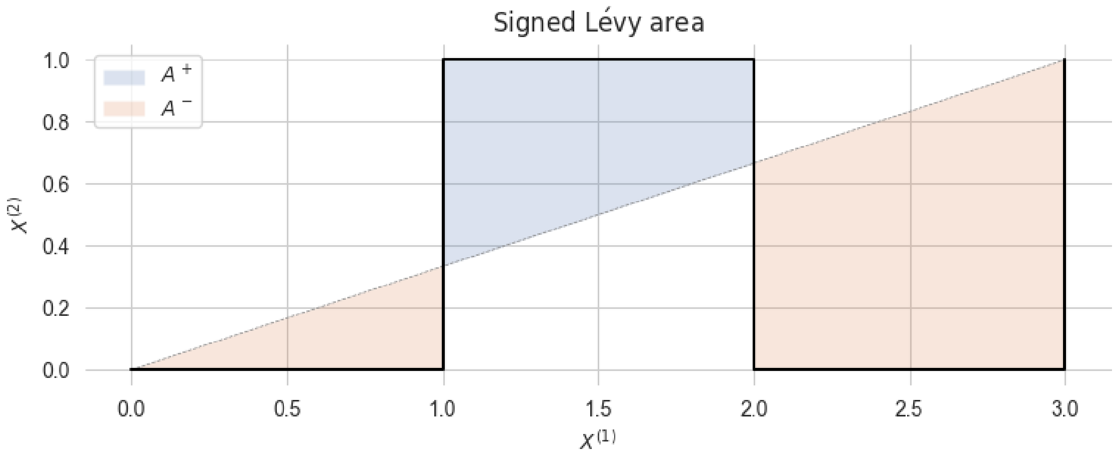

Figure 2.1: The Signed Lévy area for an axis Path

For a path X *in* $\mathbb{R}^d$ *we get that the* ij *element of the* $(d+1)\times(d+1)$ *matrix* $D_M^{(2)}(X)$ *is given by*

$$D_M^{(2)}(X)_{ij} = \begin{cases} -S(X)^{i,j} & \text{if} \quad i,j \le d \\ 0 & \text{otherwise.} \end{cases}$$

In this case, we can interpret $\frac{1}{2}(S(X)^{i,j} - S(X)^{j,i})$ *as the signed Lévy area of the path* X *projected orthogonally onto the plane defined by the vectors* e_i *and* e_j*, that is the path defined by* $(x_t^{(i)}, x_t^{(j)})$*. We then see that if two paths* $X, Y \in \mathbb{R}^d$ *have* $D_M^{(2)}(X) = D_M^{(2)}(Y)$*, then the signed Lévy areas of the paths* $(x_t^{(i)}, x_t^{(j)})$ *and* $(y_t^{(i)}, y_t^{(j)})$ *must be equal for all* $i \neq j \in \{1, \ldots, d\}$*. Note that this condition is necessary but not sufficient.*

With regards to the formulation of M above, we note that for many Lie algebras, defining a map which gives comparable information on the signature coefficients is not generally difficult. Indeed, the map used in the following sections is very similar.

2.1.2 Hyperbolic Development of a Path

We now detail the tools necessary to formulate an inveritible Cartan development for piecewise linear paths. This result hinges on the hyperbolic development of a Euclidean path detailed in Section 3.1 of [12]:

Definition 2.1.8 (Hyperbolic Space)**.** *Consider the following quadratic form on* $\mathbb{R}^{d+1}$*:*

$$Q_d(x,y) := \sum_{i=1}^{d} x_i y_i - x_{d+1} y_{d+1}.$$

The Hyperboloid Model for the d-dimensional hyperbolic space $\mathbb{H}^d$ *is given by*

$$\mathbb{H}^d := \{x \in \mathbb{R}^{d+1} : Q_d(x,x) = -1,\, x_{d+1} > 0\}.$$

This defines the upper sheet of the d-dimensional hyperboloid, a Riemann manifold with constant negative curvature, and metric given by the restriction of Q_d to the tangent spaces of $\mathbb{H}^d$.

We denote by $SO(d,1) \subset O(d,1)$ the Lie group of orientation-preserving isometries on $\mathbb{H}^d$. Here, $O(d,1) \subset GL_{d+1}(\mathbb{R})$ is the set of matrices which preserve the quadratic form Q_d, and $SO(d,1)$ is the subset with determinant 1. The Lie algebra of $SO(d,1)$ is written as $\mathfrak{so}(d,1)$ and contains $(d+1)\times(d+1)$ matrices of the form

$$\begin{pmatrix} A & x \\ x^\top & 0 \end{pmatrix}, \tag{2.1.7}$$

where A is a $d\times d$ skew-symmetric matrix and $x \in \mathbb{R}^d$

Definition 2.1.9 (Hyperbolic Development)**.** *Let $X : [0,T] \to \mathbb{R}^d$ be a Euclidean path of bounded variation and define the linear map $M : \mathbb{R}^d \to \mathfrak{so}(d,1)$ by*

$$M(x) = \begin{pmatrix} 0 & x \\ x^\top & 0 \end{pmatrix}.$$

Let Z_t be the Cartan development of X onto $SO(d,1)$ under M, which has endpoint $D_M(X)$. We refer to $D_M(X)\cdot o$ as the hyperbolic development of X, where o is the point $(0,0,\ldots,0,1)^\top \in \mathbb{R}^{d+1}$, the vertex of the hyperboloid $\mathbb{H}^d$.

From here on, any reference to a map M is to that of Definition 2.1.9 unless otherwise indicated.

We note that in [12], the Cartan development is defined such that

$$Z_t^{(k)} = \int_{0<t_1<\ldots<t_k<t} M(dX_{t_k})\cdots M(dX_{t_1}),$$

that is, the matrix product within the integral is given in the reverse order to that in (2.1.3). Under this definition, we would get that $D_M(X*Y) = D_M(Y)\cdot D_M(X)$. We choose our definition in order to match that of the signature transform, but as a consequence the interpretation of the hyperbolic development here differs slightly. In the context of piecewise linear paths, this difference amounts to the order of path segments being reversed under hyperbolic development, although the developments of linear paths themselves remain identical. By way of example, if we were to compute the hyperbolic development of a path $X*Y$, the resulting path on $\mathbb{H}^d$ would correspond to that of $Y*X$ under the definition in [12]. The main consequence of this is that our inversion procedure in 3 works forwards through the path X, whereas the procedure in works [12] backwards.

Example 2.1.3. *Let $X : [a,b] \to \mathbb{R}^d$ be a piecewise linear path such that*

$$X_t = (x_1 * \cdots * x_n)_t\ , \quad x_k : [t_{k-1}, t_k] \to \mathbb{R}^d$$

where $t_0 = a, t_n = b$. X has hyperbolic development $D_M(X)\cdot o = D_M(x_1)\cdots D_M(x_n)\cdot o$, where each matrix $D_M(x_k)$ is a transformation on $\mathbb{H}^d$ corresponding to the linear path x_k. In this sense, we iteratively develop the path onto $\mathbb{H}^d$ starting at its last segment, where

$$D_M(X_{t_{k-1},b})\cdot o = D_M(x_k)\cdot D_M(X_{t_k,b})\cdot o,$$

where we use the notation $X_{s,t}$ to denote the path X restricted to $[s,t] \subset [a,b]$. With this in mind, we can consider the path on the hyperboloid corresponding to the development of X, which we define as

$$\begin{aligned} &H : [a,b] \to \mathbb{H}^d \\ &H_s = D_M(X_{a+b-s,b})\cdot o = D_M(-\overleftarrow{X}_s)\cdot o. \end{aligned} \tag{2.1.8}$$

Defining $s_k = a+b-t_{n-k}$ for each k, we see that H is the path starting at o and finishing at $D_M(X)\cdot o$, with $H_{s_k} = D_M(X_{t_{n-k},b})\cdot o$. When we refer to the 'path' of the hyperbolic development of X as opposed to simply the hyperbolic development, it is meant in reference to this path H_s. We will also use H_a in place of $D_M(X)\cdot o$ for brevity.

This definition/notation is certainly more unwieldly than that of [12] who's equivalent notion is simply the path $Z_t \cdot o$ where Z_t in this case is their alternative definition of the Cartan development. However, as mentioned above, this enables easier comparison with the signature and has very little effect on the procedure to invert the development.

We can further study the behaviour of H_s on the hyperboloid by using Lemma 2.1.4 to derive the development of an arbitrary linear path onto $SO(d,1)$. This is calculated fully in Appendix A.1.

Proposition 2.1.10 (Development of a Linear Path onto $SO(d,1)$ [12, Equation 3.7])**.** *Consider a linear path $X \in C_1([a,b],\mathbb{R}^d)$ with direction $\dot{X} = \theta \in \mathbb{S}^{d-1}$ and length L. The matrix of its development onto $SO(d,1)$ is*

$$D_M(X) = \exp\begin{pmatrix} 0 & L\theta \\ L\theta^\top & 0 \end{pmatrix}$$

$$= \begin{pmatrix} (\cosh(L)-1)\theta\theta^\top + I_d & \sinh(L)\theta \\ \sinh(L)\theta^\top & \cosh(L) \end{pmatrix}. \tag{2.1.9}$$

Example 2.1.4 (Geometric Interpretation for the Path of the Hyperbolic Development)**.** *Let X be a linear path in $C_1([0,T],\mathbb{R}^d)$ with length L and direction θ. Consider any point $p = (p_1,\dots,p_{d+1})^\top \in \mathbb{H}^d$ and the transformed point $D_M(X)\cdot p = p' = (p'_1,\dots,p'_{d+1})^\top$. From* (2.1.9)*, for each $k \in \{1,\dots,d\}$ we have that $p'_k - p_k = \alpha\,\theta_k$ where*

$$\alpha = \left(\sum_{i=1}^{d} p_i\theta_i(\cosh(L)-1) + p_{d+1}\sinh(L)\right). \tag{2.1.10}$$

Now, denoting by $\widehat{p}$ the projection of p onto $\mathbb{R}^d$ such that $\widehat{p} = (p_1,\dots,p_d)^\top$ and likewise the point $\widehat{p}' = (p'_1,\dots,p'_d)^\top$, we have that

$$\frac{\widehat{p}' - \widehat{p}}{\|\widehat{p}' - \widehat{p}\|} = \frac{\alpha\,\theta}{|\alpha|\,\|\theta\|} = \theta$$

Namely, if H is the path of the hyperbolic development of X as in (2.1.8)*, then the shadow of H, denoted $\widehat{H}$, on $\mathbb{R}^d$ has the same direction as X and has length $|\alpha|$ in $\mathbb{R}^d$. With this in mind, we see that H itself is the path between o and $D_M(X)$, along the intersection of $\mathbb{H}^d$ with the plane containing o and $D_M(X)\cdot o$, parallel to e_{d+1}. Likewise, if H is instead the path of the hyperbolic development of a piecewise linear path $(x_1 * \cdots * x_n)_t$, then the shadow of H on $\mathbb{R}^d$ is also a piecewise linear path $(\widehat{H}_1 * \cdots \widehat{H}_n)_s$ where the direction of $\widehat{H}_k$ is equal to that of $x_{n-(k-1)}$ for each k. For a visual intuition of this in 3 dimensions, we can examine the development of the path X seen in Figure 2.2, particularly its plan view in the bottom right panel showing the piecewise linear nature of H when projected onto the x-y plane. As above, we see that the last segment of $\widehat{H}$ is in the positive x direction, the second last in the positive y direction and so on. That is, their order is the reverse of those in X. The path H itself can then be considered similarly to the case when X is linear, where H is split into distinct segments corresponding to the intersection of $\mathbb{H}^d$ with the planes parallel to e_{d+1} and containing the start and endpoints of each $\widehat{H}_k$.*

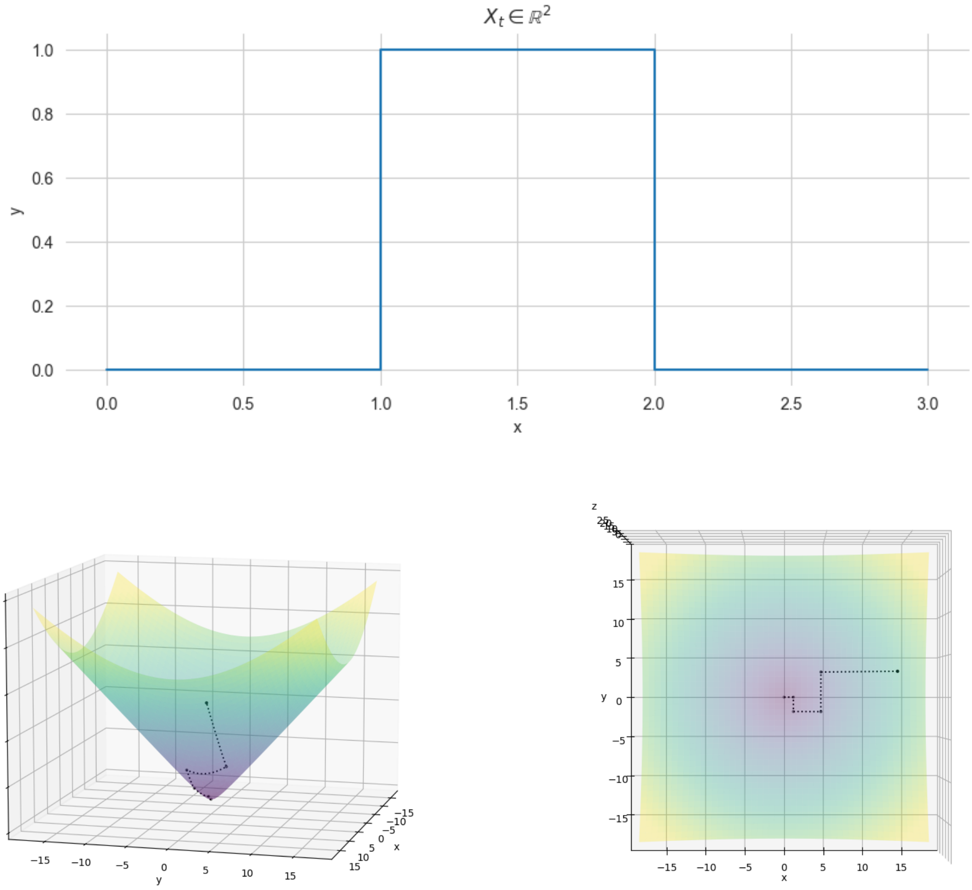

Figure 2.2: Axis path X and its hyperbolic development (plan view right).

Remark 2.1.11. *Recalling the observations made in Example 2.1.2, we note the similarity of the map M relating to the hyperbolic development with that given onto $\mathfrak{su}(d)$. Indeed, we see that for the hyperbolic development,*

$$D_M^{(1)}(X) = \begin{pmatrix} 0 & 0 & S(X)^1 \\ 0 & 0 & S(X)^2 \\ S(X)^1 & S(X)^2 & 0 \end{pmatrix}$$

and

$$D_M^{(2)}(X) = \begin{pmatrix} S(X)^{1,1} & S(X)^{1,2} & 0 \\ S(X)^{2,1} & S(X)^{2,2} & 0 \\ 0 & 0 & S(X)^{1,1} + S(X)^{2,2} \end{pmatrix},$$

so again we gain desirable geometric information about a path from the first two terms of the development. Noticing that

$$D_M(X) \cdot o = \sum_{k=0}^{\infty} \left(D_M^{(k)}(X) \cdot (0, \ldots, 0, 1)^\top \right),$$

namely that the hyperbolic development is the infinite sum of the last columns of each coefficient $D_M^{(k)}(X)$, we can extend the above relationship with signature coefficients. Letting the j^{th} coordinate of $D_M(X) \cdot o$ be $x^{(j)}$ for a path X in $\mathbb{R}^d$ and defining the set of indices

$$W_{2n} := \{w : w = (i_1, i_1, i_2, i_2, \ldots, i_m, i_m, \ldots, i_n, i_n)\},$$

we have that each coordinate is the sum of coordinate iterated integrals given by

$$x^{(j)} = \begin{cases} \sum_{n=0}^{\infty} \sum_{w \in W_{2n}} S^{w}(X) & \text{when } j = d+1 \\ \sum_{n=0}^{\infty} \sum_{w \in W_{2n}} S^{j,w}(X) & \text{when } j \leq d \quad \text{[12, Equation 3.5].} \end{cases} \tag{2.1.11}$$

As a consequence of the above, we can see that inverting the hyperbolic development also gives a method to invert the signature transform.

Chapter 3

Inverting the Path Development

Our goal in this chapter is to retrieve a piecewise linear path from its Cartan development on $SO(d,1)$. Adapted from the procedure in sections 3 and 4 of [12], we aim to obtain the direction and length of the first linear section in the path, from which we can write down the matrix of its Cartan development. We then multiply by the inverse[1] of this matrix, which thanks to its form in (2.1.8), removes the first linear section from the development. Iterating until we arrive back at the point o gives us the lengths and directions of each linear section of our piecewise linear path.

3.1 Reclaiming the First Linear Section from the Hyperbolic Development

The hyperbolic development has multiple desirable characteristics for our goal of reclaiming the path itself. Firstly, we note that since $\mathbb{H}^d$ is a surface of revolution when embedded in $\mathbb{R}^{d+1}$, every point on it can be uniquely expressed as

$$\begin{pmatrix} \eta^{(1)} \sinh(\rho) \\ \eta^{(2)} \sinh(\rho) \\ \vdots \\ \eta^{(d)} \sinh(\rho) \\ \cosh(\rho) \end{pmatrix} \tag{3.1.1}$$

where $\eta = (\eta^{(1)}, \cdots, \eta^{(d)})^\top$, is an element of the unit sphere $\mathbb{S}^{d-1}$ and $\rho \in \mathbb{R}^+$ [12, Section 3.1]. Here, η represents the point's orientation and ρ represents its hyperbolic distance from the point o. We further note that if we represent this same point as $(x_1, \ldots, x_{d+1})^\top$, we have that

$$\eta = \frac{(x_1, \ldots, x_d)^\top}{\|(x_1, \ldots, x_d)^\top\|}.$$

A particular advantage of the negative curvature of $\mathbb{H}^d$, is that as a path H on its surface gets further from o, its speed $\|\dot{H}\|$ increases, causing it to become increasingly 'stretched' in $\mathbb{R}^{d+1}$. In particular, if H is the path of the development of a piecewise linear path X, sections of H high on the hyperboloid become longer in Euclidean space relative to those lower down the surface. We can exploit this property by scaling our path X by some large $\lambda \in \mathbb{R}$, denoting the development of λX by $H(\lambda)$. This effectively pushes the endpoint $H_T(\lambda)$ higher up the surface than the rest of the path. With $\|H(\lambda)\|$ now largest close to $H_T(\lambda)$, the direction of its last section (corresponding to the first section of X) dominates that of the subsequent ones for large enough λ. Thus, as λ increases, the

[1] $SO(d,1) \subset GL_{d+1}(\mathbb{R})$

hyperbolic development of λX tends to the development of a linear path in $\mathbb{R}^d$ which has slope equal to that of the first section of X. Using this property enables us to extract the direction of the first linear piece using the coordinates of η.

The properties of $\mathbb{H}^d$ further aid us in retrieving information about the length of a path. Again scaling our path by $\lambda > 0$, denote by $\rho(\lambda)$ the hyperbolic distance from o to $D_M(\lambda X) \cdot o$. As we increase λ, the hyperbolic development stretches along the surface, increasingly close to a geodesic from o to $D_M(\lambda X) \cdot o$. The hyperbolic length of such a geodesic is $\rho(\lambda)$, which tends to the length of λX as λ increases [7, Proposition 3.8]. We can use this to approximate the length of the path X by the limiting behaviour of $\rho(\lambda)/\lambda$.

Theorem 3.1.1 (Derivative of the First Linear Piece [12, Theorem 3.4])**.** *Let X be a piecewise linear path in $C_1([0, L], \mathbb{R}^d)$ under unit-speed parameterisation and write*

$$X_t = (x_1 * x_2 * \cdots * x_n)_t$$

where each x_i is a linear path with direction $\theta_i \in \mathbb{S}^{d-1}$. Let $D_M(\lambda X) \cdot o = H_L(\lambda)$ be the end of the path of the hyperbolic development of λX for some $\lambda > 0$. Using $(\eta(\lambda), \rho(\lambda)) \in \mathbb{S}^{d-1} \times \mathbb{R}^+$ to parameterise $H_L(\lambda)$ as in (3.1.1), we have that

$$\theta_1 = \lim_{\lambda \to \infty} \eta(\lambda). \tag{3.1.2}$$

Theorem 3.1.2 (Length of the First Linear Piece [12, Theorem 4.1])**.** *Define a path X and its hyperbolic development H_L as in Theorem 3.1.1. Let the length of each x_i be l_i, such that $l_1 + l_2 + \cdots + l_n = L$. Then for large enough λ, $\exists\, c, C > 0$ such that*

$$ce^{-\lambda l_1} < |\eta(\lambda) - \theta_1| < Ce^{-\lambda l_1}, \tag{3.1.3}$$

meaning we can extract the length l_1 by calculating

$$\lim_{\lambda \to \infty} \frac{1}{\lambda} \log |\eta(\lambda) - \theta_1| = -l_1 \tag{3.1.4}$$

Example 3.1.1. *In order to build a geometric intuition for the above, consider the simple two-dimensional axis path $X \in C_1([0, 5], \mathbb{R}^2)$ shown in the top panel of Figure 2.2 and defined as*

$$X = (e_1) * (e_2) * (e_1) * (-e_2) * (e_1)$$

where $\{e_1, e_2\}$ is the standard basis of $\mathbb{R}^2$. Using lemmas 2.1.3 and 2.1.4, for $\lambda \in \mathbb{R}$ we can calculate the endpoint $D_M(\lambda X)$ on $SO(2, 1)$ as

$$D_M(\lambda X) = D_M(\lambda e_1) \cdot D_M(\lambda e_2) \cdot D_M(\lambda e_1) \cdot D_M(-\lambda e_2) \cdot D_M(\lambda e_1)$$

where

$$D_M(\lambda e_1) = \begin{pmatrix} \cosh(\lambda) & 0 & \sinh(\lambda) \\ 0 & 1 & 0 \\ \sinh(\lambda) & 0 & \cosh(\lambda) \end{pmatrix}, \; D_M(\pm\lambda e_2) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cosh(\lambda) & \pm\sinh(\lambda) \\ 0 & \pm\sinh(\lambda) & \cosh(\lambda) \end{pmatrix}.$$

As in Example 2.1.4, we note that the matrices of $D_M(e_1)$ and $D_M(e_2)$ translate points along $\mathbb{H}^2$ in the direction of e_1 and e_2 respectively. The full development onto $SO(2, 1)$ is the path Z starting at I_3 and ending at $D_M(\lambda X)$ such that $Z_t = D_M(\lambda X^t)$ where X^t is the path X stopped at time t. We can then find the path $H(\lambda)$ of the hyperbolic development of λX as the path starting at o and ending at $D_M(\lambda X) \cdot o$, defined as in Example 2.1.4. $H(\lambda)$ is shown in Figure 3.1 for $\lambda = 1$ and $\lambda = 3$. Here we get visual evidence of the behaviour described at the start of this section, whereupon increasing λ, the first segment dominates subsequent segments in the path's development. We observe in particular that the endpoint of $H(\lambda)$ is pointing in the positive x direction, close to the plane $y = 0$ when $\lambda = 3$, suggesting that the first segment of X moves also in the positive x direction, a geometric realisation of Theorem 3.1.2.

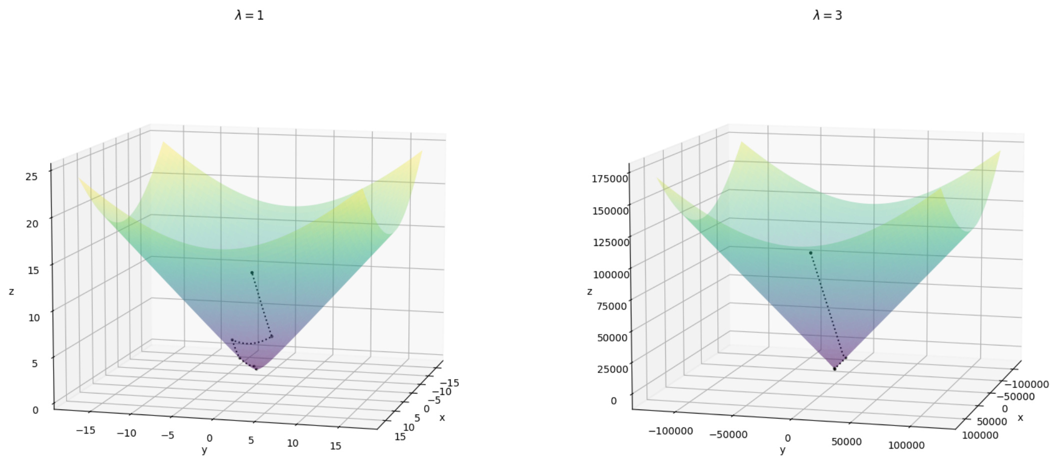

Figure 3.1: Development of the axis path λX (see figure 2.2 for $\lambda = 1$ (left) and $\lambda = 3$

Theorems 3.1.1 and 3.1.2 now give us a method to retrieve the length and direction of the first component of a piecewise linear path, using the asymptotics of its hyperbolic development under rescaling. The next step in our inversion procedure as described at the start of this section, is to remove the first linear section from the hyperbolic development. This requires obtaining and inverting the matrix of the Cartan development of the section in question.

We note that the inverse of a matrix e^A is just e^{-A}, so the inverse of (2.1.9) is simply the development of the linear path with length L and direction $-\theta$. Consequently, if $X = (x_1 * x_2 * \cdots * x_n)$ is a piecewise linear path, using (2.1.6) we obtain

$$D_M\big(x_2 * \cdots * x_n\big) = D_M\big(x_1\big)^{-1} \cdot D_M\big(x_1 * x_2 * \cdots * x_n\big)$$

where

$$D_M\big(x_1\big)^{-1} = \begin{pmatrix} (\cosh(l_1) - 1)\theta_1\theta_1^\top + I_d & -\sinh(l_1)\theta_1 \\ -\sinh(l_1)\theta_1^\top & \cosh(l_1) \end{pmatrix} \tag{3.1.5}$$

and x_1 has length l_1 and direction θ_1.

3.1.1 Obtaining the Cartan Development of a Scaled Path

Given that our goal is to find a path knowing only its development onto $SO(d, 1)$, the natural question brought up by the procedure above is how one can obtain the development of λX knowing only the development of X. Without having some knowledge of the structure of the path, we can not do this using just the single matrix $D_M(X)$. Instead, we must consider the infinite series in (2.1.2).

Example 3.1.2 (Development of a scaled path)**.** *Let G be a Lie group and M a linear map from $\mathbb{R}^d$ to its Lie algebra $\mathfrak{g}$. For a path X from $[a, b]$ to $\mathbb{R}^d$ and $\lambda \in \mathbb{R}$, the development of λX onto G under M is then $D_M(\lambda X)$. As in* (2.1.2)*, this development is given by*

$$D_M(\lambda X) = I + \sum_{k=1}^{\infty} D_M^{(k)}(\lambda X).$$

For every k, we can write

$$\begin{aligned} D_M^{(k)}(\lambda X) &= \int_{a<t_1<\ldots<t_k<b} M(\lambda dX_{t_1}) M(\lambda dX_{t_2}) \cdots M(\lambda dX_{t_k}) \\ &= \lambda^k \int_{a<t_1<\ldots<t_k<b} M(dX_{t_1}) M(dX_{t_2}) \cdots M(dX_{t_k}) \end{aligned}$$

by the linearity of M. It then follows that we can write

$$D_M(\lambda X) = I + \sum_{k=1}^{\infty} \lambda^k D_M^{(k)}(X). \tag{3.1.6}$$

Likewise, using (2.1.5) *we can write*

$$D_M(\lambda X * \lambda Y) = \sum_{k=0}^{\infty} \sum_{i=0}^{k} \left(\lambda^i D_M^{(i)}(X)\right) \left(\lambda^{k-i} D_M^{(k-i)}(Y)\right) = \sum_{k=0}^{\infty} \lambda^k D_M^{(k)}(X * Y)$$

for some path Y in $\mathbb{R}^d$.

The example above gives us the remaining theoretical piece to complete the inversion algorithm, albeit requiring having access to the full series representation as opposed to just one matrix in $SO(d,1)$. In reality, we will only ever have finitely many terms of the above series, so for this procedure to be practically feasible we must be able to truncate the series without losing important information. As mentioned at the beginning of Section 2.1, the terms of the signature transform decay factorially in k. We can show an equivalent result for the Cartan development, proceeding as in [2, Section 1.2.2] by first defining the following parameterisation for a path $X \in C_1([a,b],\mathbb{R}^d)$ with length L:

$$\tau(s) = \inf\left\{t \geq a : \|X\|_{1,[a,t]} > s\right\} \tag{3.1.7}$$

where we set $\inf \emptyset = b$ and where $\|X\|_{1,[a,t]}$ denotes the length of X on the interval $[a,t]$. Under this parameterisation, $X \circ \tau$ is a continuous path in $C_1([0,L],\mathbb{R}^d)$. Furthermore, we have that $X^\tau := X \circ \tau$ is in fact Lipschitz continuous with Lipschitz constant 1 [2, Exercise 1.6.6]. As the Cartan development is invariant under reparameterisation by any non-decreasing function, we have that $D_M(X) = D_M(X^\tau)$.

Proposition 3.1.3 (Factorial decay of the infinite sequence)**.** *Let X be a path in $C_1([a,b],\mathbb{R}^d)$, G a Lie group and M a linear map from $\mathbb{R}^d$ to $\mathfrak{g}$, the Lie algebra of G. For any matrix norm $\|\cdot\|$ we have that*

$$\left\|D_M^{(k)}(X)\right\| \leq \frac{C^k \|X\|_1^k}{k!} \tag{3.1.8}$$

for any k, where C is a constant depending only on M and the norm $\|\cdot\|$.

Proof (Adapted from [2, Proposition 1.2.3]). Let $X^\tau : [0,L] \to \mathbb{R}^d$ be the path X reparameterised under τ as defined in (3.1.7) where $L = \|X\|_1$. Both the Cartan development and the length of X are invariant under reparameterisation, so the proof for X follows from the proof for X^τ. Since X^τ is Lipschitz continuous it is also absolutely continuous, so by the Radon-Nikodym Theorem there exists an integrable function $(X^\tau)'$ such that

$$X_t^\tau = \int_0^t (X_s^\tau)' ds$$

for every $t \in [0, L]$. The derivative of X^τ_t is thus $(X^\tau_t)'$, and it exists almost everywhere on $[0, L]$. We then have that

$$\begin{aligned}
D^{(k)}_M(X) &= \int_{0<t_1<\ldots<t_k<L} M(dX^\tau_{t_1}) \cdots M(dX^\tau_{t_k}) \\
&= \int_{0<t_1<\ldots<t_k<L} M\left((X^\tau_{t_1})'\right) \cdots M\left((X^\tau_{t_k})'\right) dt_1 \ldots dt_k \\
\implies \left\| D^{(k)}_M(X) \right\| &\leq \int_{0<t_1<\ldots<t_k<L} \left\| M\left((X^\tau_{t_1})'\right) \cdots M\left((X^\tau_{t_k})'\right) \right\| dt_1 \ldots dt_k. \\
&\leq \int_{0<t_1<\ldots<t_k<L} \left\| M\left((X^\tau_{t_1})'\right) \right\| \cdots \left\| M\left((X^\tau_{t_k})'\right) \right\| dt_1 \ldots dt_k.
\end{aligned}$$

Now, since X^τ has Lipschitz constant 1, we have that $\|(X^\tau)'\|_2 \leq 1$ almost everywhere. It follows that since M is linear, there exists $C < \infty$ depending only on M and $\|\cdot\|$ such that $\|M((X^\tau)')\| < C$ almost everywhere. This allows us to write

$$\begin{aligned}
\left\| D^{(k)}_M(X) \right\| \leq\ & C^k \int_{0<t_1<\ldots<t_k<L} dt_1 \ldots dt_k. \\
=\ & \frac{C^k L^k}{k!}
\end{aligned}$$

□

Proposition 3.1.3 tells us that if we choose large enough n, $D^{(k)}_M(X)$ has negligible effect on $D_M(X)$ for all $k \geq n$. With this in mind, we introduce the notion of the step-n truncated development of a path,

$$D^{\leq n}_M(X) := I + \sum_{k=1}^{n} D^{(k)}_M(X).$$

Remark 3.1.4. *While the computation of $D^{\leq n}_M(X)$ is $\mathcal{O}(n)$ for a linear path X, we note that when X is piecewise linear, that is $X = X_1 * X_2 * \cdots * X_m$ with m linear pieces, the computation of the product $D^{\leq n}_M(X) = D^{\leq n}_M(X_1) \cdot D^{\leq n}_M(X_2) \cdots D^{\leq n}_M(X_m)$ is $\mathcal{O}(n^2)$.*

3.2 Inversion for Axis Paths

An issue with the use of the truncated sum $D^{\leq n}_M(X)$ is of course the truncation error. Due to the form of (3.1.6) and (3.1.8), we can see that this truncation error will increase in general when we scale our path by $\lambda \gg 1$. With this in mind, we would like for our inversion procedure to converge quickly with respect to λ. For axis paths in particular, there are finitely many possible directions for each linear segment, simplifying the inversion considerably. In fact, we leverage information on the possible directions of the second segment in a path in order to obtain information about the length of the first segment. As a result, we propose the method in the following example as an alternative to (3.1.4) for axis paths.

Example 3.2.1. *Let $X : [0, T] \to \mathbb{R}^d$ be an axis path defined as*

$$X = (\alpha_1 e_{i_1}) * \cdots * (\alpha_n e_{i_n})$$

where $i_k \in \{1, \dots, d\}$ and a_k is a linear path from $[t_{k-1}, t_k]$ to $\mathbb{R}^d$ for each k. As before we know that the path $H(\lambda)$ of the development of λX on the hyperboloid can be written as

$$H_s(\lambda) = \begin{pmatrix} \eta_s^{(1)}(\lambda)\sinh(\rho_s(\lambda)) \\ \vdots \\ \eta_s^{(d)}(\lambda)\sinh(\rho_s(\lambda)) \\ \cosh(\rho_s(\lambda)) \end{pmatrix}$$

Using notation and remarks from Example 2.1.4, we know that the projection of H onto $\mathbb{R}^d$ is an axis path $\widehat{H} = (\beta_1 e_{i_n}) * (\beta_2 e_{i_{n-1}}) * \cdots (\beta_n e_{i_1}) \in \mathbb{R}^d$ for some scalars $\{\beta_j\}$ where we note the reverse correspondence of the direction of each segment with that of X. As before, we define $s_k = a + b - t_{n-k}$ and see from Theorem 3.1.2 that

$$\lim_{\lambda\to\infty} \frac{\widehat{H}_{s_k}(\lambda)}{\|\widehat{H}_{s_k}(\lambda)\|} = \lim_{\lambda\to\infty} \eta_{s_k}(\lambda) = e_{i_{n-(k-1)}}$$

for all $k \in \{0, \dots, n-1\}$. That is, each vertex $\widehat{H}_{s_k}(\lambda)$ of $\widehat{H}(\lambda)$ becomes asymptotically close (in relative terms) to the $e_{i_{n-(k-1)}}$ axis, which we see an example of in Figure 3.2, comparing the path λX from Example 3.1.1 for $\lambda \in \{1, 3\}$.

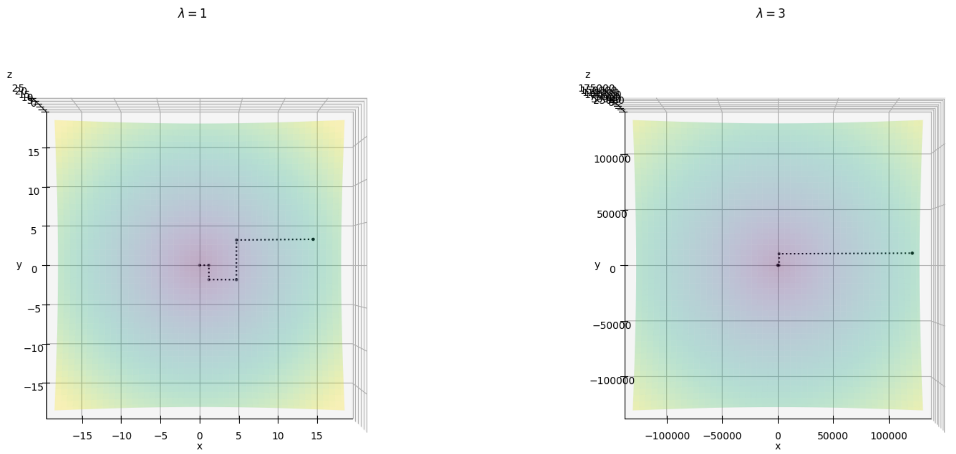

Figure 3.2: Plan view of the development in Figure 3.1, for $\lambda = 1$ (left) and $\lambda = 3$

Now, say we have obtained the direction e_{i_1} of the first segment of X. We know that

$$\begin{aligned} H_{s_1}(\lambda) &= D_M(\lambda(\alpha_2 e_{i_2} * \cdots * \alpha_n e_{i_n})) \cdot o \\ &= D_M(\lambda \alpha_1 e_{i_1})^{-1} D_M(\lambda X) \cdot o \\ &= D_M(-\lambda \alpha_1 e_{i_1}) D_M(\lambda X) \cdot o \end{aligned}$$

where $\frac{\widehat{H}_{s_1}(\lambda)}{\|\widehat{H}_{s_1}(\lambda)\|} \to e_{i_2}$ in the limit of λ and in particular, the i_1^{th} coordinate of $\frac{\widehat{H}_{s_1}(\lambda)}{\|\widehat{H}_{s_1}(\lambda)\|}$ tends to zero. As a result, if we write

$$D_M\big(-\lambda l\, e_{i_1}\big) \cdot D_M\big(\lambda X\big) \cdot o = \begin{pmatrix} \eta^{(1)}(\lambda, l)\sinh(\rho(\lambda, l)) \\ \vdots \\ \eta^{(d)}(\lambda, l)\sinh(\rho(\lambda, l)) \\ \cosh(\rho(\lambda, l)) \end{pmatrix}$$

where η and ρ are now functions of λ and some signed length $l \in \mathbb{R}$, and denote by l_λ the solution to

$$\eta^{(i_1)}(\lambda, l_\lambda) = 0,$$

we see that

$$\lim_{\lambda\to\infty} l_\lambda = \lambda\alpha_1. \tag{3.2.1}$$

We can see the existence of this solution by noting that for any point $p \in \mathbb{H}^d$ with its i_1^{th} coordinate 0 and some $l \in \mathbb{R}$, we have from (2.1.10) that if $p' = D_M(le_{i_1}) \cdot p$, then

$$\|\widehat{p}' - \widehat{p}\| = |p^{(d+1)}\sinh(l)|.$$

The existence follows from the surjectivity of the hyperbolic sine and the fact that

$$D_M(-le_{i_1}) \cdot p' = p.$$

Figure 3.3 shows the behaviour of $\eta(\lambda, l)$ for the path X from Example 2.1.4, with $\lambda = 10$. We can see that $\eta^{(1)}(10, l)$ crosses zero very close to $l = 1$, the length of the last segment of X. Additionally, $\eta^{(2)}(10, l)$ is close to 1 at the same point, leading to the (correct) conclusion that $i_2 = 2$ for this particular path. This procedure provides a new method for inverting an axis path, whereby we replace the length approximation of Theorem 3.1.2 by the solution l_λ and iterate through the path as before.

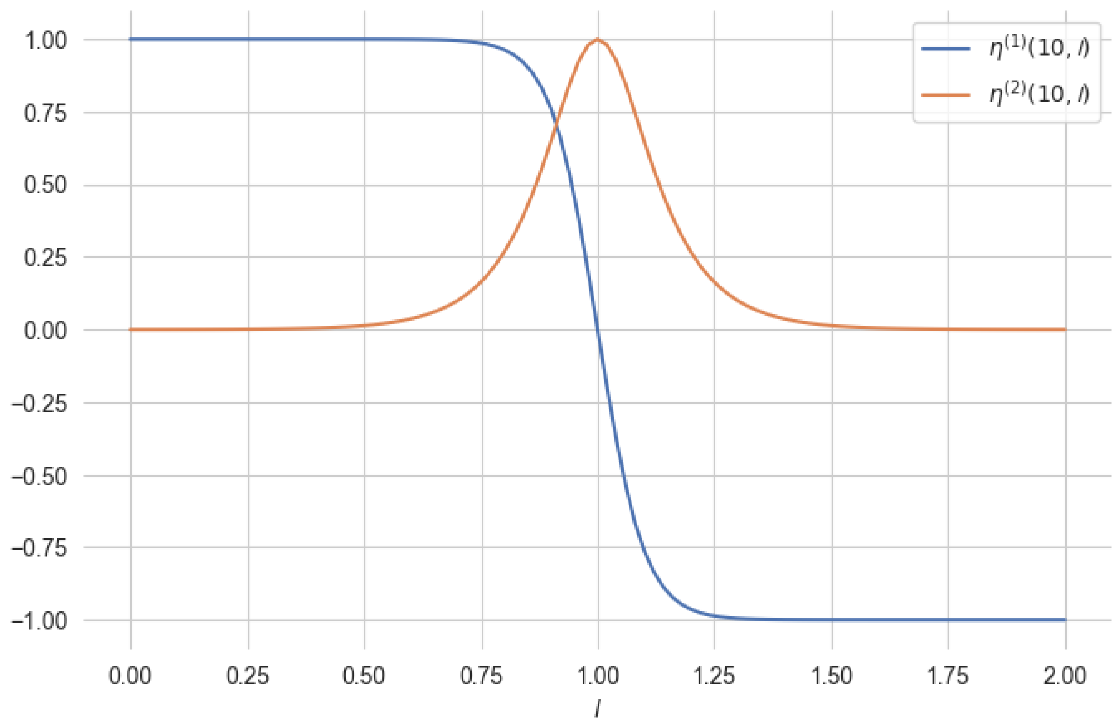

Figure 3.3: $\eta(\lambda, l)$ for the path X with $\lambda = 10$.

As mentioned at the beginning of this section, we aim for this procedure to converge quickly with respect to λ, that is, more quickly than (3.1.4). We have from Theorem 3.1.2 that if X is an axis path with its k^{th} segment having direction e_{i_k} and length α_k, then $\|\eta(\lambda) - e_{i_1}\| = \mathcal{O}(e^{-\lambda\alpha_1})$ which gives

$$\left|\frac{1}{\lambda}\log\|\eta(\lambda) - e_{i_1}\| + \alpha_1\right| = \mathcal{O}\left(\frac{1}{\lambda}\right).$$

The following shows that the decay in the error of the above procedure is much faster than this and is in fact exponential.

Proposition 3.2.1. *Let X be an axis path of the form*

$$X_t = (\alpha_1 e_{i_1}) * \cdots * (\alpha_n e_{i_n})$$

and l_λ the solution to

$$\eta^{(i_1)}(\lambda, l) = 0$$

as defined in Example 3.2.1, where $\eta^{(i_1)}(\lambda, l) \cdot \sinh(\rho(\lambda, l))$ is the i_1^{th} coordinate of $D_M(-\lambda l e_{i_1}) \cdot D_M(\lambda X) \cdot o$. Then,

$$|l_\lambda - \alpha_1| = \mathcal{O}\left(\frac{1}{\lambda} e^{-\lambda \alpha_2}\right).$$

Proof. Let $l_\lambda = \varepsilon_\lambda + \alpha_1$ for some error ε_λ where we note that

$$\begin{aligned} \lambda l_\lambda e_{i_1} &= (\lambda \varepsilon_\lambda e_{i_1}) * (\lambda \alpha_1 e_{i_1}) \\ \implies D_M(\lambda l_\lambda e_{i_1}) &= D_M(\lambda \varepsilon_\lambda e_{i_1}) \cdot D_M(\lambda \alpha_1 e_{i_1}). \end{aligned}$$

As before, $D_M(-\lambda \alpha_1 e_{i_1}) H_{s_0}(\lambda) = H_{s_1}(\lambda)$ and writing $H'_{s_1}(\lambda) = D_M(-\lambda l_\lambda e_{i_1}) \cdot H_{s_0}(\lambda)$ we see that

$$H'_{s_1}(\lambda) = \begin{pmatrix} \eta^{(1)}_{s_1}(\lambda) \sinh(\rho_{s_1}(\lambda)) \\ \vdots \\ \eta^{(i_1-1)}_{s_1}(\lambda) \sinh(\rho_{s_1}(\lambda)) \\ 0 \\ \eta^{(i_1+1)}_{s_1}(\lambda) \sinh(\rho_{s_1}(\lambda)) \\ \vdots \\ \eta^{(d)}_{s_1}(\lambda) \sinh(\rho_{s_1}(\lambda)) \\ \cosh(\rho_{s_1}(\lambda)) \end{pmatrix}$$

by the definition of l_λ and the fact that $D_M(-\lambda \varepsilon_\lambda e_{i_1})$ affects only the i_1^{th} coordinate of $\widehat{H}_{s_1}(\lambda)$. Clearly, we now have that

$$\left\|\widehat{H}'_{s_1}(\lambda) - \widehat{H}_{s_1}(\lambda)\right\| = \left|\eta^{(i_1)}_{s_1}(\lambda) \sinh(\rho_{s_1}(\lambda)\right|. \tag{3.2.2}$$

We also know that $H_{s_1}(\lambda) = D_M(\lambda \varepsilon_\lambda e_{i_1}) H'_{s_1}(\lambda)$ and Example 2.1.4 gives that

$$\begin{aligned} \left\|\widehat{H}'_{s_1}(\lambda) - \widehat{H}_{s_1}(\lambda)\right\| &= \left|H'_{s_1}(\lambda)^{(i_1)}(\cosh(\lambda \varepsilon_\lambda) - 1) + H'_{s_1}(\lambda)^{(d)} \sinh(\lambda \varepsilon_\lambda)\right| \\ &= |\cosh(\rho_{s_1}(\lambda)) \sinh(\lambda \varepsilon_\lambda)|, \end{aligned}$$

whereupon combining with (3.2.2) we get

$$\left|\eta^{(i_1)}_{s_1}(\lambda)\right| = |\coth(\rho_{s_1}(\lambda)) \sinh(\lambda \varepsilon_\lambda)|. \tag{3.2.3}$$

Now, from (3.1.3) we know that

$$\begin{aligned} \left|\eta_{s_k}(\lambda) - \operatorname{sgn}(\alpha_{k+1}) e_{i_{k+1}}\right| &= \mathcal{O}(e^{-\lambda \alpha_{k+1}}) \\ \implies \sqrt{\left(\eta^{(i_1)}_{s_1}(\lambda)\right)^2} &= \mathcal{O}(e^{-\lambda \alpha_2}) \\ \implies \left|\eta^{(i_1)}_{s_1}(\lambda)\right| &= \mathcal{O}(e^{-\lambda \alpha_2}). \end{aligned} \tag{3.2.4}$$

Since $\rho_{s_1}(\lambda) \geq 0$ we have that $\coth(\rho_{s_1}(\lambda) \geq 1$ and finally, combining this with (3.2.3), and (3.2.4) yields

$$\begin{aligned}
|\coth(\rho_{s_1}(\lambda))\sinh(\lambda\varepsilon_\lambda)| &\geq |\sinh(\lambda\varepsilon_\lambda)| \\
&= \left|\sum_{k=0}^{\infty} \frac{(\lambda\varepsilon_\lambda)^{2k+1}}{(2k+1)!}\right| \\
&\geq |\lambda\varepsilon_\lambda| \\
\implies |\varepsilon_\lambda| &= \mathcal{O}\left(\frac{1}{\lambda}e^{-\lambda\alpha_2}\right)
\end{aligned}$$

□

Chapter 4

Numerical Experiments

We note before detailing the numerical examples below, that even though we frame these experiments as an inversion procedure for the hyperbolic development, it is in fact equivalent to inverting the signature due to (2.1.11). Likewise, inverting the hyperbolic development truncated at level k is equivalent to inverting the signature truncated at level k, since again we can write $D_M^{(k)}(X) \cdot o$ explicitly in terms of the k^{th} level of the signature.

See [8] for any code related to the following numerical experiments.

4.1 Preliminary Experiments

We begin this chapter by first providing some numerical examples of Theorem 3.1.1 and Proposition 3.8 from [7], comparing the convergence of Theorem 3.1.2 with that of Proposition 3.2.1 as well as examining the factorial decay of $\|D_M^{(k)}(\lambda X)\|$ with respect to k. For the purposes of this example, we again use the axis path X shown in Figure 2.2.

The first experiments focus on one iteration of the inversion procedure described above, determining the direction and length of the first segment of X. Figure 4.1 shows the behaviour of $\eta(\lambda)$ as λ is increased, in particular the distance $\|\eta(\lambda) - e_1\|$. The exponential rate of convergence is clear, with λ not needing to be taken overly large in this case to obtain a close estimate.

We now compare the two methods given above for the estimation of the length α_1 of the first linear segment of X. In this simple case, $\alpha_1 = 1$, and Figure 4.2 shows the absolute error of the estimates for increasing values of λ. Here, 'Method 1' refers to that given in Theorem 3.1.2, and 'Method 2' to that in Proposition 3.2.1. The difference in convergence rates is stark, with the error from method 1 remaining larger at $\lambda = 30$ than that of method 2 for $\lambda = 5$. While this does not provide all that much use theoretically, the positive relationship between $\max_k \|D^{(k)}(\lambda X)\|$ and λ that was alluded to at the start of Section 3.2 means that this improvement is practically quite beneficial, meaning that the infinite series can be truncated to much fewer terms for a given accuracy in the estimate of α_1.

Proposition 3.8 from [7] states that there exists a $C > 0$ such that

$$L\lambda - \frac{C}{\lambda} \leq \rho(\lambda) \leq L\lambda$$

where L is the length of X. Recalling that the length of X in this example is 5, we see evidence of this convergence in Figure 3.3. However, with $|\rho(\lambda) - L|$ being $\mathcal{O}\left(\frac{1}{\lambda}\right)$ this takes a notably larger λ to yield reasonable accuracy than is the case with $\eta(\lambda)$. While this particular information is not explicitly used in the inversion procedure, it is certainly useful to be able to extract the length of a path from its development, and consequently its signature due to 2.1.11.

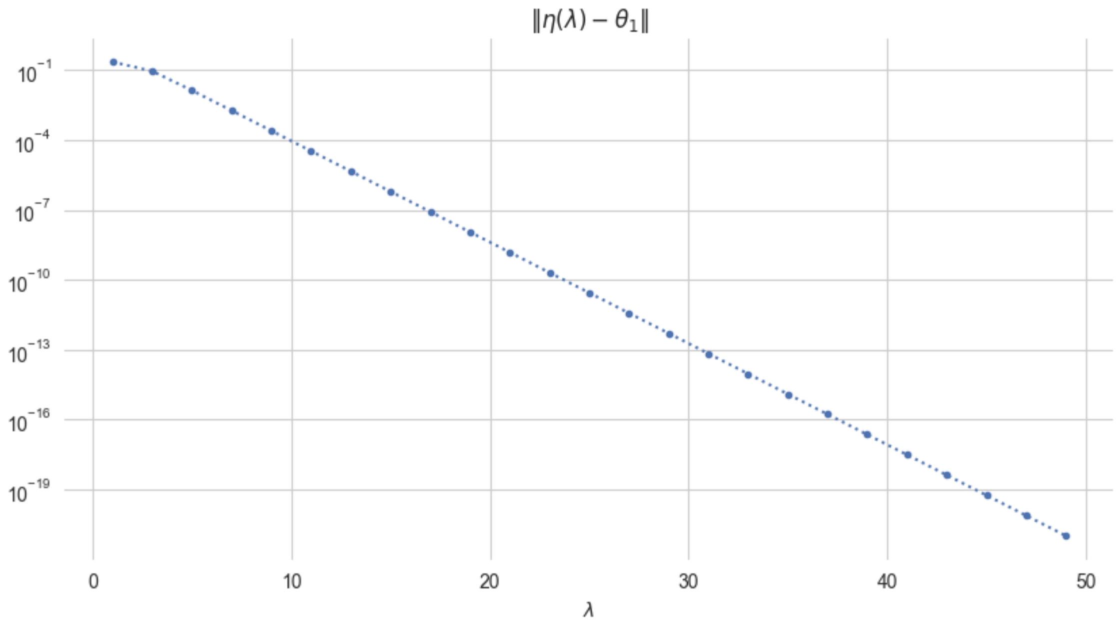

Figure 4.1: Convergence of $\|\eta(\lambda) - \theta_1\|$ for the path X.

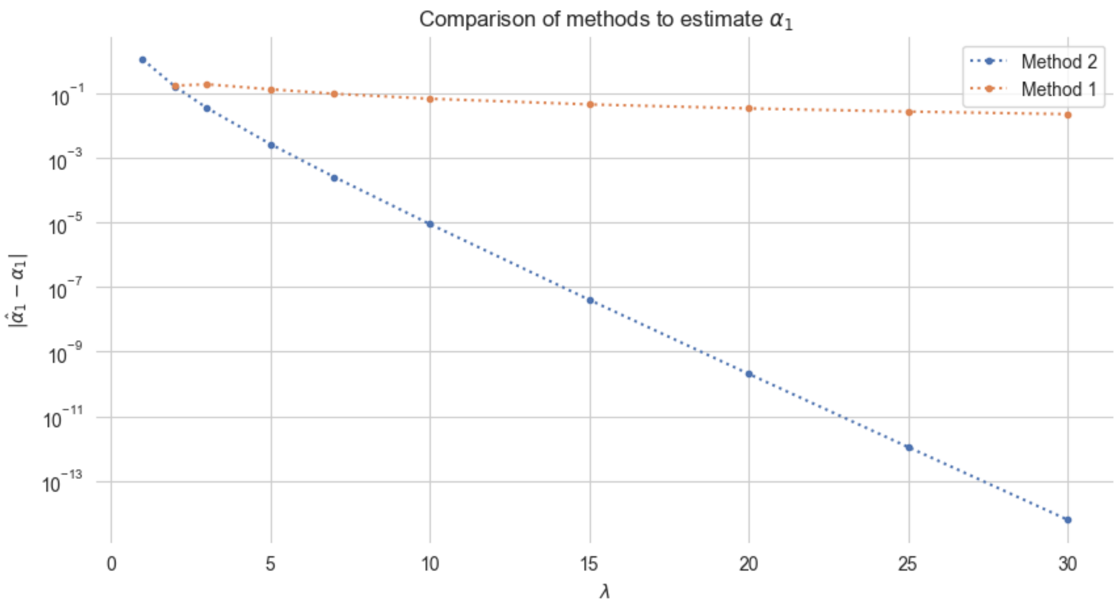

Figure 4.2: Comparing the convergence rates of (3.1.4) and (3.2.1).

The final experiment in this section concerns the behaviour of the norms $\|D_M^{(k)}(\lambda X)\|$ in relation to Proposition 3.1.3, specifically the effect of λ. Figure 4.4 shows the Frobenius norms of the first 100 coefficients of the development of λX for various values of λ on a log scale. We observe the factorial decay noted above, as well as the increase in $\arg\max_k \|D_M^{(k)}(\lambda X)\|_F$ as we increase λ. In the simplest case where we consider the hyperbolic development of a single segment of an axis path, we have that

$$\left\| D_M^{(k)}(\lambda \alpha_j e_{i_j}) \right\|_F = \frac{\left(\lambda\sqrt{|2\alpha_j|}\right)^k}{k!}$$

who's maximum argument is the mode of a Poisson distribution with parameter $\lambda\sqrt{|2\alpha_j|}$, namely $\left\{\lceil \lambda\sqrt{|2\alpha_j|}\rceil - 1, \lfloor \lambda\sqrt{|2\alpha_j|}\rfloor\right\}$. This increases linearly in λ, and although this is

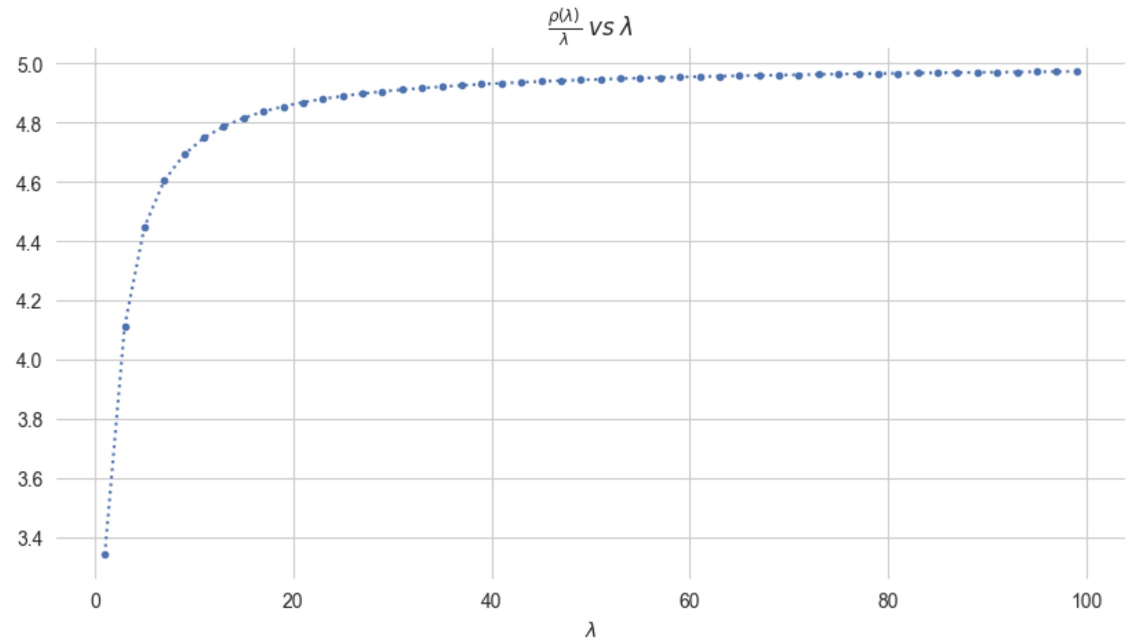

Figure 4.3: Convergence of $\frac{\rho(\lambda)}{\lambda}$ to $L = 5$.

not necessarily the case for the full path X, we can certainly see why not needing to scale the path by as large a factor could be useful computationally.

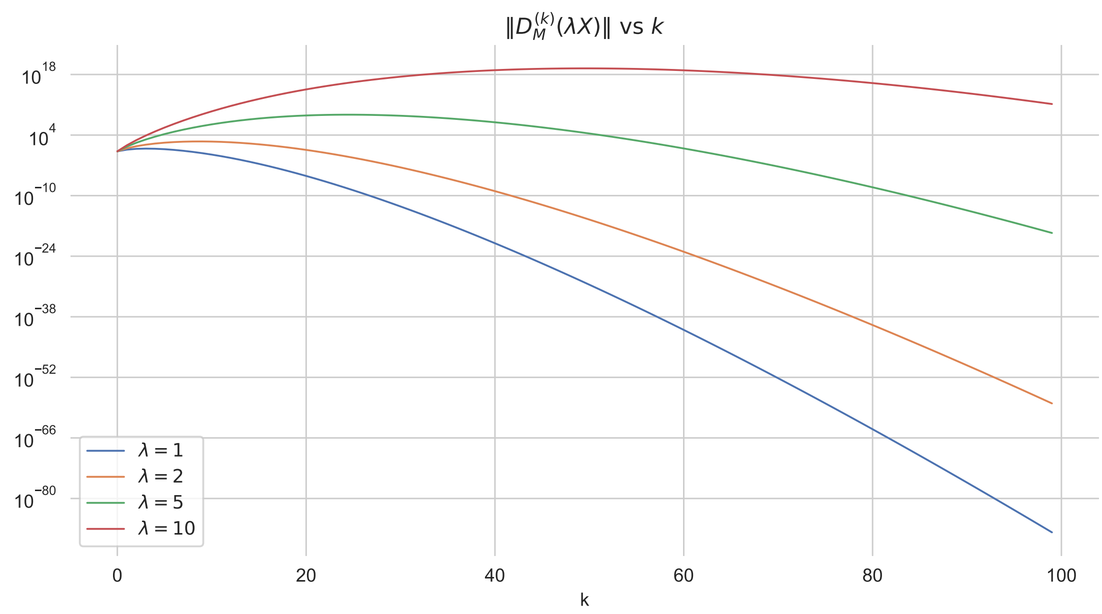

Figure 4.4: Frobenieus norms of the coefficients $D_M^{(k)}(\lambda X)$.

4.2 Axis Path Inversion

4.2.1 Paths on an Integer Lattice

Moving on to the inversion itself, we first consider a greatly simplified version of the problem, inverting the development of integer lattice paths. A d-dimensional integer lattice path is an axis path which moves in steps of length 1 between points in $\mathbb{Z}^d$. What this means in the context of inverting its development is that we may ignore completely the estimation of the length of each segment, knowing that each length is 1. We may still of course have linear segments with length greater than 1, but these can be treated as consecutive segments with the same direction when inverting. Having no need to calculate length, we only need to take λ large enough such that

$$\arg\min_j \|\eta_{s_k}(\lambda) \pm e_j\| = i_{n-(k-1)}$$

for each $k \in \{1, \ldots, n\}$. The act of rounding $\eta(\lambda)$ to the nearest basis vector means that we need much less accuracy for η, and as such can reduce the required λ further.

This simplification allows us in certain cases to reclaim a lattice path from the single matrix $D_M(\lambda X)$, without needing its full series representation. Having only this one matrix means that we are unable to vary λ, but if we possess $D_M(\lambda X)$ for a single large enough λ, this is enough in many cases to retrieve X without needing to vary λ at all. This would be beneficial in the context of generative models such as those mentioned during the introduction, allowing us to generate a single matrix in $SO(d, 1)$ rather than a point in $(M_{d+1}(\mathbb{R}))^n$ which is needed when using the truncated series representation $D_M^{\leq n}(X)$. This could be done for example by scaling the training data by a factor of λ.

The main computational issue using this procedure is that the necessary matrix operations suffer computational instability as λ increases. In particular, the formulation of (2.1.9) means that as λ becomes large its determinant becomes very unstable, flipping between 0 and 1 for small perturbations in its entries. This is clearly not a desirable characteristic given that the inversion procedure requires its inverse. This issue can be mitigated using symbolic computational packages such as `sympy` in Python, though this dramatically increases computation times for more complex paths. Using the full series representation, we are able to increase λ when estimating η, while also reducing it when performing the matrix operations necessary to remove a path segment. When we possess only the matrix $D_M(\lambda X)$ for some large, fixed λ, we are forced to perform these operations on matrices with extremely large values, leading to computational inaccuracies which are only amplified as we iterate through the procedure.

With the above in mind, paths which require a large λ for η to converge will be much more difficult to invert computationally. Paths which are 'difficult' in this respect are those which contain many twists and loops, requiring a larger λ for them to straighten out on the hyperbolic surface. If we for example constrain ourselves to the class of paths

$$\{X = (\alpha_1 e_{i_1}) * \cdots * (\alpha_n e_{i_n}) : \alpha_k > 0 \; \forall \; k\},$$

then inversion using this method (and in general) would become dramatically easier as these paths do not loop back on themselves at any point.

3-dimensional lattice path

For the first numerical example of this simplified inversion, a 3-dimensional lattice path of length 10 was generated randomly, yielding the path defined by

$$X = (e_3 * -e_2 * -e_1 * e_3 * e_2 * e_3 * -e_1 * -e_2 * -e_3 * -e_2).$$

The inversion procedure was then carried out using the matrices $D_M(X)$ and $D_M(2X)$. The results of this procedure are shown in Table 4.1 where the colour denotes whether the respective direction matches that of the original path.

	Direction									
	1	2	3	4	5	6	7	8	9	10
X	e_3	$-e_2$	$-e_1$	e_3	e_2	e_3	$-e_1$	$-e_2$	$-e_3$	$-e_2$
Estimate for $\lambda = 1$	e_3	$-e_2$	$-e_1$	e_3	e_3	$-e_1$	e_2	$-e_3$	e_2	$-e_1$
Estimate for $\lambda = 2$	e_3	$-e_2$	$-e_1$	e_3	e_2	e_3	$-e_1$	$-e_2$	$-e_3$	$-e_2$

Table 4.1: Segment directions given by the inversion procedure on $D_M(X)$

We can see for this simplified example that we actually only needed to scale the path by a factor of 2 in order to correctly reclaim the direction of each segment. In general, for any lattice path X there does in theory exist a λ such that we can correctly reclaim the path from its development $D_M(\lambda X)$. However, as above, this λ is path specific and we have no quantitative results on any global minimum requirement for a certain class of paths, say all paths of length n.

As a comparison, we also test the procedure on the path

$$Y = (e_1 * e_2 * e_3 * e_1 * e_2 * e_3 * e_1 * e_2 * e_3 * e_1),$$

a member of the class of paths mentioned above which move only in the positive direction of each axis. The results of inverting $D_M(Y)$ are shown in Figure 4.2, where we note that its lack of loops or pronounced twisting means that we can invert it perfectly while only taking $\lambda = 1$.

	Direction									
	1	2	3	4	5	6	7	8	9	10
Y	e_1	e_2	e_3	e_1	e_2	e_3	e_1	e_2	e_3	e_1
Estimate for $\lambda = 1$	e_1	e_2	e_3	e_1	e_2	e_3	e_1	e_2	e_3	e_1

Table 4.2: Segment directions given by the inversion procedure on $D_M(Y)$

A final note on this special case is that inversion is much more computationally stable when increasing the path's dimension rather than increasing the number of segments. See, for a fixed λ, as we add more segments to a path the endpoint of its hyperbolic development will generally be pushed exponentially higher up the hyperbolic surface. On the one hand, this is useful as we do not need as large a λ to be able to extract information about the first few segments, given that their position on the surface already causes them to stretch out. However, we still require λ to be large enough to retrieve the segments lower down the surface, and increasing λ when the path is already quite long leads to the numerical instability discussed at the beginning of the section. That is, the computational issues stem not just from the magnitude of λ, but more specifically the length of λX. We can though add complexity through the path dimension without similar adverse effects.

4.2.2 General Axis Paths

For more general piecewise linear paths or even just more complex lattice paths, we may require the full series representation of $D_M(X)$ in order to successfully reclaim the path X. As discussed in the previous section, the general procedure is to increase λ when approximating $\eta(\lambda)$, then reducing it when removing each segment in order to avoid the aforementioned numerical instability. We first provide an example on a lattice path to contextualise the previous section.

10-dimensional lattice path

We again use a randomly generated axis path, call it X, this time in 10 dimensions and of length 30. The inversion uses the same ideas as before, rounding each estimate of $\eta(\lambda)$ to the nearest signed basis vector, removing a segment of length one and starting again. Table 4.3 shows the results of this procedure, skipping the full form of X for brevity, though this time we compare across the level of truncation used as opposed to the value of λ. We can see that although we can correctly retrieve the directions of the first two

	Direction									
	1	2	3	4	5	$\cdots$	27	28	29	30
X	$-e_7$	e_4	$-e_9$	e_5	e_8	$\cdots$	e_4	$-e_{10}$	e_3	e_6
Estimate for $D_M^{\leq 10}(X)$	$-e_7$	e_4	e_8	e_5	$-e_9$	$\cdots$	e_4	$-e_4$	e_4	$-e_4$
Estimate for $D_M^{\leq 20}(X)$	$-e_7$	e_4	$-e_9$	e_5	e_8	$\cdots$	e_4	$-e_{10}$	e_3	e_6

Table 4.3: Segment directions given by the inversion procedure on $D_M^{\leq n}(X)$

segments from the series $D_M^{\leq 10}(X)$ in this case, we have cut out too much information to be able to retrieve any further segments (the fact that segment 27 was correctly estimated is purely coincidental). However, adding 10 more terms to the truncated series allows us to correctly retrieve all 30 segments of the path. Examining this further, we see in Figure 4.5 that the terms $D_M^{(k)}(X)$ increase up to roughly the 22^{nd} term, and that as we suspected, a considerable amount of 'information' in this sense is contained after term 15. Notably though, there are numerous terms after the 20^{th} level with sizeable norms, an interesting observation given that we were able to reclaim the path from just these first 20.

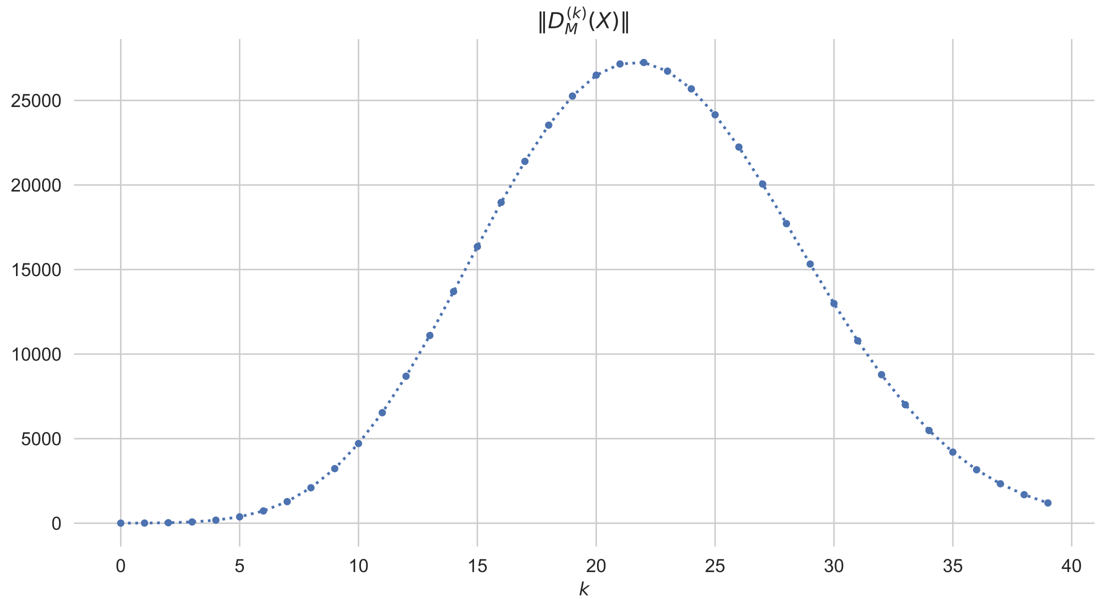

Figure 4.5: Frobenius norms of the coefficients $D_M^{(k)}(X)$ for the lattice path X.

Estimating the length of each segment

We now move on to a more general problem, where the path does not necessarily take values on an integer lattice, and we need to estimate the lengths of each segment. For ease of writing, we do in fact use a lattice path in this example, though the procedure to calculate length remains unchanged for non-integer values.

Consider the 3-dimensional axis path $X := (e_1 * e_2 * e_3)$. Without any knowledge of the length of each segment of X, the procedure to invert its truncated development $D_M^{\leq n}(X)$ is as follows:

1. Calculate $D_M^{(k)}(\lambda X) = \lambda^k D_M^{(k)}(X)$ for each k and some large λ.
2. Estimate $H(\lambda) \approx D_M^{\leq n}(\lambda X) \cdot o$ and calculate $\eta(\lambda) = \frac{\widehat{H}(\lambda)}{\|\widehat{H}(\lambda)\|}$
3. Letting e_i be the closest basis vector to $\eta(\lambda)$, calculate $H'(\lambda, l) = D^{\leq n}(-\lambda l e_i) \cdot H(\lambda)$.
4. Set $\eta(\lambda, l) = \frac{\widehat{H}'(\lambda,l)}{\|\widehat{H}'(\lambda,l)\|}$ and numerically calculate the root l_λ.
5. Calculate the new development $D^{\leq n}(-\lambda l_\lambda e_i) \cdot D^{\leq n}(X)$ by removing the estimate of the first segment.
6. Iterate until we arrive back at the point o.

Applied to the path X, for the first iteration of the above procedure (see [8]) we calculate $D_M^{\leq 50}(X)$, finding that the first direction e_{i_1} is the vector e_1. We then obtain $\eta(\lambda, l)$ as a function of l shown in Figure 4.6, where $\lambda = 10$.

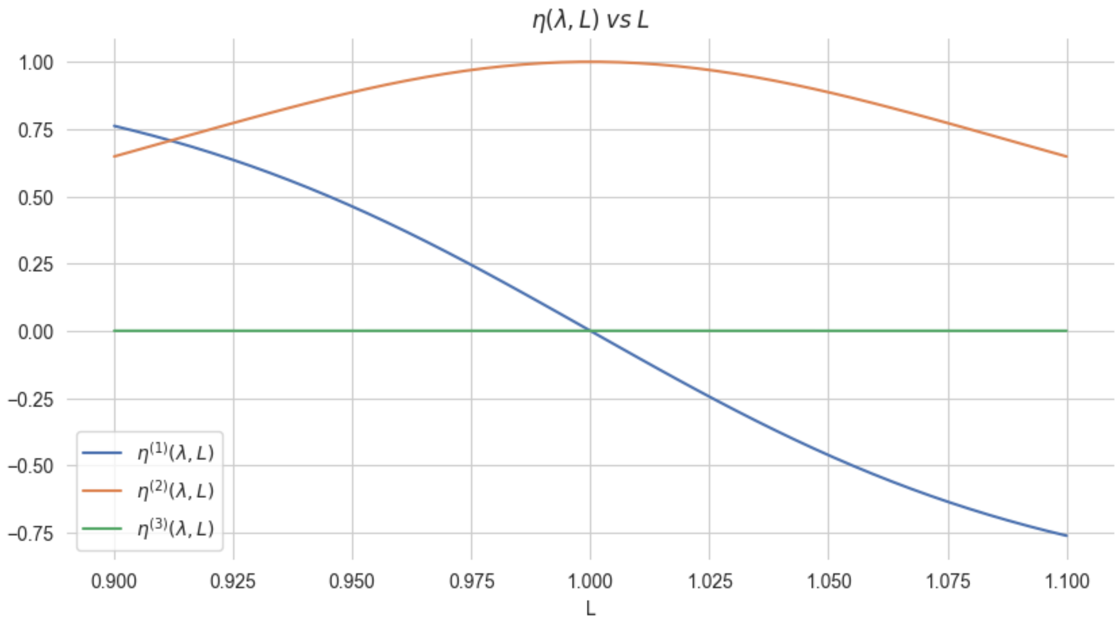

Figure 4.6: $\eta(\lambda, l)$ as a function of l.

Numerically, we calculate the root l_λ of $\eta^{(1)}(10, l)$ as $l_\lambda \approx 0.99999796$. Letting X' denote the path X with its first segment removed, we can then calculate

$$D_M^{\leq 50}(X') \approx D_M^{\leq 50}(-0.99999796\, e_1) \cdot D_M^{\leq 50}(X).$$

We then begin the procedure again on X', noting from Figure 4.6 that as the second coordinate of $\eta(10, l)$ is approximately one when its first coordinate is zero, we can already infer that the direction e_{i_2} is equal to e_2. Without explicitly detailing the remaining two iterations, we see the estimated lengths α_k and directions e_{i_k} of the three segments in Table 4.4. We note that we found it generally beneficial in our experiments to dynamically reduce the value of λ used to calculate the length of each segment as we iterated through the path. This is due to the fact that a large value of λ accentuates any slight errors in the calculations of previous segments' lengths, leading to reduced accuracy for the current segment.

	Segment		
	1	2	3
e_{i_k}	e_1	e_2	e_3
α_k	0.99999796	1.000002041	0.99999796

Table 4.4: Segment directions and lengths found by inverting $D_M^{\leq 50}(X)$
.

Clearly, this is a very simple example of the inversion procedure for a more general axis path, though serves as a proof of concept for Proposition 3.2.1 as well as the theory presented in [12]. As is the case with lattice paths, this procedure is considerably more stable to added path dimension compared with increasing the number of segments, due to the propagation of errors as we iterate through a path.

Conclusion

In this thesis, we have studied the emergence of the Cartan development as a tool for the study of continuous paths, particularly in relation to the development of a time-series when treated as a piecewise linear path. A particular focus was to study the viability of the Cartan development for use in synthesising time-series data, namely whether it would be possible to reclaim a path from a Cartan development produced as the output of a generative model.

Having detailed the relevant background concepts, we introduced the notion of developing a path onto a Lie group, examined its close relationship with the signature transform as well as presenting some crucial lemmas for the calculation of the Cartan development in practice. We focused then on a particularly useful case of the Cartan development, developing paths onto the group of orientation preserving isometries of the hyperbolic surface $\mathbb{H}^d$. This development allowed us to relate each linear path segment to an isometry on $\mathbb{H}^d$, which we could then iteratively apply starting at its origin, yielding a path on the hyperbolic surface. We then detailed the desirable geometric properties of the hyperbolic surface and how we could leverage them to invert the hyperbolic development, and consequently the signature transform of a piecewise linear path. Having established existing theory for this application, we focused in particular on axis paths, those which comprise linear segments parallel to the standard basis vectors of $\mathbb{R}^d$ and which in two dimensions can be considered as piecewise constant time-series data. We proposed a new method to reclaim the length of each segment of a general axis path from its hyperbolic development, and showed that it provided a considerable improvement in rate of convergence over the existing method.

In Chapter 4, we provided numerical support for the convergence theorems detailed in Chapter 3. We implemented a much simplified procedure to successfully invert the development (or signature) of integer lattice paths, and detailed the cases in which it proved effective as well as its computational shortcomings. We then provided a numerical implementation for the inversion of more general axis paths, enabling us to reclaim the approximate lengths and directions of each linear segment from the path's development in a simple example.

Overall, though the more computational sophistication is required for the above to be a viable inversion procedure on more general paths, it appears at the very least to be adequate as a method for inverting the development of integer lattice paths. For general piecewise linear paths, very small errors in the estimation of direction or length can amplify quickly as we iterate through the path, due to the nature of the hyperbolic surface. However, it seems that there could be scope for mitigating these effects by dynamically adjusting the λ parameter in order to balance the growth of errors with being able to obtain the necessary information.

Appendix A

Additional Proofs

A.1 Development of a Linear Path

Proposition A.1.1 (Proposition 2.1.10)**.** *Consider a linear path $X \in C_1([a,b],\mathbb{R}^d)$ with direction $\dot{X} = \theta \in \mathbb{S}^{d-1}$ and length L. The matrix of its development onto $SO(d,1)$ is*

$$D_M(X) = \begin{pmatrix} (\cosh(L)-1)\theta\theta^\top + I_d & \sinh(L)\theta \\ \sinh(L)\theta^\top & \cosh(L) \end{pmatrix}.$$

Proof. Denote by A the matrix

$$M(X_b - X_a) = \begin{pmatrix} 0 & L\theta \\ L\theta^\top & 0 \end{pmatrix}.$$

In order to calculate e^A, we use the fact that $e^{P\Lambda P^{-1}} = Pe^\Lambda P^{-1}$ for a diagonal matrix Λ and invertible matrix P [6]. When $\theta_1 \neq 0$, the first $d-2$ eigenvectors of A are

$$\left\{ \begin{pmatrix} -\frac{\theta_2}{\theta_1} \\ 1 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} -\frac{\theta_3}{\theta_1} \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, \quad \dots \quad, \begin{pmatrix} -\frac{\theta_{d-1}}{\theta_1} \\ 0 \\ 0 \\ 0 \\ \vdots \\ 1 \\ 0 \end{pmatrix} \right\}$$

and when the first i components of θ are zero, these vectors are

$$\left\{ \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, \quad \dots \quad, \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ -\frac{\theta_{i+2}}{\theta_{i+1}} \\ 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ -\frac{\theta_{i+3}}{\theta_{i+1}} \\ 0 \\ 1 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, \quad \dots \quad, \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ -\frac{\theta_{d-1}}{\theta_{i+1}} \\ 0 \\ 0 \\ \vdots \\ 1 \\ 0 \end{pmatrix} \right\}$$

where entry $i+1$ of the k^{th} vector is $\frac{\theta_k}{\theta_{i+1}}$. These $d-1$ eigenvectors have eigenvalue 0, and the remaining two eigenvectors are $(\pm\,\theta,\ 1)^\top$ which have eigenvalue $\pm L$ respectively.

Therefore, A is diagonalizable such that $A = P\Lambda P^{-1}$, where P is the matrix of eigenvectors of A and $\Lambda = \mathrm{diag}(0, \ldots, 0, -L, L)$. We can now can write

$$e^A = Pe^\Lambda P^{-1} = P \begin{pmatrix} 1 & & & & \\ & \ddots & & & \\ & & 1 & & \\ & & & e^{-L} & \\ & & & & e^L \end{pmatrix} P^{-1}$$

$$= \begin{pmatrix} (\cosh(L) - 1)\theta\theta^\top + I_d & \sinh(L)\theta \\ \sinh(L)\theta^\top & \cosh(L) \end{pmatrix}.$$

□

Bibliography

[1] B. Barancikova, Z. Huang, and C. Salvi, *Sigdiffusions: Score-based diffusion models for long time series via log-signature embeddings*, 2024.

[2] T. Cass and C. Salvi, *Lecture notes on rough paths and applications to machine learning*, arXiv preprint arXiv:2404.06583, (2024).

[3] I. Chevyrev and A. Kormilitzin, *A primer on the signature method in machine learning*, 2016.

[4] I. Chevyrev and T. Lyons, *Characteristic functions of measures on geometric rough paths*, The Annals of Probability, 44 (2016).

[5] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, *Generative adversarial nets*, in Advances in Neural Information Processing Systems, Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K. Weinberger, eds., vol. 27, Curran Associates, Inc., 2014.

[6] B. C. Hall, *Lie groups, Lie algebras, and representations*, Springer, 2015.

[7] B. Hambly and T. Lyons, *Uniqueness for the signature of a path of bounded variation and the reduced path group*, Annals of Mathematics, 171 (2010), p. 109–167.

[8] R. Kenny, *An invertible lie group embedding for piecewise linear paths*, 2024. GitHub repository, `https://github.com/Rhyskenny1/An-Invertible-Lie-Group-Embedding-for-Piecewise-Linear-Paths`.

[9] S. Li, Z. Lyu, H. Ni, and J. Tao, *On the determination of path signature from its unitary development*, 2024.

[10] H. Lou, S. Li, and H. Ni, *Path development network with finite-dimensional lie group representation*, arXiv preprint arXiv:2204.00740, (2022).

[11] H. Lou, S. Li, and H. Ni, *Pcf-gan: generating sequential data via the characteristic function of measures on the path space*, 2024.

[12] T. J. Lyons and W. Xu, *Hyperbolic development and inversion of signature*, Journal of Functional Analysis, 272 (2017), pp. 2933–2955.

[13] J. Morrill, C. Salvi, P. Kidger, J. Foster, and T. Lyons, *Neural rough differential equations for long time series*, 2021.

[14] Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole, *Score-based generative modeling through stochastic differential equations*, 2021.

[15] V. S. Varadarajan, *Lie groups, Lie algebras, and their representations*, vol. 102, Springer Science & Business Media, 2013.