MAPPING SPACES, SIGNATURES, AND DATA

Darrick Lee

A DISSERTATION
in
Applied Mathematics and Computational Sciences

Presented to the Faculties of the University of Pennsylvania
in
Partial Fulfillment of the Requirements for the
Degree of Doctor of Philosophy
2021

Supervisor of Dissertation

Robert Ghrist, Andrea Mitchell PIK Professor
Professor of Mathematics
Professor of Electrical and Systems Engineering

Graduate Group Chairperson

Robin Pemantle, Merriam Term Professor of Mathematics

Dissertation Committee:

Robert Ghrist, Professor of Mathematics and Electrical and Systems Engineering
Ryan Hynd, Associate Professor of Mathematics
Chad Giusti, Assistant Professor, Department of Mathematical Sciences, University of Delaware

Dedicated to Mom, Dad, 奶奶,
and Vanessa.

ACKNOWLEDGMENTS

The pursuit of understanding and the desire to make creative discoveries have been an enormous part of my life for the past five years. It was a dream to be in graduate school learning and doing mathematics, but I also often felt lost. I wanted to do something I would be proud of, yet afraid that this nebulous goal would never materialize. Throughout these five years, I have been incredibly fortunate to have had an amazing group of friends, colleagues, mentors, and of course family, whose unwavering support has kept this dream alive. I am grateful to all of you, for keeping me grounded, being there when things got a little rough, and for making my time in Philly truly special.

My advisor Rob Ghrist has been a constant source of inspiration, support, and encouragement. Before coming to Penn, I had the sense that applications and pure math were somewhat disparate ideas that seemed to conflict with one another. Rob's unique style of mathematics showed me otherwise and has had a profound impact on the development of my own mathematical taste. I am indebted to his belief in me, and for allowing me the intellectual freedom to explore. His dedication to teaching and exposition is unrivaled, and is something I will continually strive towards. Or maybe this is just one of those nebulous goals... *upside-down smiley emoji* Thank you Rob for being an incredibly supportive advisor, for your insight and advice, both mathematically and otherwise, and for the many group dinners over the years.

Chad Giusti has also been an incredible mentor during my time at Penn. I remember when we first started discussing path signatures in a coffee shop in West Philly. I was still unsure about how this tool for time series analysis was related to topology, and he brought out this big book on rational homotopy theory (I guess he just carries it around with him wherever he goes) and started telling me about iterated integrals and cochain models. I came out of each of those meetings thrilled and inspired to learn more. Working on projects with Chad has been an exceptionally fun and rewarding experience. Thank you Chad for sharing your deep knowledge and for your kind and patient support.

I am fortunate to have crossed paths with the other members of the applied topology research group at Penn: Ann, Brendan, Greg, Hans, Huy, Iris, Jakob, Joe, Miguel, Paige, Rachel and Zoe have all been a source of intriguing discussions and interesting ideas. In particular, I want to thank Jakob for always being willing to listen to my rambling ideas and sharing his acute insight on whatever he's stumbled upon. It has been a privilege to be part of such a talented group – thank you.

I have learned an incredible amount from the math faculty at Penn. Thank you to Mona Merling and Jonathan Block for teaching me beautiful ideas in algebraic topology; to Wolf-

gang Ziller for opening my eyes to geometry; and to Michael Steele for several wonderful courses in probability theory. Outside of Penn, I'd like to thank Yuliy Baryshnikov for several discussions during the early stages of my degree.

Furthermore, I'd like to thank Vidit Nanda for inviting me to Oxford for a brief visit in 2019 (as well as a longer visit, which was unfortunately cancelled due to the pandemic), and introducing me to the path signatures community there. The third part of this thesis is a collaboration which started with Chad and Vidit, as well as Harald Oberhauser and Ilya Chevyrev. Our project began at the onset of the pandemic, and provided me with an outlet during an intense period of uncertainty. Thank you all for sharing your deep insight and for having me as a part of such an exciting collaboration.

I am indebted to all my friends during these years – you have made Philly feel like a home, and graduate school would have been impossible without you. Jacob and Hadi, you two somehow made doing problem sets on a Friday night fun back in first year. Sammy, thank you for being an amazing friend over these years. Most of the geometry I know is because of you, and I am so happy we could share our love for our favorite Lie group (it is obviously $SU(60)$). Andrew, you have been there for me every step of the way at Penn, from exploring the city to sharing the stresses of academia. As you have said, I don't think we would've made it through without each other. Wynter and Claire, my time in Philly would have been completely different without you two – thank you for all the incredibly fun times and the warm memories.

LOTEM (shouting; this is the main mode of conversation with Lotem (not because she is angry but because she is very passionate)), it has been an adventure to go through high school, undergrad, and now graduate school together. I can't describe how nice it has been to have a familiar face around campus. You were my rock and constantly motivated me with your drive and determination – thank you.

I'd also like to thank all of my housemates during my time in Philly (all 25?! of you). I miss you Vinson. I especially want to thank my pandemic housemates Genna, Sarah, and Zoe. Our daily communal dinners and our backyard bonfires kept me sane during an otherwise insane time.

Finally, I want to thank my family for their unconditional love and support. Mom, Dad, 奶奶, and Vanessa: this thesis is yours as much as it is mine.

ABSTRACT

MAPPING SPACES, SIGNATURES, AND DATA

DARRICK LEE

ROBERT GHRIST

Time-varying phenomena are ubiquitous across pure and applied mathematics, from path spaces and stochastic differential equations to multivariate time series and dynamic point clouds. The *path signature* provides a powerful characterization of such sequential data in terms of power series of tensors, weaving together these diverse concepts. Originally defined as part of *Chen's iterated integral cochain algebra*, the path signature has since been used as the foundation for the theory of rough paths in stochastic analysis. More recently, it has been shown to be a *universal* and *characteristic* feature map for multivariate time series, providing theoretical guarantees for its application to time series analysis in the context of kernel methods in machine learning. This thesis extends the scope of the path signature to more complex parametrized data in two directions. First, we consider generalizations of the codomain of a path. We lift the theory of signatures to the setting of Lie group valued time series, adapting these tools for time series with underlying geometric constraints. Furthermore, we build a signature framework to study paths of persistence diagrams, objects which capture the evolving topological structure of dynamic data sets. Second, we consider maps parametrized by higher dimensional cubes by developing notions of the *mapping space signature*. Our approach returns to the topological origins of the signature as the 0-cochains of the Chen construction. We formulate a cubical variant of the mapping space construction, and use the resulting 0-cochains to define the mapping space signature and establish its basic properties.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

This thesis is split into three major parts.

PART I contains Chapter 1 and Chapter 2, which are expository chapters that provide the motivation and background for the remainder of the thesis. In Chapter 1, we motivate the path signature by consider lead/lag behavior in time series. We then present the core theory of path signatures in the context of machine learning applications, and provide a brief introduction to rough paths. However, one of the original formulations of the signature was in the context of algebraic topology, where they formed part of the 0-cochains of a cochain model for path spaces and loop spaces of manifolds [Che77]. Because this perspective has been largely absent in the recent literature on signature methods in machine learning, we provide an exposition of these ideas in Chapter 2.

The remainder of the thesis generalizes the theory and application of signatures in two directions. The bounded variation paths discussed in this introduction can be viewed as a subset

$$C^{1-var}(V) \subset C([0,1], V)$$

of the mapping space of continuous functions between $[0,1]$ and a Banach space V.

PART II contains Chapter 3 and Chapter 4, which generalize the path signature by considering maps with more general codomains. In particular, in Chapter 3, we show that all of the foundational theory presented in this introduction can be lifted to the setting of Lie groups. We show that the Lie group signature is also universal and characteristic, and provide two detailed numerical experiments to demonstrate these two properties.

In Chapter 4, we consider paths of persistence diagrams, which are objects that arise in topological data analysis. The standard space of persistence diagrams is simply a monoid, and lacks the algebraic and analytic structure required to define signatures. By connecting the space of persistence diagrams with the existing theory of Lipschitz-free spaces in functional analysis, we provide an isometric embedding of the space of finite persistence diagrams into a Banach space. This allows signature methods to be used to study the dynamic topological structure of temporally evolving point cloud data sets. We apply this theory to a parameter estimation problem for models of collective motion, and show that the signature method performs well, even in situations with missing data.

PART III contains Chapter 5, Chapter 6, and Chapter 7, which generalize the path signature by considering maps from higher dimension codomains. The approach in these

chapters exploit the connection between the signature and Chen's iterated integral cochain algebra from Chapter 2. We use a generalization of Chen's construction to mapping spaces [PT03; GTZ10] in order to motivate the definition of these higher dimensional signatures. In particular, the work in these chapters is restricted to the setting of *smooth maps*.

In Chapter 5, we reformulate the simplicial construction from [PT03; GTZ10] using cubical sets with connection. This cubical variant of the Chen construction is more amenable to the development of signatures from higher dimensional cubical domains. In particular, we prove that the cubical generalization of mapping space cochain algebra is a morphism of differential graded algebras to the de Rham complex of the smooth mapping space.

Next, in Chapter 6, we use the cubical construction to recover the 1-dimensional theory of path signatures. While we do not prove any new properties of the signature, this chapter shows that all of the algebraic structure of the path signature such as the shuffle product and Chen's identity, can be derived directly from the cochain model. Furthermore, the 1-dimensional setting allows us to connect the more abstract cochain construction with the more familiar path signature.

Finally, in Chapter 7, we study the cubical mapping space construction explicitly for maps

$$\gamma : \square^d \to \mathbb{R}^N$$

where the domain is the unit d-cube $\square^d := [0,1]^d$. We classify the resulting 0-cochains into several categories, and use these cochains to define notions of the *mapping space signature*. We then show that the mapping space signature satisfies several invariance properties, as well as generalized shuffle product structure. However, one of the main complications with the higher dimensional theory is the obstruction of a direct generalization of Chen's identity. Nevertheless, we provide a formula to express the signature of the concatenation of two maps using the signatures of the individual maps.

Contributions and Publications/Preprints

Parts of Chapter 1 are adapted from the following paper with Chad Giusti:

- C. Giusti, **D. Lee**, *Iterated integrals and population time series analysis*, Proceedings of the Abel Symposium, 2020. (arXiv:1811.03558).

Chapter 3 is adapted from the following preprint with Rob Ghrist:

- **D. Lee**, R. Ghrist, *Path signatures on Lie groups*, (arXiv:2007.06633).

Chapter 4 is adapted from the following paper in preparation with Chad Giusti:

- C. Giusti, **D. Lee**, *Signatures, Lipschitz-free spaces, and paths of persistence diagrams*

The content of Chapter 5, Chapter 6, and Chapter 7 is work in progress with Chad Giusti, Vidit Nanda, and Harald Oberhauser.

NOTATION

PATHS AND MAPS

- We use $\square^d := [0,1]^d$ to denote the unit d-cube.
- Throughout the thesis, paths $\gamma : [0,1] \to \mathbb{R}^N$ will be parametrized on the unit interval unless otherwise specified. In PART I and PART II, we will use subscripts to denote the time parameter

  $$\gamma_t := \gamma(t).$$

  However, in PART III, we will revert to the $\gamma(t)$ notation in order to coincide with maps from higher dimensional domains such as $\alpha(t_1, \dots t_d)$ for $\alpha : \square^d \to \mathbb{R}^N$.
- In PART I and PART II, we work with various spaces of paths. For a vector space V, we let
  1. $C^{p-var}(V)$ be the space of bounded p-variation paths;
  2. $C^{pr}(V)$ be the space of piecewise regular paths;
  3. $C^{pc}(V)$ be the space of piecewise continuous paths.

CATEGORIES

- Categories and categorical (co)limits are denoted using sans serif font.
- Let C be a category and $x, y \in \mathsf{C}$ be objects. The Hom set between x and y is denoted $\mathsf{C}(x,y)$.
- Let C and D be categories. The functor category from C to D is denoted by D^{C}.
- The cateogries used in this thesis are:
  1. Set, the category of sets
  2. Top, the category of topological spaces;
  3. Diff, the category of differentiable spaces;
  4. $\square$, the box category;
  5. $\square_c$, the box category with connection;

6. $\mathsf{C}^{\square^{\mathrm{op}}_{(c)}}$, the category of cubical objects in C (with connection);
7. $\mathsf{C}^{\square_{(c)}}$, the category of cocubical objects in C (with connection).

- Cubical objects are denoted using homological notation (subscripts), such as $K\bullet \in \mathsf{Set}^{\square^{\mathrm{op}}}$, while cocubical objects are denoted using cohomological notation (superscripts), such as $Z^\bullet \in \mathsf{Set}^{\square}$.
- All cubical sets, $K_\bullet \in \mathsf{Set}^{\square_c^{\mathrm{op}}}$, considered in this thesis are *finite*: $K_\bullet$ contains a finite number of non-degenerate cubes.

SETS AND SUBSETS

- Given $N \in \mathbb{N}$, we let $[N] \coloneqq \{1, \ldots, N\}$.
- The power set of $[N]$ is denoted by $2^{[N]}$.
- In PART III, we use the following convention for elements and subsets of $[N]$ and $2^{[N]}$:
  1. $i \in [N]$, elements of $[N]$ are denoted using lower-case characters;
  2. $I \subset [N]$, subsets of $[N]$ are denoted using upper-case characters;
  3. $\mathcal{I} \subset 2^{[N]}$, subsets of $2^{[N]}$ are denoted using calligraphic characters.

Part I

INTRODUCTION AND BACKGROUND

1

THE PATH SIGNATURE AND TIME SERIES ANALYSIS

1.1 MOTIVATION: LEAD/LAG BEHAVIOR AND SIGNED AREA

Consider the time series $\gamma^a = (\gamma_1^a, \gamma_2^a) : [0,1] \to \mathbb{R}^2$ and $\gamma^b = (\gamma_1^b, \gamma_2^b) : [0,1] \to \mathbb{R}^2$ shown in the figure below.

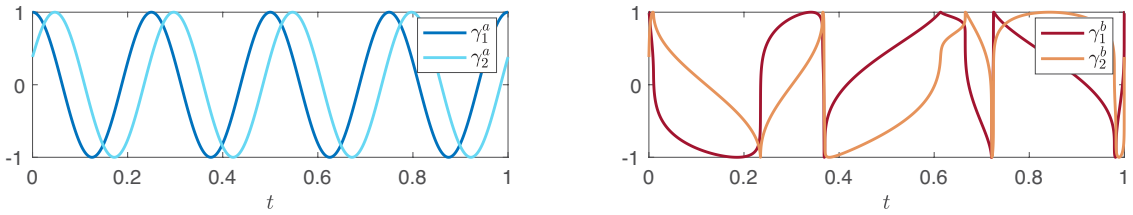


Figure 1.1: Two example time series. The time series $\gamma^a = (\gamma_1^a, \gamma_2^a)$ is plotted in blue on the left, while the time series $\gamma^b = (\gamma_1^b, \gamma_2^b)$ is plotted in red on the right.

Both of these time series qualitatively exhibit lead/lag behavior: the dark blue (resp. dark red) time series *leads* the light blue (resp. light red) time series. While the first time series γ^a is periodic, it is clear that the second time series γ^b is not. However, we may classify both time series as cyclic: the time series are of a repetitive nature, although there may not be a fixed period. More formally, we say that a time series $\gamma : [0,1] \to \mathbb{R}^N$ is *cyclic* if it factors through the circle,

$$\gamma : [0,1] \xrightarrow{\psi} S^1 \to \mathbb{R}^N,$$

with ψ monotone, continuous and winding around the circle at least twice, where the winding condition enforces nontrivial repetition. While the two time series γ^a and γ^b may qualitatively look different, they are in fact related by a reparametrization. In other words, there exists a continuous, monotone increasing bijection $\phi : [0,1] \to [0,1]$ such that $\gamma^b = \gamma^a \circ \phi$.

Our goal is to detect the lead/lag behavior exhibited by the time series. A classical approach is to use the *unbiased cross-correlation* of a time series $\gamma = (\gamma_1, \gamma_2) : [0,1] \to \mathbb{R}^2$, which is defined by

$$r[\gamma](t_d) = \frac{1}{1-t_d} \int_0^1 (\gamma_1)_t (\gamma_2)_{t-t_d} dt, \tag{1.1.1}$$

where we define $\gamma_t = 0$ when $t \notin [0,1]$. In fact, this is the convolution of the two components of the time series γ normalized by the length of the overlap between the shifted time intervals. Applying the cross-correlation to both of our example time series γ^a and γ^b, we obtain the following.

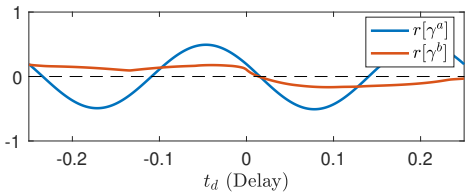


Figure 1.2: The cross correlation of the two components of the time series γ^a (in blue), nad γ^b (in red).

A large local maxima in the cross-correlation indicates the existence of lead/lag behavior, which can be clearly seen at $t_d \approx -0.05$ in the cross-correlation of γ^a. However, this peak is flattened out in the cross-correlation for γ^b, and this method is inconclusive.

The underlying reason why the cross-correlation fails to detect the lead/lag behavior in γ^b is fact that the cross-correlation is inherently *parametrization-dependent*, while the existence of lead/lag behavior is a *reparametrization-invariant* phenomenon. Indeed, one component of the time series leading another depends only on the *order* of events, and not necessarily on the exact parametrization. This suggests a reparametrization-invariant approach to time series analysis.

Instead of viewing time series as graphs of functions with respect to time, we can consider the multivariate time series as a single object: a path in $\mathbb{R}^N$. Because the parametrization does not affect the image of the path, both the time series γ^a and γ^b trace out the path in the following figure.

By considering geometric features of the resulting path, we can obtain reparametrization-invariant features for the underlying time series. One such feature which can detect lead/lag behavior is the *signed area* of a closed path. In the following discussion we will assume that $\gamma : [0,1] \to \mathbb{R}^2$ is a closed path that begins and ends at the origin, so that $\gamma_0 = \gamma_1 = (0,0)$. For arbitrary paths, we can form a closed path by connecting the end point and the initial point via linear interpolation. Furthermore, we can translate the path

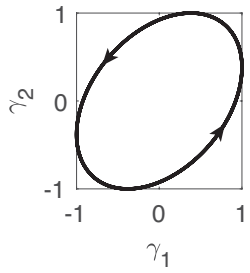


Figure 1.3: The time series γ^a and γ^b viewed as a path.

to begin and end at the origin, since geometric features such as signed area are invariant under translation.

The *signed area* of such a closed curve γ is defined to be

$$A(\gamma) = \int_{\mathbb{R}^2} \eta(\gamma; \mathbf{x}) d\mathbf{x}, \tag{1.1.2}$$

where $\eta(\gamma; \mathbf{x})$ is the *winding number* of γ about the point $\mathbf{x} \in \mathbb{R}^2$. Thus, the signed area is the area enclosed by the curve, taking into account the multiplicity as well as the orientation of winding.

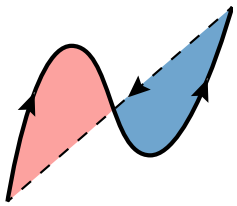


Figure 1.4: An example of the signed area. The solid black line represents the original curve, and the dashed black line is the linear interpolation between the end and initial points to create a closed curve. The area represented by the blue region is positive and the area represented by the red region is negative.

In the case of our example path in Figure 1.3, the winding number is $\eta(\gamma; \mathbf{x}) = 4$ if $\mathbf{x}$ is contained within the ellipse since it winds around 4 times (consider the original time series in Figure 1.1), and $\eta(\gamma; \mathbf{x}) = 0$ otherwise. The signed area can be reformulated using Stokes' theorem into a form which is more closely related to lead/lag behavior. In particular, the signed area can be computed by

$$A(\gamma) = \frac{1}{2} \int_0^1 (\gamma_1)_t (\gamma_2')_t - (\gamma_2)_t (\gamma_1')_t dt. \tag{1.1.3}$$

One criteria for the cyclic leading behavior can be formulated in terms of the components γ_1 and γ_2 as follows. We say that γ_1 *leads* γ_2 *locally at time* t if the following two conditions hold:

1. **positive influence of γ_1 on γ_2:** $(\gamma_1)_t(\gamma_2')_t > 0$ and
2. **negative influence of γ_2 on γ_1:** $(\gamma_2)_t(\gamma_1')_t < 0$.

The first condition says that γ_2 is changing in the same direction as the value of γ_1, while the second condition says that γ_1 is changing in the opposite direction as the value of γ_2: two properties to be expected when γ_1 leads γ_2. The signed area formula in Equation (1.1.3) is the integral of these two conditions, resulting in a *global* measure of the cyclic leading behavior over the entire path. Therefore, we say that γ_1 *leads* γ_2 if

$$A(\gamma) > 0.$$

It is clear that the signed area of our example path in Figure 1.3 is positive, and thus correctly indicates that the first component of the time series γ_1^a (resp. γ_1^b) leads the second component of the time series γ_2^a (resp. γ_2^b). As opposed to the cross-correlation, the signed area does not depend on the parametrization of the paths, and results in the same value for both example paths.

Our analysis thus far has been restricted to a fairly simplified scenario which includes only two components in the time series. In most real-world cases of time series analysis, we wish to understand the relationship between the components of a general multivariate time series $\gamma = (\gamma_1, \dots, \gamma_N) : [0,1] \to \mathbb{R}^N$. In this more general setting, we can consider the *lead matrix*, originally introduced by Baryshnikov and Schlafly in [BS16], which is a matrix of all the pairwise signed areas of the time series.

Definition 1.1.1. Let $\gamma = (\gamma_1, \dots, \gamma_N) : [0,1] \to \mathbb{R}^N$. The *lead matrix* of γ, denoted $A(\gamma)$, is an $N \times N$ matrix with entries

$$A(\gamma)_{i,j} \coloneqq \frac{1}{2} \int_0^1 \int_0^{t_2} (\gamma_i')_{t_1}(\gamma_j')_{t_2} - (\gamma_j')_{t_1}(\gamma_i')_{t_2} dt_1 dt_2. \tag{1.1.4}$$

Here, we have defined the signed area as an iterated line integral over the path, which reduces to the formula in Equation (1.1.3) when $\gamma_0 = 0$. Figure 1.5 shows an example lead matrix for a multivariate time series with 5 components, where we can clearly identify the pairwise lead/lag behavior between the components of the time series.

One important property of the lead matrix $A(\gamma)$ is that it is a real skew-symmetric matrix, which implies that its eigenvalues are purely imaginary and come in conjugate

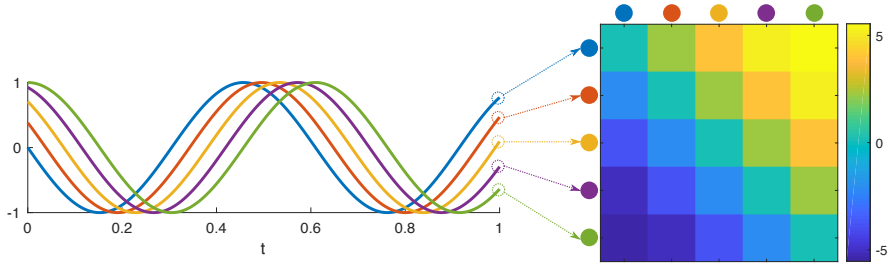


Figure 1.5: A multivariate time series with 5 components (left) along with its lead matrix (right).

pairs $\pm i\lambda_j$. Thus, if N is odd, we must have an odd number of zero eigenvalues. We will order the magnitudes of these eigenvalues from largest to smallest

$$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_{N/2} \geq 0.$$

The lead matrix cannot be diagonalized by real matrices, but it can be block-diagonalized using real matrices. In particular, if N is even, there exists a real orthogonal matrix $Q \in O(N)$, such that

$$A(\gamma) = Q\Lambda Q^{\intercal}, \tag{1.1.5}$$

where

$$\Lambda = \begin{pmatrix} 0 & \lambda_1 & 0 & 0 & \ldots & 0 & 0 \\ -\lambda_1 & 0 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 0 & \lambda_2 & \ldots & 0 & 0 \\ 0 & 0 & -\lambda_2 & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \ldots & 0 & \lambda_n \\ 0 & 0 & 0 & 0 & \ldots & -\lambda_n & 0 \end{pmatrix}. \tag{1.1.6}$$

If N is odd, the decomposition in Equation (1.1.5) still holds, but Λ has an additional row and column of zeros to account for the zero eigenvalue. Because $Q \in O(N)$, we can view Q as a matrix of orthonormal vectors $v_1, \ldots, v_N \in \mathbb{R}^N$.

The vectors v_1 and v_2 correspond to the block eigenvalue λ_1, and determine the plane on which γ achieves the maximum signed area. This is analogous to the diagonalization of covariance matrices performed in principal component analysis (PCA). In PCA, the eigenvector corresponding to the largest eigenvalue determines the direction of maximal variance. This shows that the lead matrix of a multivariate time series is a powerful tool

which can be used to study lead/lag phenomena in cyclic (and not necessarily periodic) time series [BS16; BG18; Zim+18; Sha+20].

Our approach of considering geometric properties of a path resulted in a reparametrization invariant feature set of the underlying multivariate time series. One may ask: are there more reparametrization-invariant features we can extract from the path? The answer is *yes*, and this leads to the central focus of this thesis: the *path signature*.

1.2 THE PATH SIGNATURE

The lead matrix in Definition 1.1.1 from the previous section was defined as a difference of *iterated integrals*,

$$\begin{aligned} A(\gamma)_{i,j} &= \frac{1}{2}\left(\int_0^1 \int_0^{t_2} (\gamma_i')_{t_1}(\gamma_j')_{t_2} dt_1 dt_2 - \int_0^1 \int_0^{t_2} (\gamma_j')_{t_1}(\gamma_i')_{t_2} dt_1 dt_2\right) \\ &= \frac{1}{2}\left(S^{(i,j)}(\gamma) - S^{(j,i)}(\gamma)\right). \end{aligned}$$

These iterated integrals, denoted $S^{(i,j)}(\gamma)$ and $S^{(j,i)}(\gamma)$, are components of an infinite-dimensional object called the *path signature*. The path signature is collection of reparametrization invariant features which completely characterize bounded variation paths up to *tree-like equivalence*. The path signature was originally developed by Chen [Che54; Che57; Che58] as a characterization of paths on manifolds, and was later generalized to define an iterated integral cochain algebra for path spaces and loop spaces [Che73; Che77]. We discuss the cochain algebra in Chapter 2, and this perspective forms the foundation of the generalization to mapping spaces, which is developed in Chapter 5 and Chapter 7.

More recently, the path signature was used to establish the definition of *rough paths*, a deterministic approach to the integration of highly irregular paths [Lyo98], which has led to influential developments in stochastic analysis [LQ07; LCL07; FH20; FV10]. While we do not directly work with rough paths in this thesis, we provide a brief exposition of the fundamental ideas in Section 1.4, as this is one of the primary motivations for the generalization of the signature to mapping spaces in Chapter 5 and Chapter 7.

The powerful properties of the path signature have also been exploited for time series analysis in machine learning [Lyo14; CK16; Mor+21]. This is the main application of signatures that we consider in this thesis, and we introduce the main machine learning results in Section 1.3.

In this section, we will introduce the path signature along with all of its standard properties. We will provide definitions for paths on an arbitrary Banach space V, but we will also discuss the finite-dimensional case $V = \mathbb{R}^N$, in which formulas can be more explicitly stated. We will begin by defining bounded p-variation paths, and our main focus

will be on bounded 1-variation paths. Next, since the path signature is a map which takes paths to a formal power series of tensors, we must define the Banach space of formal power series. Once these preliminary constructions are defined, we move on to discussing the path signature and its properties.

1.2.1 *Bounded p-Variation Paths*

Definition 1.2.1. Suppose $(V, \|\cdot\|)$ is a Banach space. Let $\gamma : [0,1] \to V$. The *p-variation of* γ is defined to be

$$|\gamma|_{p-var} := \sup_{\Pi} \left(\sum_{i=1}^{n} \|\gamma_{t_i} - \gamma_{t_{i-1}}\|^p \right)^{1/p} \tag{1.2.1}$$

where $\Pi = \{0 = t_0 < t_1 < \ldots < t_{n-1} < t_n = 1\}$ is a partition of the interval $[0,1]$, and the supremum is taken over all possible partitions. The space of all *bounded p-variation paths* is denoted by

$$C^{p-var}([0,1], V) := \{\gamma : [0,1] \to V \,:\, |\gamma|_{p-var} < \infty\}. \tag{1.2.2}$$

Furthermore, if we require all paths to begin at a specified base point $v \in V$, we will denote the space by $C_v^{1-var}([0,1], V)$. Unless otherwise specified, we will use the unit interval as the domain, and thus simplify the notation as

$$C^{1-var}(V) := C^{1-var}([0,1], V), \qquad C_v^{1-var}(V) := C_v^{1-var}([0,1], V).$$

Proposition 1.2.1 ([CG98])**.** *Let $(V, \|\cdot\|)$ be a Banach space and $p \geq 1$. Then $C^{p-var}([0,1], V)$ equipped with the norm*

$$\|\gamma\|_p := |\gamma_0| + |\gamma|_{p-var}$$

is a Banach space.

1.2.2 *The Tensor Algebra*

Given a Banach space $(V, \|\cdot\|)$, we denote the *algebraic tensor products* of V by

$$V^{\otimes_a k} := V \otimes_a \ldots \otimes_a V,$$

where $\otimes_a$ denotes the tensor product of the underlying vector space. There is a natural action of the symmetric group Σ_k on $V^{\otimes_a k}$ defined for $\sigma \in \Sigma_k$ by

$$\sigma \cdot (v_1 \otimes \dots v_k) = v_{\sigma(1)} \otimes \dots \otimes v_{\sigma(k)}. \tag{1.2.3}$$

Further, suppose that we are given a sequence of tensor norms $\|\cdot\|_k$ on $V^{\otimes_a k}$ which satisfy the following two conditions:

1. for all $n, m \geq 1$, we have

   $$\|a \otimes_a b\|_{n+m} \leq \|a\|_n \|b\|_m$$

   for all $a \in V^{\otimes_a n}$ and $b \in V^{\otimes_a m}$;

2. for all $n \geq 1$, the symmetric group Σ_n acts by isometries on $V^{\otimes a}$, so that

   $$\|\sigma a\|_n = \|a\|_n$$

   for all $a \in V^{\otimes_a n}$.

A sequence of tensor norms $\{\|\cdot\|_k\}_{k=1}^{\infty}$ which satisfies these two conditions are called *admissible norms* in [LCL07]. The completion of $V^{\otimes_a k}$ under the norm $\|\cdot\|_k$ is a Banach space, which we denote by $(V^{\otimes k}, \|\cdot\|_k)$. If $V = \mathbb{R}^N$ is finite dimensional, the algebraic tensor product and the completed tensor product is equivalent $V^{\otimes_a k} \cong V^{\otimes k}$. Furthermore, the action of the symmetric group also extends to the completion. By the above compatibility condition, we have

$$V^{\otimes_a n+m} \subset V^{\otimes n} \otimes_a V^{\otimes m} \subset V^{\otimes n+m},$$

and therefore, the completion of $V^{\otimes n} \otimes_a V^{\otimes m}$ can be written as

$$V^{\otimes n} \otimes V^{\otimes m} \cong V^{\otimes n+m}.$$

A *formal power series of tensors in* V is an element

$$\mathbf{s} = (\mathbf{s}_k)_{k=0}^{\infty} \in \prod_{k=0}^{\infty} V^{\otimes k}$$

which consist of a sequence of elements $\mathbf{s}_0 \in \mathbb{R}$ and $\mathbf{s}_k \in V^{\otimes k}$, for $k \geq 1$. Addition and scalar multiplication are determined degreewise, so that for $\mathbf{s}, \mathbf{t} \in T((V))$ and $\lambda \in \mathbb{R}$,

$$(\mathbf{s} + \mathbf{t})_k = \mathbf{s}_k + \mathbf{t}_k, \qquad (\lambda \mathbf{s})_k = \lambda \mathbf{s}_k, \tag{1.2.4}$$

and multiplication is defined by tensor multiplication

$$(\mathbf{s} \otimes \mathbf{t})_k = \sum_{m=0}^{\infty} \mathbf{s}_m \otimes \mathbf{t}_{k-m}. \tag{1.2.5}$$

We define a norm on $\prod_{k=0}^{\infty} V^{\otimes k}$ by

$$\|\mathbf{s}\| = \sqrt{\sum_{k=1}^{\infty} \|\mathbf{s}_k\|_k^2}, \tag{1.2.6}$$

which may be infinite for some terms. Note that there are several choices for the extension of the tensor norms to a norm for the formal power series [LCL07]. If V is a Hilbert space, our choice of power series norm coincides with the Hilbert space norm. The algebraic structure we are interested in is the following Banach space of finite norm group-like power series

$$T((V)) \coloneqq \{\mathbf{s} \in \prod_{k=0}^{\infty} V^{\otimes k} \,:\, \mathbf{s}_0 = 1,\, \|\mathbf{s}\| < \infty\}. \tag{1.2.7}$$

We will simply call this the *tensor algebra*. The condition that $\mathbf{s}_0 = 1$ ensures the existence of an inverse, and thus provides $T((V))$ with a group structure (hence the terminology *group-like elements*), and the finite norm condition $\|\mathbf{s}\| < \infty$ ensures that the space is a Banach space. Moreover, the *truncated tensor algebra at degree M* is the Banach space

$$T^{(M)}(V) \coloneqq \{\mathbf{s} \in \prod_{k=0}^{M} V^{\otimes k} \,:\, \mathbf{s}_0 = 1\},$$

where the finite norm condition is immediately satisfied since there are only finitely many terms in the series. Further details on the construction of this tensor algebra can be found in [LQ07; Boe+15].

1.2.3 *The Path Signature*

Definition 1.2.2. Let $(V, \|\cdot\|)$ be a Banach space. Let $\gamma \in C^{1-var}(V)$ be a bounded variation path, and $m \geq 1$. The *path signature of γ at level m*, $S_m : C^{1-var}(V) \to V^{\otimes m}$, is defined to be

$$S_m(\gamma) \coloneqq \int_{\Delta^m(0,1)} d\gamma_{t_1} \otimes d\gamma_{t_2} \otimes \ldots \otimes d\gamma_{t_m}, \tag{1.2.8}$$

where $\Delta^m(a,b) := \{a \leq t_1 < \ldots < t_m \leq b\}$. This is an integral in the Riemann-Stieltjes sense. By combining the information across all levels, we obtain the *path signature of* γ, denoted as

$$S : C^{1-var}(V) \to T((V)), \tag{1.2.9}$$

and defined by

$$(S(\gamma))_0 = 1, \qquad (S(\gamma))_j = S_j(\gamma),\ j \geq 1.$$

In order to understand the computation of this integral, it is instructive to use a recursive definition of the path signature. By adapting Equation (1.2.8), we define

$$S_1(\gamma)_{a,b} := \int_{\Delta^1(a,b)} d\gamma_t = \int_a^b d\gamma_t. \tag{1.2.10}$$

Next, we assume that $S_{m-1}(\gamma)_{a,b}$ is defined, and we take

$$\begin{aligned} S_m(\gamma)_{a,b} &= \int_a^b S_{m-1}(\gamma)_{a,t} \otimes d\gamma_t \\ &= \lim_{|\Pi(a,b)|\to 0} \sum_{i=1}^{r} S_{m-1}(\gamma)_{a,s_{i-1}} \otimes \left(\gamma_{s_i} - \gamma_{s_{i-1}}\right), \end{aligned} \tag{1.2.11}$$

where the limit is taken over all partitions $\Pi(a,b) = \{a = s_0 < \ldots < s_r = b\}$ of $[a,b]$, and the mesh size is $|\Pi(a,b)| = \sup_{i=1,\ldots,r}(s_i - s_{i-1})$. Then, the path signature at level m is

$$S_m(\gamma) := S_m(\gamma)_{0,1}.$$

In fact, this recursive definition shows that the signature is the solution of a controlled differential equation.

Proposition 1.2.2. *Let $\gamma \in C^{1-var}(V)$. The path signature is the solution to the differential equation*

$$\frac{dS}{dt} = S(t) \otimes \frac{d\gamma}{dt}. \tag{1.2.12}$$

We show that the norm of the path signature is finite, so that its codomain is indeed $T((V))$, which consists of finite norm power series.

Proposition 1.2.3. *For any $\gamma \in C^{1-var}(V)$, $\|S(\gamma)\| < \infty$.*

Proof. We claim that at each level m, the norm of the level m signature can be bounded by

$$\|S_m(\gamma)\|_m \leq \frac{|\gamma|_{1-var}}{m!}, \tag{1.2.13}$$

which we will prove by induction. At level $m = 1$, we have

$$\|S_1(\gamma)\|_1 = \|\gamma_1 - \gamma_0\|_1 \leq |\gamma|_{1-var}.$$

Now, suppose Equation (1.2.13) holds for level $m-1$. Using the recursive definition of the signature in Equation (1.2.11), we have

$$\begin{aligned}\|S_m(\gamma)\|_m &= \left\| \int_0^1 S_{m-1}(\gamma)_{0,t} \otimes d\gamma_t \right\|_m \\ &\leq \int_0^1 \|S_{m-1}(\gamma)_{0,t}\| \cdot \|d\gamma_t\|.\end{aligned}$$

Note that $S_{m-1}(\gamma)_{0,t}$ is the level $m-1$ signature of the restricted path $\gamma|_{[0,t]}$. Thus, by the induction hypothesis, we have

$$\|S_{m-1}(\gamma)_{0,t}\| \leq \frac{\left|\gamma|_{[0,t]}\right|_{1-var}^{m-1}}{(m-1)!}.$$

Applying this to the above integral, and performing a change of variable, we get

$$\begin{aligned}\|S_m(\gamma)\|_m &\leq \int_0^1 \frac{\left|\gamma|_{[0,t]}\right|_{1-var}^{m-1}}{(m-1)!} \cdot \|d\gamma_t\| \\ &\leq \int_0^{|\gamma|_{1-var}} \frac{s^{m-1}}{(m-1)!} ds \\ &= \frac{|\gamma|_{1-var}^m}{m!}.\end{aligned}$$

Now, we can bound the norm of the full signature,

$$\|S(\gamma)\|^2 = \sum_{m=1}^{\infty} \|S(\gamma)\|_m^2 \leq \sum_{m=1}^{\infty} \frac{|\gamma|_{1-var}^{2m}}{(m!)^2} \leq \sum_{m=1}^{\infty} \frac{|\gamma|_{1-var}^{2m}}{m!} \leq \exp(|\gamma|_{1-var}^2).$$

□

In computational settings, we will also work with truncated path signatures.

Definition 1.2.3. Let $\gamma \in C^{1-var}(V)$, and $m \geq 1$. The *truncated path signature of γ at level m*, $S_{\leq m} : C^{1-var}(V) \to T^{(m)}(V)$, is defined by

$$(S_{\leq m}(\gamma))_0 = 1, \qquad (S_{\leq m}(\gamma))_j = S_j(\gamma),\ 1 \leq j \leq m. \tag{1.2.14}$$

Finally, we will also give a separate definition in the finite-dimensional setting of $V = \mathbb{R}^N$, which allows for a more concrete definition.

Definition 1.2.4. Let $V = \mathbb{R}^N$ with an ordered basis $(e_1, \dots, e_N)$, and $\gamma = (\gamma_1, \dots, \gamma_N) \in C^{1-var}(V)$. Let $I = (i_1, \dots, i_m)$ be a multi-index valued in $[N] = \{1, \dots, N\}$, so that $i_j \in [N]$. The *path signature of γ with respect to I is*

$$S^I(\gamma) \coloneqq \int_{\Delta^m(0,1)} (\gamma'_{i_1})_{t_1} \dots (\gamma'_{i_m})_{t_m} dt_1 \dots dt_m. \tag{1.2.15}$$

The tensor product $V^{\otimes k}$ has a canonical basis $\{e_I\}$, where $I = (i_1, \dots, i_m)$ is a multi-index valued in $[N]$. The *path signature of γ* is defined by

$$S(\gamma) \coloneqq 1 + \sum_{m=1}^{\infty} \sum_{|I|=m} S^I(\gamma)\, e_I. \tag{1.2.16}$$

The finite dimensional definition coincides with the Banach space definition in Definition 1.2.2, but the Banach space definition is for the entire level m signature, while the finite dimensional definition is term-by-term. The following proposition provides three basic properties of the path signature.

Proposition 1.2.4. *Suppose $\gamma \in C^{1-var}(V)$, $\phi : [0,1] \to [0,1]$ a continuous monotone increasing bijection, and $a \in \mathbb{R}^N$. For $\lambda > 0$, we define the multilinear scaling operator*

$$\begin{aligned} \sigma_\lambda : T((V)) &\to T((V)) \\ (\mathbf{s}_m)_{m=0}^{\infty} &\mapsto (\lambda^m \mathbf{s}_m)_{m=0}^{\infty}. \end{aligned} \tag{1.2.17}$$

The path signature satisfies the following three properties.

1. ***Translation invariance:***

$$S(\gamma + a) = S(\gamma). \tag{1.2.18}$$

2. ***Reparametrization invariance:***

$$S(\gamma \circ \phi) = S(\gamma). \tag{1.2.19}$$

3. ***Multilinear scaling:***

$$S(\lambda\gamma) = \sigma_\lambda S(\gamma). \tag{1.2.20}$$

Proof. All three properties are straightforward to show using the definition of path signatures. Translation invariance is due to fact that the path signature depends only on the derivatives of the path (or the fact that Riemann-Stieltjes integrals are translation-invariant with respect to the integrator). Reparametrization invariance of the first level is given by

$$S_1(\gamma \circ \phi) = \int_0^1 d(\gamma \circ \phi)_t = \int_0^1 d\gamma_t = S_1(\gamma),$$

since the Riemann-Stieltjes integral is reparametrization-invariant. Then assuming invariance holds for level $m-1$, and using the recursive definition for the signature from Equation (1.2.11),

$$S_m(\gamma \circ \phi) = \int_0^1 S_{m-1}(\gamma \circ \phi)_{0,t} \otimes d(\gamma \circ \phi)_t = \int_0^1 S_{m-1}(\gamma)_{0,t} \otimes d\gamma_t = S_m(\gamma).$$

Finally, the multilinear scaling property is clear for level $m = 1$, and we assume it holds for level $m-1$. Then using the recursive definition of the signature again,

$$S_m(\lambda\gamma) = \int_0^1 S_{m-1}(\lambda\gamma)_{0,t} \otimes d(\lambda\gamma)_t = \lambda^m \int_0^1 S_{m-1}(\gamma)_{0,t} \otimes d\gamma_t = \lambda^m S_m(\gamma).$$

□

Remark 1.2.1. Path signatures can be defined for paths with an arbitrary closed interval $[a, b] \subset \mathbb{R}$ as a domain. However, without loss of generality due to reparametrization invariance, we only consider paths defined on $[0, 1]$.

These path signatures characterize classes of paths in $\mathbb{R}^N$ up to a *tree-like equivalence*, originally defined using a reduction procedure in [Che58] for piecewise regular paths, and later defined in its modern formulation in [HL10] for bounded variation paths. In order to define the relation, we first consider concatenation of paths. Suppose $\alpha, \beta \in C^{1-var}(V)$, then define the concatenation of the two paths, $\alpha * \beta \in C^{1-var}(V)$ by

$$(\alpha * \beta)_t = \begin{cases} \alpha_{2t} & : t \in [0, \frac{1}{2}) \\ (\alpha_1 - \beta_0 + \beta_{2t-1} & : t \in [\frac{1}{2}, 1] \end{cases} \tag{1.2.21}$$

The inverse of a path $\gamma \in C^{1-var}$ is defined to be the same path but running in the opposite direction, namely

$$(\gamma^{-1})_t = \gamma_{1-t}. \tag{1.2.22}$$

Definition 1.2.5 ([HL10]). A path $\gamma \in C^{1-var}(V)$ is a *tree-like path* in $\mathbb{R}^N$ if there exists some positive real-valued continuous function h defined on $[0,1]$ such that $h(0) = h(1) = 0$ and such that

$$\|\gamma_t - \gamma_s\| \leq h(s) + h(t) - 2 \inf_{u \in [s,t]} h(u), \tag{1.2.23}$$

where $\|\cdot\|$ is the Euclidean norm on $\mathbb{R}^N$. The function h is called a *height function* for γ and if h is of bounded variation, then γ is a *Lipschitz tree-like path*.

Definition 1.2.6. Two paths $\alpha, \beta \in C^{1-var}(V)$ are *tree-like equivalent*, denoted by $\alpha \sim_t \beta$, if $\alpha * \beta^{-1}$ is a Lipschitz tree-like path.

Lemma 1.2.1. *Tree-like equivalence is an equivalence relation.*

Proposition 1.2.5. *The tree-like equivalence classes of paths*

$$\widetilde{C}^{1-var}(V) := C^{1-var}(V)/\sim_t \tag{1.2.24}$$

is a group under concatenation, where the inverse is given by Equation (1.2.22).

Now we state the characterization theorem, which was proved by Chen [Che58] for irreducible piecewise regular continuous paths, and generalized in [HL10] to bounded variation paths on a finite dimensional space $V = \mathbb{R}^N$. It was then further generalized to rough paths on an arbitrary Banach space in [Boe+16]. The following statement is a special case of the most general result in [Boe+16].

Theorem 1.2.1. *Suppose $\alpha, \beta \in C^{1-var}(V)$. Then $S(\alpha) = S(\beta)$ if and only if $\alpha \sim_t \beta$.*

In fact, this statement is even stronger when we consider the algebraic structures of the group of equivalence classes $\widetilde{C}^{1-var}(V)$ and the tensor algebra $T((V))$. The following result is a fundamental property of the path signature, which is crucial to the definition of rough paths. This was originally proved in [Che54] and is called *Chen's identity*.

Theorem 1.2.2. *Suppose $\alpha, \beta \in C^{1-var}(V)$. Then*

$$S(\alpha * \beta) = S(\alpha) \otimes S(\beta). \tag{1.2.25}$$

Proof. We begin by noting that it suffices to prove this property level-wise

$$S_m(\alpha * \beta) = \sum_{j=0}^{m} S_j(\alpha) \otimes S_{m-j}(\beta). \tag{1.2.26}$$

This can be proved using the following decomposition of the m-simplex:

$$\Delta^m(0,1) = \coprod_{j=0}^{m} \Delta^j(0,1/2) \times \Delta^{m-j}(1/2,1). \tag{1.2.27}$$

Let $\gamma = \alpha * \beta$ to simplify notation. Then, decomposing the domain of integration in the definition of the signature from Equation (1.2.8), we get

$$\begin{aligned}
S_m(\gamma) &= \int_{\Delta^m(0,1)} d\gamma_{t_1} \otimes \ldots \otimes d\gamma_{t_m} \\
&= \sum_{j=0}^{m} \int_{\Delta^j(0,1/2)\times\Delta^{m-j}(1/2,1)} d\gamma_{t_1} \otimes \ldots \otimes d\gamma_{t_m} \\
&= \sum_{j=0}^{m} \int_{\Delta^j(0,1)\times\Delta^{m-j}(0,1)} d\alpha_{t_1} \otimes \ldots \otimes d\alpha_{t_j} \otimes d\beta_{t_{j+1}} \otimes \ldots \otimes d\beta_{t_m} \\
&= \sum_{j=0}^{m} \left(\int_{\Delta^j(0,1)} d\alpha_{t_1} \otimes \ldots \otimes d\alpha_{t_j} \right) \otimes \left(\int_{\Delta^{m-j}(0,1)} d\beta_{t_{j+1}} \otimes \ldots \otimes d\beta_{t_m} \right) \\
&= \sum_{j=0}^{m} S_j(\alpha) \otimes S_{m-j}(\beta),
\end{aligned}$$

where we use the definition of concatenation and a change of variables in the third line. □

Then, combining Theorem 1.2.1 and Theorem 1.2.2, we obtain the following theorem.

Theorem 1.2.3. *The signature map $S : \widetilde{C}^{1-var}(V) \to T((V))$ is an injective group homomorphism.*

In the finite dimensional setting of $V = \mathbb{R}^N$, this theorem implies the collection of signatures with respect to all multi-indices I of $[N]$ provide a complete set of features for paths up to tree-like equivalence. In the general setting, where V is an arbitrary Banach space, we define

$$L_V := \bigoplus_{m=0}^{\infty} (V')^{\otimes m}, \tag{1.2.28}$$

where V' denotes the topological dual of V. The space L_V is identified with a dense subset of the topological dual of the tensor algebra $T((V))'$ by setting

$$\langle \boldsymbol{\ell}, \mathbf{t} \rangle := \sum_{m=0}^{\infty} \langle \boldsymbol{\ell}_m, \mathbf{t}_m \rangle, \tag{1.2.29}$$

where $\boldsymbol{\ell} \in L_V$ and $\mathbf{t} \in T((V))$. Note that the $\langle \boldsymbol{\ell}, \mathbf{t} \rangle$ is simply the notation for $\boldsymbol{\ell}(\mathbf{t})$; since V is an arbitrary Banach space, there doesn't necessarily exist an inner product. However, this notation coincides with the setting where V is a Hilbert space, in which case $V' \cong V$, and can be interpreted as an inner product. Thus, we can similarly view the elements of L_V as feature maps for paths using the above pairing, and the collection of feature maps over L_V also provides a complete set of tree-like invariant features.

In fact, as functions on the path space, the elements of L_V are equipped with an algebraic structure.

Definition 1.2.7. Let p and q be non-negative integers. A (p, q)-*shuffle* is a permutation $\sigma \in \Sigma_{p+q}$ such that

$$\sigma^{-1}(1) < \sigma^{-1}(2) < \ldots < \sigma^{-1}(p)$$

and

$$\sigma^{-1}(p+1) < \sigma^{-1}(p+2) < \ldots < \sigma^{-1}(p+q).$$

We denote by $Sh(p, q)$ the set of (p, q)-shuffles.

Let $N \in \mathbb{N}$. Given two finite multi-indices $I = (i_1, \ldots, i_p)$ and $J = (j_1, \ldots, j_q)$ of N, let $R = (r_1, \ldots, r_p, r_{p+1}, \ldots r_{p+q}) = (i_1, \ldots, i_p, j_1, \ldots, j_q)$ be the concatenated multi-index. The *shuffle product* of I and J is defined to be the multiset of multi-indices

$$I \sqcup\!\sqcup J = \left\{ \left(r_{\sigma(1)}, \ldots r_{\sigma(p+q)} \right) \mid \sigma \in Sh(p, q) \right\}.$$

As an example, suppose $I = (1, 2)$ and $J = (2, 3)$. Then

$$I \sqcup\!\sqcup J = \{(1,2,2,3), (1,2,2,3), (2,1,2,3), (1,2,3,2), (2,1,3,2), (2,3,1,2)\}.$$

There exists a shuffle product structure on the path signature.

Proposition 1.2.6. *Suppose $\gamma \in C^{1-var}(V)$. Then,*

$$S_p(\gamma) \otimes S_q(\gamma) = \sum_{\sigma \in Sh(p,q)} \sigma^{-1} \cdot S_{p+q}(\gamma), \tag{1.2.30}$$

where $\sigma^{-1} \in \Sigma_{p+q}$ acts by permutation on $V^{\otimes(p+q)}$ on the right hand side.

Proof. The proof of this result relies on the following decomposition of the product of simplices

$$\Delta^p \times \Delta^q = \coprod_{\sigma \in Sh(p,q)} \Delta_\sigma^{p+q}, \tag{1.2.31}$$

where

$$\Delta_\sigma^{p+q} = \{(t_1, \dots, t_{p+q}) \,:\, 0 \leq t_{\sigma(1)} \leq \dots t_{\sigma(p+q)} \leq 1\}. \tag{1.2.32}$$

Applying this to the definition to the tensor product of the level p and level q signatures, we get

$$\begin{aligned} S_p(\gamma) \otimes S_q(\gamma) &= \left(\int_{\Delta^p} d\gamma_{t_1} \otimes \dots d\gamma_{t_p} \right) \otimes \left(\int_{\Delta^q} d\gamma_{t_1} \otimes \dots d\gamma_{t_q} \right) \\ &= \sum_{\sigma \in Sh(p,q)} \int_{\Delta_\sigma^{p+q}} d\gamma_{t_1} \otimes \dots \otimes d\gamma_{t_{p+q}} \\ &= \sum_{\sigma \in Sh(p,q)} \int_{\Delta^{p+q}} d\gamma_{t_{\sigma^{-1}(1)}} \otimes \dots \otimes d\gamma_{t_{\sigma^{-1}(p+q)}} \\ &= \sum_{\sigma \in Sh(p,q)} \sigma^{-1} \cdot S_{p+q}(\gamma), \end{aligned}$$

where we apply the change of variables $t_{\sigma^{-1}(j)} \mapsto t_j$ in the third line in order to transform the domain of integration from Δ_σ^{p+q} to Δ^{p+q}. □

This leads to an algebraic structure on the space of functions L_V.

Corollary 1.2.1. *Suppose* $\mathbf{f} \in (V')^{\otimes p}$ *and* $\mathbf{g} \in (V')^{\otimes q}$*, and we treat both* $\mathbf{f}$ *and* $\mathbf{g}$ *as elements of* L_V*. Let* $\gamma \in C^{1-var}(V)$*. Then,*

$$\langle \mathbf{f}, S(\gamma) \rangle \cdot \langle \mathbf{g}, S(\gamma) \rangle = \sum_{\sigma \in Sh(p,q)} \langle \sigma \cdot (\mathbf{f} \otimes \mathbf{g}), S(\gamma) \rangle. \tag{1.2.33}$$

Thus, L_V *is a subalgebra of the algebra of continuous functions* $C(C^{1-var}(V), \mathbb{R})$*.*

Proof. First, since $\mathbf{f} \in (V')^{\otimes p}$ and $\mathbf{g} \in (V')^{\otimes q}$, we can simplify the expression on the left by

$$\begin{aligned} \langle \mathbf{f}, S(\gamma) \rangle \cdot \langle \mathbf{g}, S(\gamma) \rangle &= \langle \mathbf{f}, S_q(\gamma) \rangle \cdot \langle \mathbf{g}, S_q(\gamma) \rangle \\ &= \langle \mathbf{f} \otimes \mathbf{g}, S_p(\gamma) \otimes S_q(\gamma) \rangle. \end{aligned}$$

Then, we apply the shuffle product of the signature in Equation (1.2.30) to get

$$\begin{aligned}\langle \mathbf{f} \otimes \mathbf{g}, S_p(\gamma) \otimes S_q(\gamma) \rangle &= \left\langle \mathbf{f} \otimes \mathbf{g}, \sum_{\sigma \in Sh(p,q)} \sigma^{-1} \cdot S_{p+q}(\gamma) \right\rangle \\ &= \sum_{\sigma \in Sh(p,q)} \langle \sigma \cdot (\mathbf{f} \otimes \mathbf{g}), S_{p+q}(\gamma) \rangle,\end{aligned}$$

where $\langle \sigma \cdot \mathbf{s}, \mathbf{t} \rangle = \langle \mathbf{s}, \sigma^{-1} \cdot \mathbf{t} \rangle$ for $\mathbf{s} \in (V')^{\otimes n}$ and $\mathbf{t} \in V^{\otimes n}$ since the tensor product of linear functionals act component-wise. □

In the finite dimensional setting of $V = \mathbb{R}^N$, the product structure can be made more explicit. In this setting, $V' \cong V$, and thus L_V is the space of polynomials of tensors. In particular, given $e_I = e_{i_1} \otimes \ldots \otimes e_{i_m}$ for some multi-index I of $[N]$, the functional corresponding to $e_I \in (V')^{\otimes m} \subset L_V$ is

$$\langle e_I, S(\gamma) \rangle = S^I(\gamma). \tag{1.2.34}$$

Therefore, the product structure in Corollary 1.2.1 can be restated as a product structure on signatures with respect to specific multi-indices.

Corollary 1.2.2. *Suppose $V = \mathbb{R}^N$. Let $I = (i_1, \ldots, i_p)$ and $J = (j_1, \ldots, j_q)$ be multi-indices valued in $[N]$. Then for any $\gamma \in C^{1-var}(V)$, we have*

$$S^I(\gamma) S^J(\gamma) = \sum_{P \in I \sqcup\!\sqcup J} S^P(\gamma). \tag{1.2.35}$$

Proof. Taking $\mathbf{f} = e_I$ and $\mathbf{g} = e_J$ in Corollary 1.2.1, we have

$$\begin{aligned}\langle e_I, S(\gamma) \rangle \cdot \langle e_J, S(\gamma) \rangle &= \sum_{\sigma \in Sh(p,q)} \langle \sigma \cdot (e_I \otimes e_J), S(\gamma) \rangle \\ S^I(\gamma) S^J(\gamma) &= \sum_{P \in I \sqcup\!\sqcup J} S^P(\gamma).\end{aligned}$$

□

The last property of the path signature that we consider in this section is continuity. We begin by computing a bound for the modulus of continuity for the level m-signature. The proof given here is from [FV10], which states the result for a finite-dimensional V, but their proof also holds in the general setting.

Proposition 1.2.7. *Let $\alpha, \beta \in C^{1-var}(V)$ and $\ell \geq \max\{|\alpha|_{1-var}, |\beta|_{1-var}\}$. For all $m \geq 1$,*

$$\|S_m(\alpha) - S_m(\beta)\|_m \leq E_{m-1} \ell^{m-1} |\alpha - \beta|_{1-var}, \tag{1.2.36}$$

where

$$E_m := \sum_{0=1}^{m} \frac{1}{j!}. \tag{1.2.37}$$

Proof. We will prove this by induction. At level $m = 1$, we have

$$\begin{aligned}\|S_1(\alpha) - S_1(\beta)\|_1 &= \left\| \int_0^1 d(\alpha - \beta)(t) \right\|_1 \\ &\leq |\alpha - \beta|_{1-var}.\end{aligned}$$

Now, suppose the result holds at level $m-1$. Using the recursive definition of the signature in Equation (1.2.11), we have

$$\begin{aligned}&S_m(\alpha) - S_m(\beta) \\ &\quad = \int_0^1 (S_{m-1}(\alpha)_{0,t} - S_{m-1}(\beta)_{0,t}) \otimes d\alpha_t + \int_0^1 S_{m-1}(\beta)_{0,t} \otimes d(\alpha - \beta)(t).\end{aligned}$$

By the induction hypothesis,

$$\|S_{m-1}(\alpha)_{0,t} - S_{m-1}(\beta)\|_{m-1} \leq E_{m-2}\ell^{m-2}|\alpha - \beta|_{1-var}$$

holds for all $t \in [0,1]$. Then, we can bound the first integral as

$$\left\| \int_0^1 (S_{m-1}(\alpha)_{0,t} - S_{m-1}(\beta)_{0,t}) \otimes d\alpha_t \right\|_m \leq E_{m-2}\ell^{m-1}|\alpha - \beta|_{1-var}.$$

For the second integral, we use the bound from Equation (1.2.13) to get

$$\left\| \int_0^1 S_{m-1}(\beta)_{0,t} \otimes d(\alpha - \beta)(t) \right\| \leq \frac{1}{m-1}\ell^{m-1}|\alpha - \beta|_{1-var}.$$

Combining the bounds on these two integrals, we obtain the desired inequality.

$$\begin{aligned}\|S_m(\alpha) - S_m(\beta)\|_m &\leq E_{m-2}\ell^{m-1}|\alpha - \beta|_{1-var} + \frac{1}{m-1}\ell^{m-1}|\alpha - \beta|_{1-var} \\ &= E_{m-1}\ell^{m-1}|\alpha - \beta|_{1-var}.\end{aligned}$$

□

This result shows that the truncated path signature is Lipschitz continuous for paths with a fixed maximum length.

Corollary 1.2.3. *Let $\ell > 0$ be a constant such that $\ell \neq 1$ and $\alpha, \beta \in C^{1-var}(V)$ be bounded variation paths such that $|\alpha|_{1-var}, |\beta|_{1-var} \leq \ell$. For all $m \geq 1$,*

$$\|S_{\leq m}(\alpha) - S_{\leq m}(\beta)\| \leq \exp(1) \left(\frac{1-\ell^m}{1-\ell}\right)^{1/2} |\alpha - \beta|_{1-var}. \tag{1.2.38}$$

Proof. Applying the previous proposition,

$$\begin{aligned}
\|S_{\leq m}(\alpha) - S_{\leq m}(\alpha)\|^2 &= \sum_{j=1}^{m} \|S_j(\alpha) - S_j(\beta)\|_j^2 \\
&\leq \sum_{j=1}^{m} E_{m-1}^2 \ell^{m-1} |\alpha - \beta|_{1-var}^2 \\
&\leq \exp(1) |\alpha - \beta|_{1-var} \sum_{j=1}^{m} \ell^{m-1} \\
&= \exp(2) |\alpha - \beta|_{1-var}^2 \left(\frac{1-\ell^m}{1-\ell}\right).
\end{aligned}$$

Then, we get the desired bound by taking the square root. □

In fact, one can show that the signature map is continuous, though it requires a more detailed analysis. The following theorem is a special case of Theorem 3.1.3 in [LQ07], which provides a continuity result for rough paths.

Theorem 1.2.4. *Let $\alpha, \beta \in C^{1-var}(V)$ and β be a constant such that*

$$\beta \geq 2\left(1 + \sum_{r=3}^{\infty} \left(\frac{2}{r-2}\right)^2\right).$$

If ℓ is a constant such that

$$|\alpha|_{1-var}, |\beta|_{1-var} \leq \frac{\ell}{\beta}, \quad \text{and} \quad |\alpha - \beta|_{1-var} \leq \epsilon \frac{\ell}{\beta}$$

for some $\epsilon > 0$. Then for all $m \geq 1$,

$$\|S_m(\alpha) - S_m(\beta)\|_m \leq \epsilon \frac{\ell^m}{\beta(m!)}.$$

This implies that

$$\|S(\alpha) - S(\beta)\| \leq \frac{\epsilon}{\beta} \exp(\ell^2/2).$$

In particular, this shows that the path signature is continuous.

1.3 THE SIGNATURE KERNEL

In this section, we will discuss the application of the path signature for time series analysis in machine learning in the context of kernel methods. Suppose $\mathcal{X}$ is a topological space of data that we wish to study. A standard method in machine learning is to first apply a feature map, which is a continuous function

$$\Phi : \mathcal{X} \to V \tag{1.3.1}$$

into a topological vector space V. The feature map allows us to extract relevant properties of the data, and also to transform the data into elements of a linear space which can be more easily manipulated computationally. Feature maps which are universal and characteristic (see Definition 1.3.2) provide theoretical guarantees for problems related to learning both functions and measures on $\mathcal{X}$. However, a vector space V rich enough to provide such features may itself be unavoidably infinite dimensional, in which case we have not mitigated the computational issues. When V is a Hilbert space, the theory of *kernel methods* in machine learning allow us to circumvent this difficulty by using efficient algorithms to compute inner products $\langle \Phi(x), \Phi(y) \rangle_V$ *without* computing the infinite dimensional representations $\Phi(x)$. This is often referred to as the *kernel trick*.

Some of the feature maps discussed in this thesis are not valued in Hilbert spaces; rather they will be valued in a Banach space. However, the concept of universal and characteristic feature maps still hold in this setting.

We begin with a brief introduction to kernel methods, followed by a discussion of universal and characteristic feature maps. A common method to prove the universality of a feature map is by using the Stone-Weierstrass theorem. However, the standard Stone-Weierstrass result only holds for locally compact spaces, and the primary spaces of interest in this thesis, such as path spaces and mapping spaces, are not locally compact. This leads into our discussion of the strict topology [Gil71] on the space of bounded continuous functions on non-locally compact spaces, which admits a Stone-Weierstrass type result.

We will then consider the signature as a feature map for the path space and define the corresponding kernel. We discuss the results in [CO18], which first utilized the strict topology to prove the universal and characteristic properties of the path signature. We will end the section by considering the discrete path signature [KO19; DET20] which leads to a more efficient algorithms to compute the signature kernel.

1.3.1 *Kernel Methods*

Let $\mathcal{X}$ be a topological space and V be a Hilbert space. Given a feature map $\Phi : \mathcal{X} \to V$, we define a corresponding *kernel function*:

$$\kappa : \mathcal{X} \times \mathcal{X} \to \mathbb{R}, \quad \kappa(x, x') \coloneqq \langle \Phi(x), \Phi(x') \rangle_V.$$

Note that we can treat the dual V^* as a space of functions on $\mathcal{X}$ by viewing an element $\ell \in V^*$ as $\langle \ell, \Phi(\cdot) \rangle_V : \mathcal{X} \to \mathbb{R}$. Conversely, associated to any such positive definite kernel is a Hilbert space of functions on $\mathcal{X}$, denoted $\mathcal{H}_\kappa$, defined as the completion of the linear span of the elements $K_x = \kappa(\cdot, x) = \langle \Phi(\cdot), \Phi(x) \rangle_V : \mathcal{X} \to \mathbb{R}$. The Hilbert space $\mathcal{H}_\kappa$ is a *reproducing kernel Hilbert space (RKHS)*, and its inner product is defined by

$$\langle K_x, K_y \rangle_{\mathcal{H}_\kappa} = \kappa(x, y).$$

This construction for arbitrary positive definite kernels is the content of the Moore-Aronszajn theorem [Aro50]. At the heart of the theoretical results from kernel methods is the following representer theorem.

Theorem 1.3.1 ([SHS01])**.** *Suppose we are given a nonempty set $\mathcal{X}$, a positive definite real-valued kernel $\kappa : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ with its associated RKHS $\mathcal{H}_\kappa$, a training sample $(x_1, y_1), \ldots, (x_m, y_m) \in \mathcal{X} \times \mathbb{R}$, a strictly monotonically increasing real valued function $g : [0, \infty) \to \mathbb{R}$ to be used as a regularizer, and an arbitrary error function $E : (\mathcal{X} \times \mathbb{R}^2)^m \to \mathbb{R} \cup \{\infty\}$.*

Then any minimizer of the empirical risk

$$f^* = \operatorname{argmin}_{f \in \mathcal{H}_\kappa} \left\{ E\left((x_1, y_1, f(x_1)), \ldots, (x_m, y_m, f(x_m))\right) + g(\|f\|_{\mathcal{H}_\kappa}) \right\}$$

admits a representation of the form

$$f^* = \sum_{i=1}^{m} c_i K_{x_i},$$

where $c_i \in \mathbb{R}$.

This theorem allows us to reframe an infinite dimensional optimization problem over $\mathcal{H}_\kappa$ into a finite dimensional problem over the span of $\{K_{x_i}\}_{i=1}^m$, which is determined by the training set. This can be applied to two large classes of problems in machine learning:

1. **(Learning functions on $\mathcal{X}$)** Suppose we wish to learn a function $f : \mathcal{X} \to \mathbb{R}$. If $f \in \mathcal{H}_\kappa$, then we can use the representer theorem to recast this as a linear optimization problem, significantly simplifying the computation. This is used to learn the decision

boundary for binary classification problems using support vector machines, or to perform regression using support vector regression.

2. **(Learning measures on $\mathcal{X}$)** Suppose we wish to understand a probability measure μ on $\mathcal{X}$. Under mild assumptions, we can compute the *kernel mean embedding* (KME) of μ defined by

$$\begin{aligned}\overline{\Phi} : \mathcal{P}(\mathcal{X}) &\to \mathcal{H}_\kappa \\ \mu &\mapsto \int_{\mathcal{X}} \Phi(x) d\mu(x).\end{aligned}$$

Because V is a Hilbert space, we can use the norm to define a pseudometric for $\mathcal{P}(\mathcal{X})$, the space of Borel probability measures on $\mathcal{X}$. When the KME $\overline{\Phi}$ is injective, this becomes a metric. This induced metric is equivalent to a classical metric for measures called the maximum mean discrepancy (MMD), and can be used to perform hypothesis testing on measures [Gre+12].

From the representer theorem in Theorem 1.3.1, we can use the kernel κ to perform *nonlinear* classification tasks using *linear* methods such as the support vector machine (SVM). In fact, we can use the norm on $\mathcal{H}_\kappa$ to define a notion of distance on $\mathcal{P}(\mathcal{X})$. Although this only provides a pseudometric since $\overline{\Phi}$ may not be injective, it coincides with a well known measure of discrepancy between probability measures.

Definition 1.3.1. Let $\mathcal{G} \subset \mathbb{R}^{\mathcal{X}}$ be a class of functions and $\mu, \nu \in \mathcal{P}(\mathcal{X})$. The *maximum mean discrepancy* (MMD) of μ and ν with respect to $\mathcal{G}$ is

$$\mathrm{MMD}[\mathcal{G}, \mu, \nu] \coloneqq \sup_{f \in \mathcal{G}} \left(\mathbb{E}_\mu[f] - \mathbb{E}_\nu[f] \right). \tag{1.3.2}$$

When we take the function class $\mathcal{G}$ to be the unit ball in the RKHS $\mathcal{H}_\kappa$, the MMD can be written as the distance between the mean embeddings, with respect to the norm on $\mathcal{H}$.

Lemma 1.3.1 ([Bor+06]). *Suppose the KME map $\overline{\Phi}$ is well-defined and suppose $\mu, \nu \in \mathcal{P}(\mathcal{X})$. Let $\mathcal{G} = \{f \in \mathcal{H} \subset \mathbb{R}^{\mathcal{X}} : \|f\| \leq 1\}$. Then,*

$$\mathrm{MMD}^2[\mathcal{G}, \mu, \nu] = \|\overline{\Phi}(\mu) - \overline{\Phi}(\nu)\|^2. \tag{1.3.3}$$

This simplifies the study of probability measures by considering them as elements of a linear space, and also provides a straightforward method to compute an unbiased finite sample estimate of the MMD in terms of the kernel.

Lemma 1.3.2 ([Gre+12]). *Suppose the KME map $\overline{\Phi}$ is well-defined and suppose $\mu, \nu \in \mathcal{P}(\mathcal{X})$. Let $\mathcal{G} = \{f \in \mathcal{H}_\kappa \subset \mathbb{R}^{\mathcal{X}} : \|f\| \leq 1\}$. Let $X = (x_1, \ldots, x_n)$ and $Y = (y_1, \ldots, y_m)$ be i.i.d. samples*

from μ and ν respectively. An unbiased estimate of $\mathrm{MMD}^2[\mathcal{G}, \mu, \nu]$ *is given as the MMD of the empirical distributions of X and Y,*

$$\begin{aligned}\mathrm{MMD}_u^2[\mathcal{G}, X, Y] =& \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \kappa(x_i, x_j) + \frac{1}{m(m-1)} \sum_{i=1}^{m} \sum_{j \neq i}^{m} \kappa(y_i, y_j) \\ & - \frac{2}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} \kappa(x_i, y_j). \end{aligned} \tag{1.3.4}$$

While the requirement that V is a Hilbert space may not be satisfied in certain situations, recent work has generalized the Hilbert space kernel methods to the setting of Banach spaces [ZXZ09; FLS11]. In particular, there are generalizations to the representer theorem for reflexive Banach spaces [Sch19b], where the necessity of the reflexive hypothesis is also discussed. In the non-reflexive setting, approximate representer theorems have been developed in [Sch20]. Although we do not pursue this approach in the present paper, this suggests a potential avenue for future work.

1.3.2 *Universal and Characteristic Feature Maps*

In order to effectively use the feature map to study the two classes of machine learning problems listed above, the feature map must satisfy additional properties. In the case of learning functions, we must first specify the function class $\mathcal{G} \subset \mathbb{R}^{\mathcal{X}}$ of interest, and show that $\mathcal{H}_\kappa = \mathcal{G}$. This means that the linear span of functionals $\langle \ell, \Phi(\cdot) \rangle_V : \mathcal{X} \to \mathbb{R}$ for $\ell \in V^*$ is a dense subset of $\mathcal{G}$. In the case of learning measures, we must ensure that the KME map is injective. These two properties are respectively called being *universal* and *characteristic*. The discussion in this section is adapted from [CO18].

In this section, we will only suppose that V is a topological vector space. Let V' denote the topological dual of continuous linear functionals, and let V^* denote the algebraic dual of all linear functionals. Recall that these duals are equivalent when V is finite dimensional.

Fix a function class $\mathcal{G} \subset \mathbb{R}^{\mathcal{X}}$. We will require a generalization of the KME defined on the dual $\mathcal{G}'$, which may include distributions. Rather than defining the KME using an integral, the KME will send a distribution $D \in \mathcal{G}'$ to a linear functional on V' as follows

$$\begin{aligned} \overline{\Phi} : \mathcal{G}' &\to (V')^* \\ D &\mapsto \big(\ell \mapsto D(\langle \ell, \Phi(\cdot) \rangle_V)\big), \end{aligned}$$

where $\ell \in V'$. With this generalized KME, we can now define universal and characteristic maps.

Definition 1.3.2. Fix an input topological space $\mathcal{X}$, and a function space $\subset \mathbb{R}^{\mathcal{X}}$. Consider a feature map

$$\Phi : \mathcal{X} \to V.$$

Suppose that $\langle \ell, \Phi(\cdot) \rangle \in \mathcal{G}$ for all $\ell \in V'$. We say that Φ is

1. *universal to* $\mathcal{G}$ if the map

$$\iota : V \to \mathcal{G}, \quad \ell \mapsto \langle \ell, \Phi(\cdot) \rangle$$

has a dense image in $\mathcal{G}$; and

2. *characteristic to* $\mathcal{G}'$ if the KME map

$$\overline{\Phi} : \mathcal{G}' \to (V')^*, \quad D \mapsto \big(\ell \mapsto D(\langle \ell, \Phi(\cdot) \rangle_V)\big)$$

is injective.

Remark 1.3.1. Note that when V is a finite dimensional Hilbert space, we have $(V')^* = V$, and can define the KME for distributions in $\mathcal{G}'$ using the Pettis integral, as described in [SS18].

In the case where V is a Hilbert space, the characteristic property of the feature map implies that the MMD between measures defined by the norm of V is a metric.

Corollary 1.3.1. *Suppose V is a Hilbert space and $\Phi : \mathcal{X} \to V$ is a feature map which is characteristic with respect to $\mathcal{P}(\mathcal{X})$. The maximum mean discrepancy with respect to $\mathcal{G} = \{f \in V \subset \mathbb{R}^{\mathcal{X}} : \|f\| \leq 1\}$ is a metric.*

We have the following equivalence between universal and characteristic maps.

Theorem 1.3.2 ([CO18])**.** *Suppose that $\mathcal{G}$ is a locally convex topological vector space. A feature map Φ is universal to $\mathcal{G}$ if and only if Φ is characteristic to $\mathcal{G}'$.*

Remark 1.3.2. This equivalence may seem to stand in contrast with previous work relating universality and characteristicness, such as in [Kwi+15], which states that a characteristic feature map is not necessarily universal. This is due to variations in the definition of characteristicness; for example in [Kwi+15], a feature map is called characteristic if its corresponding KME is injective on probability measures. An extensive discussion on the various notions of universal and characteristic maps is provided in [SS18].

1.3.3 *The Strict Topology*

The equivalence between universal and characteristic maps significantly simplifies the study of such properties, as we only need to prove one of the two, and in many cases, universality can be shown using Stone-Weierstrass type results. Our focus in this paper is to show universality with respect to $\mathcal{G} = C_b(\mathcal{X})$, the space of continuous bounded real-valued functions on $\mathcal{X}$. Furthermore, we also wish to show characteristicness with respect to probability measures on $\mathcal{X}$.

An issue that arises is that the input spaces $\mathcal{X}$ that we consider in this paper are not locally compact, and the classical Stone-Weierstrass theorem does not apply. However, as was first noted in [CO18] for applications to signatures, when $\mathcal{G} = C_b(\mathcal{X})$ is equipped with the *strict topology* [Gil71], there exists a Stone-Weierstrass result, and $\mathcal{G}'$ includes the space of probability measures on $\mathcal{X}$.

Definition 1.3.3 ([Gil71])**.** Let $\mathcal{X}$ be a topological space. We say that a function $\psi : \mathcal{X} \to \mathbb{R}$ vanishes at infinity if for each $\epsilon > 0$, there exists a compact set $K \subset \mathcal{X}$ such that $\sup_{x \in \mathcal{X} - K} |\psi(x)| < \epsilon$. Denote by $C_0(\mathcal{X})$ the set of functions that vanish at infinity. The *strict topology* on $C_b(\mathcal{X})$ is the topology generated by the seminorms

$$p_\psi(f) = \sup_{x \in \mathcal{X}} |f(x)\psi(x)|, \quad \psi \in C_0(\mathcal{X}).$$

Theorem 1.3.3 ([Gil71])**.** *Let $\mathcal{X}$ be a metrizable topological space.*

1. *The strict topology on $C_b(\mathcal{X})$ is weaker than the uniform topology and stronger than the topology of uniform convergence on compact sets.*
2. *If $\mathcal{G}_0$ is a subalgebra of $C_b(\mathcal{X})$ such that for all $x, y \in \mathcal{X}$, there exists some $f \in \mathcal{G}_0$ such that $f(x) \neq f(y)$ ($\mathcal{G}_0$ separates points), and for all $x \in \mathcal{X}$, there exists some $f \in \mathcal{G}_0$ such that $f(x) \neq 0$, then $\mathcal{G}_0$ is dense in $C_b(\mathcal{X})$ under the strict topology.*
3. *The topological dual of $C_b(\mathcal{X})$ equipped with the strict topology is the space of finite regular Borel measures on $\mathcal{X}$.*

Specifically, the space of finite regular Borel measures on $\mathcal{X}$ includes all probability measures on $\mathcal{X}$.

1.3.4 *The Signature Kernel*

The perspective in this section is to view the path signature $S : C^{1-var}(V) \to T((V))$ as a feature map and consider the universal and characteristic property with respect to $\mathcal{G} = C_b(C^{1-var}(V))$. As described in Definition 1.3.2, we must first ensure that $\langle \ell, \Phi(\cdot) \rangle_V \in \mathcal{G}$

for all $\ell \in T((V))'$. In particular, we must show that $\langle \ell, \Phi(\cdot) \rangle_V : \mathcal{X} \to \mathbb{R}$ is bounded. This will be achieved using tensor normalization.

Definition 1.3.4. A *tensor normalization* is a continuous injective map of the form

$$\begin{aligned} \Lambda : T((V)) &\to \{\mathbf{t} \in T((V)) \,:\, \|\mathbf{t}\| \leq K\} \\ \mathbf{t} &\mapsto \sigma_{\lambda(\mathbf{t})}\mathbf{t} \end{aligned}$$

where $K > 0$ is a constant, $\lambda : T((V)) \to (0, \infty)$ is a function, and σ_λ is multilinear scaling operator from Equation (1.2.17).

The following proposition, proved in Appendix A of [CO18], shows the existence and construction of the tensor normalization

Proposition 1.3.1 ([CO18]). *Let V be a Banach space. Let $\psi : [1, \infty) \to [1, \infty)$ with $\psi(1) = 1$. For $\mathbf{t} \in T((V))$ and define $\lambda : T((V)) \to (0, \infty)$ to be the unique non-negative number such that $\|\sigma_{\lambda(\mathbf{t})}\mathbf{t}\|^2 = \psi(\|\mathbf{t}\|)$, where σ_λ is the multi-linear scaling function in Equation (1.2.17). Define*

$$\begin{aligned} \Lambda : T((V)) &\to T((V)) \\ \mathbf{t} &\mapsto \sigma_{\lambda(\mathbf{t})}\mathbf{t}. \end{aligned}$$

Denote $\|\psi\|_\infty = \sup_{x \geq 1} \psi(x)$. Then the following holds.

1. *The function Λ takes values in the set $\left\{\mathbf{t} \in T((V)) \,:\, \|\mathbf{t}\| \leq \sqrt{\|\psi\|_\infty}\right\}$.*
2. *If ψ is injective, then so is Λ.*
3. *Suppose that $\sup_{x \geq 1} \psi(x)/s^2 \leq 1$, $\|\psi\|_\infty < \infty$, and that ψ is K-Lipschitz for some $K > 0$. Then*

$$\|\Lambda(\mathbf{s}) - \Lambda(\mathbf{t})\| \leq (1 + \sqrt{K} + 2\sqrt{\|\psi\|_\infty})(\sqrt{\|\mathbf{s} - \mathbf{t}\|} \vee \|\mathbf{s} - \mathbf{t}\|).$$

Corollary 1.3.2 ([CO18]). *Let $\psi : [1, \infty) \to [1, \infty)$ be injective satisfying $\psi(1) = 1$ and the conditions of item (3.) in Proposition 1.3.1. Then, the function Λ constructed in Proposition 1.3.1 is a tensor normalization.*

This method of constructing a tensor normalization is done by defining the normalized norm $\|\sigma_{\lambda(\mathbf{t})}\mathbf{t}\|$ of an element $\mathbf{t}$, given its original norm $\psi(\|\mathbf{t}\|)$. In practice, we will be working with a truncated element of the tensor algebra. Thus, given an element $\mathbf{t} \in$

$T^{(M)}((V))$, we can obtain the value of $\lambda(\mathbf{t})$ by solving the equation $\|\sigma_{\lambda(\mathbf{t})}\mathbf{t}\|^2 = \psi(\|\mathbf{t}\|)$. This reduces to finding the zero of the polynomial equation

$$\sum_{k=1}^{M} \lambda^{2k}\|\mathbf{t}_k\|^2 - \psi(\|\mathbf{t}\|) = 0.$$

The following result shows that the path signature is a universal and characteristic feature map, and we reproduce the proof here for completeness.

Theorem 1.3.4 ([CO18]). *Let $\Lambda : T((V)) \to T((V))$ be a tensor normalization. The normalized signature*

$$\Phi : \widetilde{C}^{1-var}(V) \to T((V)), \quad \Phi = \Lambda \circ S,$$

1. *is a continuous injection from $\widetilde{C}^{1-var}(V)$ into a bounded subset of $T((V))$,*
2. *is universal to $\mathcal{G} := C_b(\widetilde{C}^{1-var}(V))$ equipped with the strict topology, and*
3. *is characteristic to the space of finite regular Borel measures on $\widetilde{C}^{1-var}(V)$.*

Proof. From Theorem 1.2.3 and Theorem 1.2.4, the path signature is a continuous injection on $\widetilde{C}^{1-var}(V)$. Furthermore, since the tensor normalization is continuous and injective, Φ is a continuous injection.

To prove the second statement, let $\mathcal{G}_0 = L_V$, as defined in Equation (1.2.28), which we view as a subset of $C_b(\widetilde{C}^{1-var}(V))$ by setting

$$\ell(\gamma) := \langle \ell, \Phi(\gamma) \rangle,$$

for any $\ell \in L_V$ and $\gamma \in \widetilde{C}^{1-var}(V)$. Because Φ is injective, L_V separates points. Next, for each $\gamma \in \widetilde{C}^{1-var}(V)$, there is some $\ell \in L_V$ such that $\langle \ell, \Phi(\gamma) \rangle \neq 0$. Moreover, L_V is a subalgebra of $C_b(\widetilde{C}^{1-var}(V))$ by Corollary 1.2.1. By the second part of Theorem 1.3.3, Φ is universal with resepct to $C_b(\widetilde{C}^{1-var}(V))$ equipped with the strict topology.

Finally, to prove the third statement, the equivalence in Theorem 1.3.2 shows that Φ is characteristic with respect to the topological dual of $C_b(\widetilde{C}^{1-var}(V))$. By the third part of Theorem 1.3.3, the topological dual is the space of finite regular Borel measures. □

1.3.5 *The Discrete Path Signature*

While the path signature is naturally defined for continuous paths, we often work with discrete time series in data science applications. There is a discrete approximation of the continuous signature for finite dimensional discrete paths which allows for more efficient

computation, originally introduced in [KO19]. In this subsection, we recall main concepts of this approximation.

In this subsection, we let V denote a Hilbert space with orthonormal basis $\{e_\alpha\}_{\alpha \in A}$ for some index set A. For $L \in \mathbb{N}$, we let $[L] := \{1, \dots, L\}$. A *discrete path* is a map $\hat{\gamma} : [L+1] \to V$, and its *discrete derivative* $\hat{\gamma}' : [L] \to V$ is defined to be

$$\hat{\gamma}'(t) = \hat{\gamma}_{t+1} - \hat{\gamma}_t$$

for all $t \in [L]$. We let $\mathrm{Seq}(V)$ denote the space of discrete paths of arbitrary finite length on V. Furthermore, for $\alpha \in A$, we denote $\gamma_\alpha : [L+1] \to \mathbb{R}$ to be the α-component of the path.

Definition 1.3.5. Let $\hat{\gamma} : [L+1] \to V$. Let $I = (i_1, \dots, i_m) \in A^m$ be a multi-index. We define the discrete m-simplex with length L to be

$$\hat{\Delta}_L^m = \{(t_1, \dots, t_m) \in [L]^m \quad : \quad 0 \leq t_1 < t_2 < \dots < t_m \leq L\}.$$

The discrete path signature of $\hat{\gamma}$ with respect to I is defined to be

$$\hat{S}^I(\hat{\gamma}) := \sum_{(t_1,\dots,t_m) \in \hat{\Delta}_L^m} \hat{\gamma}'_{i_1}(t_1) \dots \hat{\gamma}'_{i_m}(t_m).$$

The discrete path signature can be viewed as an approximation to the continuous path signature. Let $\gamma \in C^{1-var}(V)$ be a bounded variation path. Given a partition $\Pi = (0 = t_1 < t_2 < \dots < t_{L+1} = 1)$, the discretization of γ with respect to Π, denoted $\hat{\gamma}^{(\Pi)} : [L+1] \to G$, is defined to be

$$\hat{\gamma}^{(\Pi)}(i) := \gamma_{t_i}.$$

The following proposition in [KO19] shows that the discrete signature indeed approximates the continuous path signature and provides an explicit error estimate.

Proposition 1.3.2 (Corollary 4.3, [KO19])**.** *Let $\gamma \in C^{1-var}(V)$, and define a partition $\pi = (0 = t_1 < t_2 < \dots < t_{L+1} = 1)$. Then,*

$$\|\hat{S}(\hat{\gamma}^{(\pi)}) - S(\gamma)\| \leq |\gamma|_{1-var} e^{|\gamma|_{1-var}} \max_{i=1,\dots,L} |\gamma|_{[t_i,t_{i+1}]}|_{1-var}.$$

The discrete signature can be viewed as a map

$$\hat{S} : \mathrm{Seq}(V) \to T((V)),$$

or in truncated form as

$$\hat{S}_M : \mathrm{Seq}(V) \to T^{(M)}(V).$$

We will be using the truncated variant of the signature in data analysis applications due to its computability. We can bound the truncation error as follows.

Lemma 1.3.3. *Let $\hat{\gamma} : [L+1] \to V$, and define $\ell := |\hat{\gamma}|_{1-var}$. Given $M \in \mathbb{N}$, we have*

$$\|\hat{S}(\hat{\gamma}) - \hat{S}_{\leq M}(\hat{\gamma})\| \leq \left(\frac{\exp(\ell^2)\ell^{M+1}}{(M+1)!}\right)^{1/2}.$$

Proof. First, we write a recursive definition

$$\hat{S}_M(\hat{\gamma})(t) = \sum_{t' \in [t]} \hat{S}_{M-1}(\hat{\gamma})(t-1) \otimes \hat{\gamma}'(t)$$

For $t' \in \mathbb{N}$, let $\hat{\gamma}_{[0,t']}$ denote the restriction of $\hat{\gamma}$ to $[t']$. We claim that

$$\hat{S}_M(\hat{\gamma})(t') \leq \frac{|\hat{\gamma}_{[0,t']}|_{1-var}^M}{M!}$$

The case of $M = 1$ is straightforward since

$$\hat{S}_1(\hat{\gamma})(t') = \|\hat{\gamma}_{t'} - \hat{\gamma}_0\| \leq |\hat{\gamma}|_{[0,t']}|_{1-var}.$$

Then, we will use induction to show the case for $M > 1$. By expanding definitions, we have

$$\begin{aligned}
\|\hat{S}_M(\hat{\gamma})(t)\| &= \left\| \sum_{t' \in [t]} \hat{S}_{M-1}(\hat{\gamma})(t'-1) \otimes \hat{\gamma}'(t') \right\| \\
&\leq \sum_{t' \in [t]} \|\hat{S}_{M-1}(\hat{\gamma})(t'-1) \otimes \hat{\gamma}'(t')\| \\
&\leq \sum_{t' \in [t]} \|\hat{S}_{M-1}(\hat{\gamma})(t'-1)\| \cdot \|\hat{\gamma}'(t')\| \\
&\leq \sum_{t' \in [t]} \frac{|\hat{\gamma}_{[0,t']}|_{1-var}^{M-1}}{(M-1)!} \cdot \|\hat{\gamma}'(t')\|,
\end{aligned}$$

where we use the compatibility condition of the tensor norms in the third line and the induction hypothesis in the last line. Now, let $\gamma : [0, L+1] \to V$ continuous path defined by linear interpolation of $\hat{\gamma}$, and note that

$$|\gamma_{[0,t']}|_{1-var} = |\hat{\gamma}_{[0,t']}|_{1-var}.$$

Then, since $|\gamma_{[0,t']}|_{1-var}$ a monotone increasing function of t', we can use an integral to bound the sum

$$\sum_{t' \in [t]} \frac{|\hat{\gamma}_{[0,t']}|_{1-var}^{M-1}}{(M-1)!} \cdot \|\hat{\gamma}'(t')\| \leq \frac{1}{(M-1)!} \int_0^t |\gamma_{[0,t']}|_{1-var}^{M-1} \cdot \|\gamma'(t')\| dt'$$

Applying a change of variables with $s = |\gamma_{[0,t']}|_{1-var}$ with $ds = \|\gamma'(t')\| dt'$, we get

$$\begin{aligned} \frac{1}{(M-1)!} \int_0^t |\gamma_{[0,t']}|_{1-var}^{M-1} \cdot \|\gamma'(t')\| dt' &\leq \frac{1}{(M-1)!} \int_0^{|\gamma_{[0,t]}|_{1-var}} s^{M-1} ds \\ &= \frac{|\gamma_{[0,t]}|_{1-var}^{M}}{M!}, \end{aligned}$$

which proves the claim. Now, let $\ell := |\hat{\gamma}|_{1-var}$, so we have

$$\hat{S}_M(\hat{\gamma}) \leq \frac{\ell^M}{M!}$$

for all $M \in \mathbb{N}$. The truncation error can then by given by

$$\begin{aligned} \|\hat{S}(\hat{\gamma}) - \hat{S}_{\leq M}(\hat{\gamma})\|^2 &= \sum_{k=M+1}^{\infty} \|\hat{S}_k(\hat{\gamma})\| \\ &\leq \sum_{k=M+1}^{\infty} \frac{\ell^{2k}}{(k!)^2} \\ &\leq \sum_{k=M+1}^{\infty} \frac{\ell^{2k}}{k!} \\ &\leq \frac{\exp(\ell^2)\ell^{M+1}}{(M+1)!}. \end{aligned}$$

□

We define the truncated discrete signature kernel to be

$$\begin{aligned} \hat{K}_M : \mathrm{Seq}(V) \times \mathrm{Seq}(V) &\to \mathbb{R} \\ (\hat{\alpha}, \hat{\beta}) &\mapsto \langle \hat{S}_M(\hat{\alpha}), \hat{S}_M(\hat{\beta}) \rangle. \end{aligned}$$

By expanding the definition of the discrete signature, we can formulate the following expression for the signature kernel.

Proposition 1.3.3 ([KO19]). *The truncated discrete signature kernel can be computed as*

$$\hat{K}_M(\hat{\beta}, \hat{\gamma}) = \sum_{m=0}^{M} \sum_{(s,t)\in \hat{\Delta}_T^m \times \hat{\Delta}_T^m} \prod_{i=1}^{m} \langle \hat{\beta}'(s_i), \hat{\gamma}'(t_i) \rangle. \tag{1.3.5}$$

When V is a finite-dimensional Hilbert space, this expression can be explicitly computed, with an efficient algorithm given in [KO19]. Furthermore, we may be interested in situations where the time series is valued in a topological space $\mathcal{X}$, which is then mapped into a Hilbert space V using a feature map $\Phi : \mathcal{X} \to V$. This is equipped with the associated kernel function $\kappa : \mathcal{X} \times \mathcal{X} \to V$, and we are interested in the composition of the the feature map Φ with the discrete signature kernel

$$\begin{aligned} K_{\Phi,m} : \mathrm{Seq}(\mathcal{X}) \times \mathrm{Seq}(\mathcal{X}) &\to \mathbb{R} \\ (\beta, \gamma) &\mapsto \langle S_m \circ \Phi(\beta), S_m \circ \Phi(\gamma) \rangle_V. \end{aligned}$$

By once again expanding the definitions, we obtain a similar expression for the signature kernel.

Proposition 1.3.4 ([KO19]). *Let $\Phi : \mathcal{X} \to V$ be a feature map into a Hilbert space, and suppose $\kappa : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be the corresponding kernel. The truncated discrete signature kernel can be computed as*

$$\hat{K}_M(\hat{\beta}, \hat{\gamma}) = \sum_{m=0}^{M} \sum_{(s,t)\in \hat{\Delta}_T^m \times \hat{\Delta}_T^m} \prod_{i=1}^{m} \nabla_{s_i,t_i} \kappa(\hat{\beta}, \hat{\gamma}), \tag{1.3.6}$$

where

$$\nabla_{j,k}\kappa(\hat{\beta}, \hat{\gamma}) := \kappa(\hat{\beta}_{j+1}, \hat{\gamma}_{k+1}) - \kappa(\hat{\beta}_j, \hat{\gamma}_{k+1}) - \kappa(\hat{\beta}_{j+1}, \hat{\gamma}_k) + \kappa(\hat{\beta}_j, \hat{\gamma}_k).$$

This result implies that even if the feature space V is infinite dimensional, as long as we have a computable kernel κ, we can compute the signature kernels for discrete paths in $\mathrm{Seq}(\mathcal{X})$.

Remark 1.3.3. There is another approach to computing signature kernels which is discussed in [Cas+21]. In this approach, the untruncated signature kernel is formulated in terms of the solution to a PDE. This approach can be also be extended to time series composed with a feature map $\Phi : \mathcal{X} \to V$ by using explicit computations for its corresponding kernel $\kappa : \mathcal{X} \times \mathcal{X} \to V$. In particular, this method would allow us to compute the *untruncated* signature kernel composed with the *untruncated* moment map. In this case, the approximation error would be purely due to the PDE solution error rather than a truncation error.

1.3.6 *Path Transformations*

We turn to a discussion of several transformations which add information to paths, usually used as a preprocessing step for signature methods in machine learning. This section is from [LG20], but is also discussed in [Fer21; Mor+21].

1.3.6.1 *Time transformation*

The *time transformation* is a simple method to remove reparametrization invariance (common in the path signatures literature such as in [CK16]), defined by appending the time parameter to the path:

$$\begin{aligned} T_{\text{Time}} : C_0^{1-var}(V) &\to C_0^{1-var}(V \times \mathbb{R}) \\ \gamma_t &\mapsto (\gamma_t, t). \end{aligned} \tag{1.3.7}$$

Lemma 1.3.4. *The map $T_{\text{Time}} : C_0^{1-var}(V) \to \widetilde{C}_0^{1-var}(V \times \mathbb{R})$ is injective.*

Proof. First, note that T_{Time} is an injective map into a subset of $C_0^{1-var}(V \times \mathbb{R})$ where at least one component is monotone increasing. The proof of Theorem 5 in [LO14] shows that the signature is injective on paths with a monotone increasing component. Then, since the signature of two paths is equivalent if and only if the two paths are tree-like equivalent, this implies that T_{Time} is injective into $\widetilde{C}_0^{1-var}(V \times \mathbb{R})$. □

This lemma implies that by first embedding a path $\gamma \in C_0^{1-var}(V)$ into $\widetilde{C}_0^{1-var}(V \times \mathbb{R})$, we obtain a parametrization-dependent feature map for $\gamma \in C_0^{1-var}(V)$ using the path signature. More generally, this also removes tree-like invariance, so in fact we obtain an injective feature map for $C_0^{1-var}(V)$.

1.3.6.2 *Identity initialized transformation*

The *identity initialized (IdInit) transformation* is a simple method to remove translation invariance. Suppose $v \in V$ and we define $\ell_v \in C^{1-var}(V)$ to be the linear interpolation between the origin of V to the point $v \in V$. Then the IdInit transformation is defined to be

$$\begin{aligned} T_{\text{IdInit}} : C^{1-var}(V) &\to C_0^{1-var}(V) \\ \gamma &\mapsto \ell_{\gamma_0} * \gamma. \end{aligned} \tag{1.3.8}$$

For discrete time series, this simply amounts to appending the identity element to the beginning of the time series. The following lemma is clear by definition.

Lemma 1.3.5. *The map $T_{\text{IdInit}} : C^{1-var}(V) \to C_0^{1-var}(V)$ is injective.*

The implication of this lemma is that by embedding $C^{1-var}(V)$ into $C_0^{1-var}(V)$, we can obtain a translation-dependent feature map for $C^{1-var}(V)$ using the path signature. Moreover, by combining the time and identity start transformations, the path signature provides an injective feature map for $C^{1-var}(V)$.

Corollary 1.3.3. *The composition $T_{\text{Time}} \circ T_{\text{IdInit}} : PG \to \widetilde{P}(G \times \mathbb{R})$ is injective.*

1.3.6.3 *Sliding window transformation*

The sliding window transformation (often called the lead-lag transformation in the path signature literature) has been used for several applications of the path signatures, such as [Gyu+13] and [Yan+19], and has been empirically found to produce good results in classification tasks [Fer21]. The exact definition may vary slightly between these papers, and we give the definition from [Yan+19] and [Fer21].

Let $\tau \in \mathbb{R}^+$, and $m \in \mathbb{N}$. Given a path $\gamma \in C^{1-var}(V)$, we can extend its definition to all of $\mathbb{R}$ as

$$\gamma_t = \begin{cases} \gamma_0 & : t < 0 \\ \gamma_t & : t \in [0,1] \\ \gamma_1 & : t > 1. \end{cases}$$

Then, we define the *sliding window transformation with m lags* to be

$$\begin{aligned} T_{\text{SWin},m} : C^{1-var}(V) &\to (C^{1-var}(V))^{m+1} \\ \gamma_t &\mapsto (\gamma_t, \gamma_{t-\tau}, \gamma_{t-2\tau}, \dots, \gamma_{t-m\tau}). \end{aligned} \tag{1.3.9}$$

For discrete time series, we assume that the data is temporally uniformly sampled, and we choose τ to be the time in between samples. Then, if we consider a discrete time series to be $\hat{\gamma} : [L+1] \to V$, then the sliding window transformation will be

$$T_{\text{SWin},m}(\hat{\gamma})_i := (\hat{\gamma}_i, \hat{\gamma}_{i-1}, \dots, \hat{\gamma}_{i-m}).$$

In the context of Euclidean path signatures, [Fer21] empirically shows that the sliding window embedding often performs well on classification tasks, though there is no theoretical explanation. We note that due to the choice of padding the start of the delayed time series with the identity, this transformation breaks translation invariance. We suggest that breaking the translation symmetry is one reason the sliding window transformation performs well in practice. This is discussed in Remark 3.5.1 in Section 3.5.1.

1.4 ROUGH PATHS

In this section, we provide a brief exposition of the basic ideas behind rough paths. Although they are not explicitly used in this thesis, the field of rough paths has played a major role in the development of the theory and applications of path signatures in the past two decades. Furthermore, one of the main motivations for the development of the mapping space signature in Chapter 5 and Chapter 7 is to work towards a generalization of rough maps to higher dimensional domains.

The goal of this section is to motivate rough paths, provide the definition, and show that the path signature can be defined for rough paths. We will follow the exposition in [LCL07] in this section; in particular, we will only consider the finite dimensional case of $V = \mathbb{R}^N$. Another reference for the finite dimensional theory, including more applications to stochastic analysis is [FV10], while [LQ07] works in the setting of arbitrary vector spaces. Note that there is an alternate formulation of rough paths from an algebraic perspective [Gub04] which is also discussed in [FH20].

Let V and W be Banach spaces. Suppose $X : [0,1] \to V$ and $f : W \to L(V,W)$, where $L(V,W)$ is the Banach space of continuous linear maps between V and W. Consider the controlled differential equation

$$\frac{dY}{dt} = f(Y(t))\frac{dX}{dt}, \qquad Y(0) = Y_0 \tag{1.4.1}$$

where X is the control, f is the vector field controlled by X, and Y is the response. We say that $Y : [0,1] \to W$ is a *solution* of Equation (1.4.1) if

$$Y(t) = Y_0 + \int_0^t f(Y(s))dX(s). \tag{1.4.2}$$

In order to make sense of the solution, we must be precise with the type of integral used. To begin, we suppose $V, W = \mathbb{R}$ and $f(Y(t))dX(t) = Y(t)dX(t)$. Then consider the integral

$$\int_0^t Y(s)dX(s). \tag{1.4.3}$$

in isolation. If Y is continuous and X is bounded 1-variation, then the classical Riemann-Stieltjes integration theory tells us this integral exists. In fact, Young [You36] generalized the Riemann-Stieltjes result and showed that if $Y \in C^{q-var}(V)$ and $X \in C^{p-var}(V)$, with $\frac{1}{p} + \frac{1}{q} > 1$, then the resulting integral exists, and as a function of t, the integral in Equation (1.4.3) is bounded p-variation.

Returning to the integral equation in Equation (1.4.2) with the same f, Young's result implies that if $X \in C^{p-var}(V)$, then the solution $Y \in C^{p-var}(V)$. Now, Young's condition

for the existence of the integral reduces to $\frac{2}{p} > 1$, or $p < 2$. Young's integral can be used to prove the following existence and uniqueness theorem for controlled differential equations.

Theorem 1.4.1 ([LCL07]). *Let p and r be such that $1 \leq p < 2$ and $p < r$. Suppose $X \in C^{p-var}(V)$ and f is $\mathrm{Lip}(r)$. The, for every $Y_0 \in W$, the differential equation Equation (1.4.3) admits a unique solution. Let $I_f(X, Y_0) = Y$ denote the unique solution to Equation (1.4.3) starting at Y_0. Then the mapping*

$$I_f : C^{p-var}(V) \times W \to C^{p-var}(W) \tag{1.4.4}$$

is continuous.

Here, $\mathrm{Lip}(r)$ denotes a Lipschitz continuous function with additional regularity conditions defined in [LCL07]. The main property of $\mathrm{Lip}(r)$ functions required in this result is if $Y \in C^{p-var}(W)$ and $f \in \mathrm{Lip}(r)$, then $f(Y) \in C^{\frac{p}{r}-var}(W)$.

However, there are fundamental obstructions to the extension of Young's integral to lower regularity paths, as well as extensions of Theorem 1.4.1 to $p \geq 2$. Two such reasons are given in [LCL07]:

1. There are no continuous extensions of the Young integral rich enough to handle Brownian paths in a deterministic fashion;
2. The signed area of a path is not a continuous functional with respect to the 2-variation metric.

Thus, ordinary paths of low regularity do not contain enough information to establish a deterministic integration theory. Note in particular that this implies that the path signature cannot be defined for bounded p-variation paths with $p \geq 2$. The perspective of rough paths is to focus on the path signature as the fundamental object. We will require two preliminary definitions to define rough paths.

Definition 1.4.1. A *control function* is a continuous non-negative function $\omega : \Delta^2 \to \mathbb{R}$, which is super-additive in the sense that

$$\omega(s,t) + \omega(t,u) \leq \omega(s,u),$$

for all $0 \leq s \leq t \leq u \leq 1$ and for which $\omega(t,t) = 0$ for all $t \in [0,1]$. For a path $\gamma : [0,1] \to V$, if there exists a control function ω such that

$$\|\gamma|_{[s,t]}\|_{p-var} \leq \omega(s,t)^{1/p},$$

then we say that the p-variation of γ is *controlled by* ω.

Example 1.4.1. As a simple example, let $\gamma \in C^{p-var}(V)$. Define $\omega : \Delta^2 \to \mathbb{R}$ by

$$\omega(s,t) := \|\gamma|_{[s,t]}\|^p_{p-var}.$$

It can be easily checked that ω satisfies the conditions to be a control function, and furthermore, the p-variation of γ is controlled by ω.

Definition 1.4.2. Suppose $n \in \mathbb{N}$. Let $\mathbf{X} : \Delta^2 \to T^{(n)}(V)$ be a continuous map. For each $(s,t) \in \Delta^2$, we denote $\mathbf{X}_{s,t} := \mathbf{X}(s,t)$, and write

$$\mathbf{X}_{s,t} = (\mathbf{X}^0_{s,t}, \mathbf{X}^1_{s,t}, \dots, \mathbf{X}^n_{s,t}),$$

where $\mathbf{X}^m_{s,t} \in V^{\otimes m}$. The function $\mathbf{X}$ is a *multiplicative functional of degree n in V* if $\mathbf{X}^0_{s,t} = 1$ and

$$\mathbf{X}_{s,u} \otimes \mathbf{X}_{u,t} = \mathbf{X}_{s,t} \tag{1.4.5}$$

for all $s \leq u \leq t$.

Example 1.4.2. Suppose $\gamma \in C^{1-var}(V)$ and let $n \in \mathbb{N}$. Define a map $\mathbf{X} : \Delta^2 \to T^{(n)}(V)$ by

$$\mathbf{X}_{s,t} := S_{\leq n}(\gamma|_{[s,t]}).$$

By definition, we have $\mathbf{X}^0_{s,t} = 1$ for all $(s,t) \in \Delta^2$, and by Chen's identity in Theorem 1.2.2, the multiplicative property in Equation (1.4.5) holds:

$$\mathbf{X}_{s,t} = S_{\leq n}(\gamma|_{[s,u]} * \gamma|_{[u,t]}) = S_{\leq n}(\gamma|_{[s,u]}) \otimes S_{\leq n}(\gamma|_{[u,t]}) = \mathbf{X}_{s,u} \otimes \mathbf{X}_{u,t}.$$

Definition 1.4.3. Let $p \geq 1$ be a real number. A *p-rough path in V* is a multiplicative functional of degree $\lfloor p \rfloor$ in V such that there exist a control function $\omega : \Delta^2 \to \mathbb{R}$ where

$$\|\mathbf{X}^m_{s,t}\|_m \leq \frac{\omega(s,t)^{\frac{m}{p}}}{\beta\left(\frac{m}{p}\right)!} \tag{1.4.6}$$

for all $(s,t) \in \Delta^2$ and $m = 1, \dots, \lfloor p \rfloor$, where

$$\beta = p^2 \left(1 + \sum_{r=3}^{\infty} \left(\frac{2}{r-2}\right)^{\frac{\lfloor p \rfloor + 1}{p}}\right). \tag{1.4.7}$$

Example 1.4.3. Consider the multiplicative function $\mathbf{X} : \Delta^2 \to T^{(n)}(V)$ defined in Example 1.4.2. We claim that $\mathbf{X}$ is an n-rough path. Define the control function $\omega : \Delta^2 \to \mathbb{R}$ by

$$\omega(s,t) := \beta |\gamma|_{[s,t]}|_{1-var}^n.$$

Then, by the levelwise signature norm bound in Equation (1.2.13), we have

$$\|\mathbf{X}_{s,t}^m\|_m = \|S_m(\gamma|_{[s,t]})\|_m \leq \frac{|\gamma|_{[s,t]}|^m}{m!} = \frac{\omega(s,t)^{\frac{m}{n}}}{\beta(m!)} \leq \frac{\omega(s,t)^{\frac{m}{n}}}{\beta\left(\frac{m}{n}\right)!}.$$

We can see that the first level of a p-rough path contains the same information as an ordinary path, up to translation. From this example, we see that

$$\mathbf{X}_{0,t}^1 = S_1(\gamma|_{[0,t]}) = \gamma_t - \gamma_0.$$

Furthermore, we can interpret the m^{th} level of an arbitrary p-rough path $\mathbf{X} : \Delta^2 \to T^{(\lfloor p \rfloor)}(V)$ as the analogue of the m^{th} level signature. For an arbitrary p-rough path with $p \geq 2$, the signature cannot be directly computed, as we have previously discussed. Thus, the higher level signature terms up to level $\lfloor p \rfloor$ are prescribed as part of the definition of a rough path. In fact there exist distinct p-rough paths which agree at the first level.

Example 1.4.4. We will construct a family of 2-rough paths which coincide at the first level. Let $\mathbf{w} \in V^{\otimes 2}$, and define a rough path $\mathbf{X}_\mathbf{w} : \Delta^2 \to T^{(2)}(V)$ by

$$(\mathbf{X}_\mathbf{w})_{s,t}^0 = 1, \qquad (\mathbf{X}_\mathbf{w})_{s,t}^1 = 0, \qquad (\mathbf{X}_\mathbf{w})_{s,t}^2 = (t-s)\mathbf{w}.$$

The multiplicative property in Equation (1.4.5) holds at the first level since it is trivial, and at level 2, we have

$$((\mathbf{X}_\mathbf{w})_{s,u} \otimes (\mathbf{X}_\mathbf{w})_{u,t})^2 = \mathbf{X}_{s,u}^2 + \mathbf{X}_{u,t}^2 = (u-s)\mathbf{w} + (t-u)\mathbf{w} = (t-s)\mathbf{w} = (\mathbf{X}_\mathbf{w})_{s,t}^2.$$

Now, we define a control function $\omega : \Delta^2 \to \mathbb{R}$ by

$$\omega(s,t) := (t-s)\beta\|\mathbf{w}\|_2,$$

and the regularity condition in Equation (1.4.6) is satisfied. Therefore, $\mathbf{X}_\mathbf{w}$ is a 2-rough path for any $\mathbf{w} \in V^{\otimes 2}$. Note that while $\mathbf{X}_\mathbf{w}$ is a distinct path for distinct choices of $\mathbf{w} \in V^{\otimes 2}$, the first level of $\mathbf{X}_\mathbf{w}$ is trivial, and coincides with a constant path. One can interpret the second level of these 2-rough paths as the iterated integral of infinitesimal information which cannot be represented by the underlying path.

To understand this idea, we consider a specific example where $V = \mathbb{R}^2$. We fix

$$\mathbf{w} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

Let $\gamma_n : [0,1] \to \mathbb{R}^2$ be a sequence of paths defined by

$$\gamma_n(t) = \left(\frac{\cos(2\pi n^2 t)}{\sqrt{\pi} n}, \frac{\sin(2\pi n^2 t)}{\sqrt{\pi} n} \right),$$

so that γ_n winds around a circle of radius $\frac{1}{\sqrt{\pi} n}$ a total of n^2 times. Each path in the sequence is bounded 1-variation, so we let $\mathbf{X}_n : \Delta^2 \to T^{(2)}(V)$ be the p-rough path, where $2 < p < 3$, defined by the 2-truncated path signature of γ_n,

$$(\mathbf{X}_n)_{s,t} := S_{\leq 2}(\gamma_n|_{[s,t]}),$$

as in Example 1.4.2. The second level of $\mathbf{X}_n$ can be directly computed and written as a sum of two matrices

$$\begin{aligned}(\mathbf{X}_n)^2_{s,t} =& (t-s) \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \\ &+ \frac{1}{4\pi n^2} \begin{pmatrix} \cos(4\pi n^2 t) - \cos(4\pi n^2 s) & \sin(4\pi n^2 t) - \sin(4\pi n^2 s) \\ -(\sin(4\pi n^2 t) - \sin(4\pi n^2 s)) & -(\cos(4\pi n^2 t) - \cos(4\pi n^2 s)) \end{pmatrix}.\end{aligned}$$

The second matrix converges to 0 as $n \to \infty$, and thus

$$(\mathbf{X}_n)^2_{s,t} \to (\mathbf{X}_{\mathbf{w}})^2_{s,t}$$

as $n \to \infty$. Furthermore,

$$(\mathbf{X}_n)^1_{s,t} \to 0 = (\mathbf{X}_{\mathbf{w}})^1_{s,t}$$

in the p-variation metric with $2 < p < 3$.

While the signature of a bounded p-variation path with $p \geq 2$ cannot be defined, once we lift the path to a p-rough path by prescribing its iterated integrals up to level $\lfloor p \rfloor$, we are able to perform the necessary integration steps to determine the signature.

Theorem 1.4.2 (Extension Theorem [LCL07]). *Let $p \geq 1$ be a real number and let $\mathbf{X} : \Delta^2 \to T^{(n)}(V)$ be a p-rough path controlled by ω. Then, there exists a unique extension of $\mathbf{X}$ to a multiplicative functional $\widetilde{\mathbf{X}} : \Delta^2 \to T((V))$, where*

$$\widetilde{\mathbf{X}}_{s,t} = (1, \mathbf{X}^1_{s,t}, \dots, \mathbf{X}^{\lfloor p \rfloor}_{s,t}, \mathbf{X}^{\lfloor p \rfloor + 1}_{s,t}, \dots), \tag{1.4.8}$$

such that

$$\|\mathbf{X}^m_{s,t}\| \leq \frac{\omega(s,t)^{\frac{m}{p}}}{\beta\left(\frac{m}{p}\right)!}$$

for all $m \in \mathbb{N}$.

Note that the unique extension $\widetilde{\mathbf{X}}$ coincides with the original rough path $\mathbf{X}$ up to level $\lfloor p \rfloor$. Then, we can define the *path signature* of the p-rough path $\mathbf{X}$ as

$$S(\mathbf{X}) := \widetilde{\mathbf{X}}_{0,1}. \tag{1.4.9}$$

This result is merely the beginning of the rich theory of rough paths. Using this theory, we are able to define generalizations of the 1D integral in Equation (1.4.3) where the integrator is a p-rough path $\mathbf{X}$. This, in turn, allows us to study controlled differential equations of the form in Equation (1.4.1) where the control can be a p-rough path of arbitrarily low regularity. For further details, we refer the reader to [LCL07; LQ07; FH20].

2

PATH SPACE COCHAINS AND SIGNATURES

In this chapter, we describe the path signature from the perspective of differential forms on the path space. This perspective was initiated by Chen [Che77] as a way to study the cohomology of loop spaces using geometric methods, akin to how the de Rham theorem tells us that the de Rham cohomology of a manifold (derived from the *geometry* of the manifold via differential forms) is isomorphic to the singular cohomology (derived from the *topology* of the manifold).

The main tool that Chen used to build this connection between the geometry and topology of loop spaces is the construction of the *iterated integral cochain algebra* of the loop space, ΩX, which is also called the *Chen complex of the loop space*, $\mathrm{Chen}^\bullet(\Omega X)$. The same construction can be defined for the path space of a manifold, PX, yielding the *Chen complex of the path space*, $\mathrm{Chen}^\bullet(PX)$, which is the object we will study in this chapter.

The Chen complex is a subcomplex of the de Rham complex for path spaces,

$$\mathrm{Chen}^\bullet(PX) \subset \Omega^\bullet(PX),$$

so we can interpret elements of $\mathrm{Chen}^\bullet(PX)$ as *differential forms* on the path space. However, the path space of a manifold PX is infinite-dimensional, and therefore is not a manifold. Thus, we must first establish a more general notion of a "smooth space" on which we can define notions such as differential forms. Chen introduced and used the notion of *differentiable spaces*, which is based on the following principle: we understand the concepts of smooth maps and differential forms on Euclidean spaces, so we will use Euclidean spaces to model the smooth structure of more general spaces.

If we fix $X = \mathbb{R}^N$, we find that the path signature $S^I : PX \to \mathbb{R}$ for a multi-index $I = (i_1, \dots, i_n)$ valued in $[N]$ is an element of the degree 0 part of the Chen complex

$$S^I \in \mathrm{Chen}^0(P\mathbb{R}^N) \subset \Omega^0(P\mathbb{R}^N).$$

Such an object is a 0-*cochain* or 0-*form* on PX. This connection shows that the path signature only uses the objects in the degree 0 part of the construction, where the higher cochains have yet to be exploited in the context of stochastic analysis or machine learning. While we

do not explicitly discuss the application of higher cochains in this thesis, the perspective of the path signature being the 0-cochains of the iterated integral cochain algebra is the primary motivation behind the development of the mapping space signature in the latter part of the thesis.

This chapter will introduce Chen's construction and the corresponding connection to path signatures in an expository manner. We begin with a brief discussion on differential forms on $\mathbb{R}^N$ in order to get an intuition into what it means to be a "differential form on the path space." Then, we consider diffeological spaces and differential forms on those spaces, followed by a description of Chen's construction along with the derivation of the path signature from the 0-cochains.

2.1 FROM FORMS ON EUCLIDEAN SPACES TO FORMS ON PATH SPACES

Throughout this section, we will treat X as a manifold. Recall that for each point $x \in X$, the *tangent space of X and x* is the vector space denoted $T_x X$. For $X = \mathbb{R}^N$, the tangent space at each point $x = (x_1, \dots, x_N) \in \mathbb{R}^N$ is the N-dimensional vector space spanned by the basis vectors

$$\frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_N}, \tag{2.1.1}$$

which formally describes the space of derivatives at the point $x \in \mathbb{R}^N$. For instance, given a smooth path $\gamma = (\gamma_1, \dots, \gamma_N) : [0,1] \to \mathbb{R}^N$, the derivative of γ at $t \in [0,1]$ is contained in the tangent space

$$\gamma'(t) \in T_{\gamma(t)}\mathbb{R}^N.$$

In the case of $X = \mathbb{R}^N$, tangent vectors from one tangent space $T_x\mathbb{R}^N$ can be *parallel transported* to another tangent space $T_y\mathbb{R}^N$ via the identity isomorphism $\mathrm{id} : T_x\mathbb{R}^N \cong T_y\mathbb{R}^N$. Thus, when we are working on Euclidean space, we can simply work with a single tangent space, which we take to be the tangent space at the identity $\mathfrak{r} := T_0\mathbb{R}^N \cong \mathbb{R}^N$, which is also the *Lie algebra* of $\mathbb{R}^N$. We will consistently use the vector space $\mathfrak{r}$ for tangent vectors in order to distinguish them from the underlying manifold $X = \mathbb{R}^N$. Note that for arbitrary manifolds, while the tangent spaces over any two points are isomorphic, the specific isomorphism will depend on the choice of *affine connection* on the manifold, as well as a path which connects the two points.

The identification of every tangent space on $\mathbb{R}^N$ with $\mathfrak{r}$ allows for a simple definition of vector fields and differential forms. A *vector field* is a smooth function $F : \mathbb{R}^N \to \mathfrak{r}$ which assigns a tangent vector to every point on $\mathbb{R}^N$.

Next, the linear dual of a tangent space of X at x is called the *cotangent space of X at x*, denoted $T_x^*X := (T_xX)^*$. For $X = \mathbb{R}^N$, the cotangent space at $x \in \mathbb{R}^N$ is the N-dimensional vector space spanned by the basis vectors

$$dx_1, \dots dx_N,$$

which are dual to the basis vectors in Equation (2.1.1). As with tangent spaces, we can identify the cotangent spaces over any point with the dual of $\mathfrak{r}$, denoted $\mathfrak{r}^*$. A *differential 1-form* on $\mathbb{R}^N$ is a smooth function $\omega : \mathbb{R}^N \to \mathfrak{r}^*$. Thus, a 1-form on $\mathbb{R}^N$ can be written as

$$\omega(x_1, \dots, x_N) = \sum_{i=1}^{N} f_i(x_1, \dots, x_N) dx_i, \tag{2.1.2}$$

where each $f_i : \mathbb{R}^N \to \mathbb{R}$ is a smooth function. The space of all 1-forms is denoted $\Omega^1(\mathbb{R}^N)$.

Recall that the k^{th} *exterior power* of a vector space V, denoted $\bigwedge^k V$, is the vector space spanned by the elements

$$v_1 \wedge \dots \wedge v_k,$$

where $v_i \in V$. The wedge product satisfies the alternating property: for any permutation $\sigma \in \Sigma_k$,

$$v_{\sigma(1)} \wedge \dots \wedge v_{\sigma(k)} = \operatorname{sgn}(\sigma) v_1 \wedge \dots \wedge v_k. \tag{2.1.3}$$

In particular, the k^{th} exterior power of $\mathfrak{r}^*$, denoted $\bigwedge^k \mathfrak{r}^*$, is the vector space with the basis vectors

$$dx_{i_1} \wedge \dots \wedge dx_{i_k},$$

where $1 \leq i_1 < \dots < i_k \leq N$. Then, a *differential k-form* on $\mathbb{R}^N$ is a smooth function $\omega : \mathbb{R}^N \to \bigwedge^k \mathfrak{r}^*$, which can be written as

$$\omega(x_1, \dots, x_N) = \sum_{I=(i_1<\dots<i_k)} f_I(x_1, \dots, x_N) dx_{i_1} \wedge \dots \wedge dx_{i_k}, \tag{2.1.4}$$

where the sum is taken over all indices where $1 \leq i_1 < \dots < i_k \leq N$. The space of all k-forms is denoted $\Omega^k(\mathbb{R}^N)$.

Intuitively, a differential k-form assigns an infinitesimal signed k-volume element to every point in $\mathbb{R}^N$, where the orientation is taken into account by the alternating property in Equation (2.1.3). The signed k-volume element represents some quantity of interest, and we can integrate the k-form over a k-dimensional region $C \subset \mathbb{R}^N$ (or the manifold X in general) to obtain the total quantity.

A k-form $\nu \in \Omega^k(C)$, where $C \subset \mathbb{R}^k$ can be written as

$$\nu(x_1, \ldots, x_k) = g(x_k, \ldots, x_k) dx_1 \wedge \ldots \wedge dx_k, \tag{2.1.5}$$

since $\bigwedge^k(\mathfrak{r}^k)^*$ is 1-dimensional. The integral of an k-form over an k-dimensional region $C \subset \mathbb{R}^k$ is defined to be

$$\int_C \nu := \int_C g(x_1, \ldots, x_k) dx_1 \ldots dx_k, \tag{2.1.6}$$

where the right hand side is interpreted as an ordinary multidimensional integral.

The integral of a k-form ω on $\mathbb{R}^N$ over a k-dimensional region $C \subset \mathbb{R}^N$ requires a parametrization of the region C by a function $\gamma = (\gamma_1, \ldots, \gamma_N) : D \to \mathbb{R}^N$, where $D \subset \mathbb{R}^k$ and the image is C. The integral can be defined using the *pullback of ω along γ*, denoted $\gamma^*\omega \in \Omega^k(D)$. Let $\mathbf{t} = (t_1, \ldots, t_k)$ describe the coordinates of D. Suppose ω is given as in Equation (2.1.4), then the pullback $\gamma^*\omega$ is defined by

$$\gamma^*\omega(t_1, \ldots, t_k) = \sum_{I=(i_1<\ldots<i_k)} (f_I \circ \gamma)(\mathbf{t}) J[\gamma_I](\mathbf{t}) dt_1 \wedge \ldots \wedge dt_k, \tag{2.1.7}$$

where

$$J[\gamma_I] = \begin{vmatrix} \frac{\partial \gamma_{i_1}}{\partial t_1} & \ldots & \frac{\partial \gamma_{i_1}}{\partial t_k} \\ \vdots & \ddots & \vdots \\ \frac{\partial \gamma_{i_k}}{\partial t_1} & \ldots & \frac{\partial \gamma_{i_d}}{\partial t_k} \end{vmatrix} \tag{2.1.8}$$

is the determinant of the Jacobian of $\gamma_I = (\gamma_{i_1}, \ldots, \gamma_{i_k}) : D \to \mathbb{R}^k$. Then, since $\gamma^*\omega \in \Omega^k(D)$, where $D \subset \mathbb{R}^k$, the integral is given by Equation (2.1.6).

To summarize, the perspective on differential forms, which we state for an arbitrary manifold X, is the following:

A differential k-form $\omega \in \Omega^k(X)$ assigns an infinitesimal signed k-volume elements to every point in X. Integrals of ω can be computed using the pullback $\gamma^*\omega$ along the map $\gamma : D \to X$ which parametrizes the domain of integration.

This perspective places the pullback of forms along maps with Euclidean domains as the central object needed to perform explicit computations such as integration. In fact, a k-form $\omega \in \Omega^k(X)$ can be *defined* by the forms $\{\phi^*\omega\}_\phi$ over all smooth maps $\phi : D \to X$, in which $D \subset \mathbb{R}^n$, where the dimension of the Euclidean space $n \in \mathbb{N}$ is permitted to vary. If we define a form using $\{\phi^*\omega\}_\phi$ over all such ϕ, we must ensure that the forms satisfy compatibility conditions, which is discussed in the next section.

This definition of a differential form in terms of pullbacks along maps with Euclidean domains allows an immediate generalization to spaces more general than manifolds, such as the path space $P\mathbb{R}^N$. As we will make precise in the next section, a k-form on $P\mathbb{R}^N$ is defined as a collection of forms $\{\omega_\phi\}_\phi$, where $\omega_\phi \in \Omega^k(D)$ and $\phi : D \to P\mathbb{R}^N$ is a smooth map. This requires the notion of smooth maps $\phi : D \to P\mathbb{R}^N$, which will be achieved by viewing $P\mathbb{R}^N$ as a smooth space using Chen's notion of a differentiable space.

Once this is established, we can treat differential forms on $P\mathbb{R}^N$ is the same manner as forms on manifolds. In particular, a k-form on $P\mathbb{R}^N$ assigns an infinitesimal signed k-volume element to each point in $P\mathbb{R}^N$. However, we describe the infinitesimal volume element via pullbacks to Euclidean domains, which can then be integrated as differential forms on $\mathbb{R}^N$.

2.2 CHEN'S DIFFERENTIABLE SPACES

In this section, we introduce Chen's notion of a differentiable space [Che77], which provides notions of smooth maps for spaces much more general than manifolds. A smooth structure is defined on an N-dimensional manifold X by using a collection of charts, which are homeomorphisms $\phi : U \to X$, where $U \subset \mathbb{R}^N$ is open. Charts can be viewed as probes into the local structure of a manifold, and allow us to define smooth maps between manifolds by exploiting the well-understood notion of smoothness for Euclidean spaces. Differentiable spaces are defined similarly by using a collection of *plots*, the analogue of charts, which relax the homeomorphism condition of charts, and also allow for maps from convex Euclidean regions of arbitrary dimension.

Definition 2.2.1. A *differentiable space* is a set X equipped with, for every Euclidean convex set $C \subseteq \mathbb{R}^n$ with nonempty interior and for any dimension n, a collection of functions $\phi : C \to X$ called *plots*, satisfying the following:

1. (Closure under pullback) If ϕ is a plot and $f : C' \to C$ is a smooth map between convex subsets $C' \subset \mathbb{R}^{n'}$ and $C \subset \mathbb{R}^n$, then $\phi \circ f$ is a plot.

2. (Covering condition) Suppose the collection of convex sets $\{C_j\}$ form an open cover of the convex set C, with inclusions $i_j : C_j \hookrightarrow C$. If ϕi_j is a plot for all j, then ϕ is a plot.

3. (Constant plots) Every map $f : \mathbb{R}^0 \to X$ is a plot.

A function $f : X \to Y$ between two differentiable spaces is a *smooth map* if for every plot ϕ in X, the composite $f \circ \phi$ is a plot in Y. The category of differentiable spaces and smooth maps is denoted Diff.

Remark 2.2.1. We can also define a differentiable space as a *generalized space* from the perspective of category theory. Let Conv be the category where the objects are all convex subsets of $\mathbb{R}^n$ for any $n \geq 0$, and the morphisms are smooth maps. There is a natural Grothendieck toplogy on Conv defined using open coverings via inclusion maps. Thus, Conv is a concrete site. Then, a differentiable space is a concrete sheaf $X : \mathsf{Conv}^{\mathrm{op}} \to \mathsf{Set}$. In fact, given X as a set, a differentiable structure on X is a subsheaf of

$$\mathsf{Set}(-, X) : \mathsf{Conv}^{\mathrm{op}} \to \mathsf{Set}, \tag{2.2.1}$$

which provides the collection of plots in Definition 2.2.1. The closure under pullback and covering conditions are inherent in the definition of a sheaf, and the constant plot condition is equivalent to the sheaf being concrete. For further details on this perspective, see [BH11],

Note that *a priori*, a differentiable space is simply a set and is not equipped with a topology. However, a differentiable space inherits a topology by considering the final topology with respect to all plots: the finest topology on X such that all plots are continuous. Now, we consider several examples and constructions of differentiable spaces. These examples are taken from [BH11].

Example 2.2.1. Any smooth manifold X is a differentiable space by letting $\phi : C \to X$ be a plot if and only if ϕ is smooth (in the sense of manifolds). This can be achieved by letting every chart from a convex domain be a plot, and adding all other plots required by Definition 2.2.1.

Example 2.2.2. Any subset $Y \subset X$ of a differentiable space X is a differentiable space where $\phi : C \to Y$ is a plot if and only if its composite with the inclusion $\iota : Y \hookrightarrow X$ is a plot in X. This is called the *subspace smooth structure*. Note that the subspace smooth structure can used to define equalizers in Diff.

Example 2.2.3. Suppose X is a differentiable space and $\sim$ is any equivalence relation on X. The quotient space $Y = X/\sim$ is a differentiable space where $\phi : C \to Y$ is a plot if and only if there exists an open cover $\{C_i\}$ fo C and a collection of plots in X

$$\{\phi_i : C_i \to X\}_i$$

such that the following diagram commutes:

$$\begin{array}{ccc} C_i & \xrightarrow{\phi_i} & X \\ {\scriptstyle \iota_i}\downarrow & & \downarrow{\scriptstyle q} \\ C & \xrightarrow[\phi]{} & Y, \end{array}$$

where $q : X \to Y$ is the quotient map and $\iota_i : C_i \to C$ is the inclusion. This is the *quotient smooth structure*. Note tha the quotient smooth structure is used to define coequalizers in Diff.

Example 2.2.4. Suppose X and Y are differentiable spaces. The coproduct (disjoint union) $X \coprod Y$ is a differentiable space where $\phi : C \to X \coprod Y$ if and only if ϕ is the composite of a plot of X or Y, and one of the inclusion maps

$$\iota_X : X \hookrightarrow X \coprod Y, \qquad \iota_Y : Y \hookrightarrow X \coprod Y.$$

This is called the *coproduct smooth structure*.

Example 2.2.5. Suppose X and Y are differentiable spaces. The product $X \times Y$ is a differentiable space where $\phi : C \to X \times Y$ is a plot if and only if composition with the projections

$$\pi_X : X \times Y \to X, \qquad \pi_Y : X \times Y \to Y$$

are plots in X and Y respectively. This is called the *product smooth structure*.

Example 2.2.6. Suppose X and Y are differentiable spaces. The set $\mathsf{Diff}(X, Y)$ of smooth maps between X and Y is a differentiable space, where $\phi : C \to \mathsf{Diff}(X, Y)$ is a plot if and only if the adjoint function $\tilde{\phi} : C \times X \to Y$ defined by

$$\tilde{\phi}(x, y) = \phi(x)(y),$$

is smooth. Note in particular that this implies that the smooth path space $\mathsf{Diff}([0,1], \mathbb{R}^N)$ and the smooth mapping space $\mathsf{Diff}(\square^d, \mathbb{R}^N)$, where $\square^d := [0,1]^d$, are examples of differentiable spaces.

The final two examples lead to the following property of differentiable spaces.

Proposition 2.2.1 ([BH11])**.** *Let X, Y, and Z be differentiable spaces. The map*

$$\begin{aligned} \Phi : \mathsf{Diff}(X \times Y, Z) &\to \mathsf{Diff}(X, \mathsf{Diff}(Y, Z)) \\ \tilde{f} &\mapsto f \end{aligned} \tag{2.2.2}$$

defined by $f(x)(y) = \tilde{f}(x, y)$ is smooth and has a smooth inverse. Thus, the category Diff *is cartesian closed.*

There is one more category-theoretic property which we will need in the mapping space construction.

Proposition 2.2.2 ([BH11])**.** *The category* Diff *is complete (has all small limits) and cocomplete (has all small colimits).*

Next, we will define differential forms on differentiable spaces.

Definition 2.2.2. A *differential k-form* ω on a differentiable space X is an assignment of a k-form ω_ϕ on C to each plot $\phi : C \to X$ such that if $f : C' \to C$ is smooth a smooth map between convex subsets $C' \subset \mathbb{R}^{n'}$ and $C \subset \mathbb{R}^n$, then $\omega_{\phi \circ f} = f^* \omega_\phi$. The collection of k-forms on X is denoted $\Omega^k(X)$, and the graded collection of all forms on X is $\Omega^\bullet(X)$.

The linear structure, wedge product, and exterior derivative are all defined on differential forms plot-wise. Given $\omega, \omega_1, \omega_2 \in \Omega^\bullet(X)$, $\lambda \in \mathbb{R}$, and any plot $\phi : C \to X$,

- $(\omega_1 + \lambda \omega_2)_\phi = (\omega_1)_\phi + \lambda (\omega_2)_\phi$,
- $(\omega_1 \wedge \omega_2)_\phi = (\omega_1)_\phi \wedge (\omega_2)_\phi$, and
- $(d\omega)_\phi = d\omega_\phi$.

2.3 THE CHEN COMPLEX

In this section, we will introduce Chen's iterated integral cochain algebra for the path space PX of a manifold X. Chen's construction can be summarized as the composition of the following two maps

$$\Omega^\bullet(X)^{\otimes m} \xrightarrow{\mathrm{ev}_m^*} \Omega^\bullet(\Delta^m \times PX) \xrightarrow{\int_{\Delta^m}} \mathrm{Chen}^\bullet(PX) \subset \Omega^\bullet(PX). \tag{2.3.1}$$

In other words, a q-form in $\mathrm{Chen}^q(PX)$ is constructed following three steps.

1. Choose m differential forms $\omega_1, \ldots, \omega_m$ such that $\omega_j \in \Omega^{q_j}(X)$ such that

$$q = \sum_{j=1}^{m} q_j - m. \tag{2.3.2}$$

2. Compute the pullback of $\omega_1 \otimes \ldots \otimes \omega_m \in \Omega^\bullet(X^m)$ along the evaluation map $\mathrm{ev}_m : \Delta^m \times PX \to X^m$ defined by

$$\mathrm{ev}_k((t_1, \ldots, t_k), \gamma) \coloneqq (\gamma(t_1), \ldots, \gamma(t_k)). \tag{2.3.3}$$

3. Integrate the resulting form on $\Delta^m \times PX$ along Δ^m to obtain a form on PX, which is denoted

$$\int \omega_1 \ldots \omega_k \in \mathrm{Chen}^q(PX).$$

Because differential forms on differentiable spaces such as PX are defined plotwise, we fix a plot $\phi : C \to PX$, where $C \subset \mathbb{R}^n$. Then, the adjoint map $\tilde{\phi} : [0,1] \times C \to X$ is a plot on X by the smooth structure on PX in Example 2.2.6. The resulting differential form with respect to the plot ϕ, $\left(\int \omega_1 \dots \omega_m\right)_\phi$, is then given by the composition of the following maps for differential forms on ordinary manifolds

$$\Omega^\bullet(X)^{\otimes m} \xrightarrow{\mathrm{ev}_m^*} \Omega^\bullet(\Delta^m \times C) \xrightarrow{\int_{\Delta^m}} \Omega^\bullet(C). \tag{2.3.4}$$

Given the differential forms $\omega_1, \dots, \omega_m$ where $\omega_j \in \Omega^{q_j}(X)$, consider the pullback of each ω_i along $\tilde{\phi}$, which is a form $\tilde{\phi}^*\omega_j \in \Omega^{q_j}([0,1] \times C)$. We will use the coordinates $t \in [0,1]$ and $\mathbf{x} = (x_1, \dots, x_n) \in C$. We say that a form $\nu \in \Omega^k([0,1] \times C)$ is a $\Omega^m(C)$*-valued function on* $[0,1]$ if

$$\nu(t, \mathbf{x}) = \sum_{I=(i_1<\dots<i_m)} g_I(t, \mathbf{x}) dx_{i_1} \wedge \dots \wedge dx_{i_k}.$$

In other words, ν is a $\Omega^k(C)$-valued function on $[0,1]$ if there is no dt component in the differential form. The pullback can be decomposed as

$$\tilde{\phi}^*\omega_j(t, \mathbf{x}) = \omega_j'(t, \mathbf{x}) \wedge dt + \omega_j''(t, \mathbf{x}), \tag{2.3.5}$$

where ω_j' is a $\Omega^{q_j-1}(C)$-valued function on $[0,1]$ and ω_j'' is a $\Omega^{q_i}(C)$-valued function on $[0,1]$. The pullback along ev_m in Equation (2.3.4) can be written as

$$\mathrm{ev}_k^*(\omega_1 \otimes \dots \otimes \omega_m) = \tilde{\phi}^*\omega_1(t_1, \mathbf{x}) \wedge \dots \wedge \tilde{\phi}^*\omega_m(t_m, \mathbf{x}). \tag{2.3.6}$$

Integrating this pullback along the Δ^m coordinates, we obtain the expression for the Chen q-form

$$\left(\int \omega_1 \dots \omega_k\right)_\phi = \int_{\Delta^m} \omega'(t_1, \mathbf{x}) \wedge \dots \wedge \omega'(t_m, \mathbf{x}) dt_1 \wedge \dots \wedge dt_m. \tag{2.3.7}$$

Note that because we have integrated out k of the dimensions, the degree of this form is

$$q = \sum_{j=1}^m q_j - m.$$

In order to recover the path signature, we will set $X = \mathbb{R}^N$, and consider the Chen 0-forms in $\mathrm{Chen}^0(P\mathbb{R}^N)$. We begin by choosing m differential forms $\omega_1, \dots, \omega_m$ with $\omega_k \in \Omega^{q_j}(\mathbb{R}^N)$ such that $\sum_{j=1}^m q_= m$. The only nontrivial 0-forms are obtained by letting

each ω_i be a 1-form. Furthermore, we choose each ω_i to be a standard 1-form on $\mathbb{R}^N$, so we choose $\omega_j = dx_{i_j}$, where $I = (i_1, \ldots, i_m)$ is a multi-index valued in $[N]$.

Next, we must fix a plot in $P\mathbb{R}^N$. Because we are constructing a 0-form, we may use the constant plot $\gamma : * \to P\mathbb{R}^N$, where the adjoint $\tilde{\gamma} = (\tilde{\gamma}_1, \ldots, \tilde{\gamma}_N) : [0,1] \to \mathbb{R}^N$ is simply a path. In this case, the pullback in Equation (2.3.5) is

$$\tilde{\gamma}^* dx_{i_j}(t) = \tilde{\gamma}'_{i_j}(t)dt,$$

where the prime denotes the time derivative of $\tilde{\gamma}$ here. Then, then resulting Chen 0-form is

$$\left(\int dx_{i_1} \ldots dx_{i_m}\right)_{\tilde{\gamma}} = \int_{\Delta^m} \tilde{\gamma}'_{i_1}(t_1) \ldots \tilde{\gamma}'_{i_m}(t_m) dt_1 \ldots dt_m, \tag{2.3.8}$$

which is equivalent to the path signature $S^I(\tilde{\gamma})$.

Part II

GENERALIZATIONS OF THE CODOMAIN

3

PATH SIGNATURES ON LIE GROUPS

3.1 INTRODUCTION

Time series data is ubiquitous in modern data science, and may take values in a variety of forms. Perhaps the most common is a collection of simultaneous multivariate real-valued time series $\{\gamma^i\}_{i=1}^N$, where $\gamma^i : [0,1] \to \mathbb{R}$. In this case, we may consider the entire collection $\gamma = (\gamma^1, \dots, \gamma^N)$ as a path through Euclidean space, $\gamma : [0,1] \to \mathbb{R}^N$. The *path signature* is a feature set that completely characterizes such paths, and has recently been applied to several tasks in machine learning [CK16; Lyo14]. Recent work has provided the path signature with strong theoretical properties; namely that it is a universal and characteristic kernel for time series in Euclidean space $\mathbb{R}^N$ [CO18].

However, in many scenarios, the data may have some geometric constraints, and may be better represented by elements of a (non-Euclidean) manifold. In this case, the time-varying data can be modelled as path on such a manifold, rather than on Euclidean space. *Lie groups* are smooth manifolds equipped with a compatible group structure. Paths (or time series) valued in Lie groups model a number of natural phenomena, including the following.

- The *special Euclidean group* $SE(n)$ is the Lie group of all rigid body motions in $\mathbb{R}^n$. The group $SE(3)$ is often used to model the position and pose of a rigid body, such as a component of a robotic arm or an element of a drone swarm, with k such components or elements collectively giving rise to a path $SE(3)^k$ [Selo4].
- The *special orthogonal group* $SO(n)$ is the Lie group of all rotations in $\mathbb{R}^n$; this is a Lie subgroup of $SE(n)$. The Lie group $SO(3)^k$ has recently been used to represent the pose of a human by recording the relative rotations of k pairs of body parts [VC16]. Thus, human movement can be represented as a path in $SO(3)^k$. This representation has been used in the computer vision problem of human action recognition, and Lie group methods have achieved state-of-the-art results in this domain [Hua+17].

- The state of an oscillator may be described as an element of the *circle* S^1, and collective behavior of a network of oscillators can be describe by an element of the *n-torus*, $T^n = (S^1)^n$. The time evolution of oscillator networks can therefore be modelled as a path on T^n [Stroo].
- The Euclidean space $\mathbb{R}^N$ is the simplest example of a Lie group, where the group operation is addition. The classical path signature for Euclidean space can be viewed as a special case of path signatures on Lie groups.

In this chapter, we extend path signatures to time series valued in Lie groups, and show that this extension is also a universal and characteristic kernel.

3.1.1 *Contributions*

We lift the theory of path signatures for time series valued in Euclidean space to the setting of time series valued in Lie groups, restricting ourselves to the class of piecewise regular paths on Lie groups.

Definition 3.1.1. Let G be a Lie group. A path $\gamma : [a, b] \to G$ is *regular* if γ'_t is continuous and nonvanishing on the entire interval $[a, b]$. Such a path is *piecewise regular* if there exists a partition $a = t_0 < t_1 < \ldots < t_n = b$ such that γ is regular on each open subinterval (t_i, t_{i+1}) for all i. The *pathspace* – the space of all piecewise regular paths on the unit interval, $\gamma : [0, 1] \to G$ – will be denoted $C^{pr}(G)$.

Let G be a Lie group of dimension N, and let $\mathfrak{g}$ be its Lie algebra (the tangent space at the identity). We denote the underlying vector space of $\mathfrak{g}$ by $\overline{\mathfrak{g}} \cong \mathbb{R}^N$. The *path signature* is a function on paths,

$$S : C^{pr}(G) \to T((\overline{\mathfrak{g}})),$$

valued in a formal power series of tensors, $T((\overline{\mathfrak{g}}))$, where we may view the coefficients as descriptors (or features) of the underlying path (or time series). Path signatures for general manifolds were originally defined by [Che58], but not in a manner conducive to data analysis. Path signatures for Lie group valued data have been previously considered by [CLT19] in a preliminary empirical study, showing promising qualitative classification results, but extensions of theoretical results and detailed quantitative comparisons were not provided. This paper gives a computationally clean derivation for path signatures on Lie groups tuned for use in data analysis, and provides a thorough discussion of its theoretical properties in the context of kernel methods.

Our generalization is designed to be analogous to the Euclidean case as much as possible, for ease of applicability. For example, the definition of the path signature for $\gamma : [0, 1] \to G$

depends only on the derivative $\gamma' : [0,1] \to \mathfrak{g}$. We exploit one of the key properties of Lie groups — that tangent vectors at a point correspond to elements of its Lie algebra $\mathfrak{g}$, a vector space. This will permit a signature construction making use of iterated integrals as per the Euclidean case.

In the Euclidean case $G = \mathbb{R}^N$, the Lie group is often conflated with its Lie algebra $\mathfrak{r} = \mathbb{R}^N$, and the fact that the integration is performed in the Lie algebra is often not made. By clarifying and emphasizing this point, the generalization to Lie groups illuminates understanding of the classical Euclidean case.

From a machine learning perspective, the basic properties of the path signature as a feature map provide several benefits.

- The signature is a feature set for a path as a whole, and can be used to compare time series with varying numbers of time points.
- Defined as iterated line integrals, the path signature is invariant under reparametrization, and thus only depends on the *order* in which events occur.
- The signature is left translation invariant, meaning the signatures of paths that differ by a constant element $g \in G$ will be the same. This implies that the signature only depends on the dynamics of the time series and is unconcerned with the initial point.
- The antisymmetrization of the second degree signature tensor can be viewed as an indicator of lead-lag behavior in the time series. In the case of Lie groups, the interpretation will be considered in terms of left-invariant vector fields.

However, the most crucial property is that the path signature fully characterizes paths up to tree-like equivalence; that is, the map S is injective, up to quotienting $C^{pr}(G)$ out by an equivalence relation. This fact is originally due to [Che58] for the case of piecewise regular paths on Lie groups, and later generalized by [HL10] to the case of bounded variation paths in $\mathbb{R}^n$.

Our main contribution is to apply this injectivity result to prove that a normalized variant of the signature, $S : C^{pr}(G) \to T((\overline{\mathfrak{g}}))$, is a *universal* and *characteristic* feature map for time series in G, when we equip $T((\overline{\mathfrak{g}}))$ with the structure of a Hilbert space. This is proved in Section 3.4.1. This was originally shown for the Euclidean case by [CO18]. Such feature maps can be used to two large classes of machine learning problems, in the context of kernel methods.

1. **(Studying functions on $C^{pr}(G)$)** Solving a classification problem on $C^{pr}(G)$ can be reduced to finding a function $f : C^{pr}(G) \to \mathbb{R}$ such that the level set $f = 0$ provides the decision boundary. The *universality* of the normalized signature map states that any continuous bounded function $f : C^{pr}(G) \to \mathbb{R}$ can be approximated using a

linear functional $f(\cdot) \approx \langle \ell, \tilde{S}(\cdot) \rangle$. This allows us to reduce a nonlinear optimization problem into a linear one, greatly reducing the complexity.

2. **(Studying measures on $C^{pr}(G)$)** Two-sample hypothesis testing on $C^{pr}(G)$ requires the computation of a set of statistics that is rich enough to distinguish any two probability measures on $C^{pr}(G)$. The *characteristicness* of the normalized signature map states that the kernel mean embedding (KME) is injective with respect to the normalized signature

$$\overline{\Phi} : \mathcal{M}(C^{pr}(G)) \to T((\overline{\mathfrak{g}})), \quad \overline{\Phi}(\mu) = \mathbb{E}_{\mu}[S],$$

where $\mathcal{M}(C^{pr}(G))$ denotes all finite regular Borel measures on $C^{pr}(G)$, and S is appropriately normalized. This allows us to consider probability measures as elements of a linear space; furthermore, the norm induced by the Hilbert space structure coincides with the maximum mean discrepancy (MMD) between measures.

We perform two experiments that demonstrate the efficacy of the path signature for these two classes of problems. First, we consider the computer vision problem of human action recognition in Section 3.5.1. We show that the path signature method is much easier to use than shallow learning methods previously applied to this problem [VAC14; VC16] while providing comparable results. Second, in Section 3.5.2, we consider a hypothesis testing problem for simulated random walks on the Lie group $SO(3)$. Here, we show that the Lie group valued path signature vastly outperforms the Euclidean path signature.

Along the way, we will establish extensions of other properties of the path signatures to Lie groups and discuss several concepts related to path signatures and data analysis on Lie groups more broadly. A summary of these contributions is given below.

1. We provide a detailed exposition of Lie group valued time series, and discuss a notion of scaling for such time series in Section 3.2.2. Scaling of data is sometimes required when the data needs to be normalized, and we discuss how scaling affects the path signature in Section 3.3.1. We also discuss the continuous interpretation of discrete time series on Lie groups in Section 3.2.3.

2. For G an N-dimensional Lie group, we give a signature-preserving bijection between $C^{pr}(G)$ and $C^{pr}(\mathbb{R}^N)$ in Section 3.3.2, which provides a Euclidean representation of Lie group valued time series. This bijection allows the exportation of Euclidean data analysis tools to Lie group valued data. With the metric introduced in Section 3.3.3, this bijection is an isometry.

3. It is well known that the Euclidean path signature is equivariant with respect to linear transformations [FV10]. We show that path signatures are equivariant under Lie

group homomorphisms in general. Namely, given a homomorphism of Lie groups $F : G_1 \to G_2$, where $\mathfrak{g}_1$ and $\mathfrak{g}_2$ are the respective Lie algebras, we define the action of this homomorphism on the tensor algebra $F_* : T((\overline{\mathfrak{g}}_1)) \to T((\overline{\mathfrak{g}}_2))$, and show in Section 3.3.4 that

$$S(F\gamma) = F_* S(\gamma)$$

for all $\gamma \in C^{pr}(G)$.

4. An important feature of the path signature is the interpretability of lower level signature terms. We discuss the extension of the lead-lag interpretation of second level signature terms for Euclidean paths, as well as a topological interpretation of the first level signature terms for abelian Lie groups in Section 3.3.6.

5. Path transformations, such as appending the time parameter or using a sliding window, are often used as a preprocessing step for Euclidean path signatures [CK16]. We discuss these transformations in the context of breaking reparametrization or left-translation invariance in Section 1.3.6. Empirical studies [Fer21] have shown that the sliding window transformation (also called the lead-lag transformation) provides good classification results, despite the lack of a theoretical explanation. We propose one explanation, which is that the sliding window transformation breaks left-translation invariance, and we provide empirical evidence in the experiments in Section 3.5.1.

6. We provide both algorithmic details and a Julia package for the computation of path signatures valued in Lie groups, which can be found at `https://github.com/ldarrick/PathSignatures`.

3.1.2 *Previous and Related Work*

The concept of path signatures is relatively new in data science and machine learning [Lyo14; CK16; GL20], but has deep roots in topology and geometry. Chen originally defined the path signature for piecewise regular paths on manifolds and proved several basic properties in a sequence of papers [Che54; Che57; Che58]. He later studied the geometry and topology of path spaces and loop spaces by constructing a rational cochain model of these spaces, in which path signatures constitute 0-cochains [Che77].

[Lyo98] developed the concept of the path signature in a different direction, using the path signature as a construction to lift bounded variation paths on $\mathbb{R}^N$ to paths of power series of tensors $T((\mathbb{R}^N))$. This initiated the study of *rough paths*, which can be thought of as a generalization of the path signature to highly irregular paths. This theory was

then used to study stochastic processes and stochastic differential equations [LQ07; LCL07; FV10].

Within machine learning, path signatures have been used to study real-valued time series data in a variety of settings. Examples can be found in the study of financial time series [Gyu+13; LNO14], handwritten character recognition [YJL16], human action recognition using position data [Yan+19], identifying psychological or neurological disorders [Moo+19; Zim+18; PA+18] and featurizing the output of persistent homology in topological data analysis [CNO20]. Additionally, experiments with path signatures on Lie groups have previously been performed [CLT19], though theoretical results were not provided, and thus suggests further study.

The theoretical aspects of path signatures in the context of kernel methods were developed in [KO19] and [CO18]. The present paper is largely inspired by these two papers. The concept of using the path signature as a kernel for time series was first proposed in [KO19], and efficient algorithms for computing the kernel were developed. The path signature for Euclidean space was shown to be a universal and characteristic feature map in [CO18]. This exploits the recently formalized duality between universal and characteristic kernels in [SS18].

It is well known that Euclidean path signatures are translation invariant, and we will show that Lie group path signatures are left translation invariant. [DR19] has considered the related problem of determining the Euclidean path signature terms which are invariant under some matrix Lie group action.

We begin in Section 3.2 by reviewing basic facts on Lie groups and Lie algebras, and provide an exposition on continuous and discrete time series on Lie groups. We then define the path signature for Lie groups in Section 3.3, and discuss the bijection between $C^{pr}(G)$ and $C^{pr}(\mathbb{R}^N)$, the equivariance of the path signature, detecting lead-lag behavior in time series, and path transformations. In Section 3.4, we provide a brief overview of kernel methods and prove our main result, which shows that the path signature kernel is universal and characteristic. Finally, in Section 3.5, we apply the path signature on Lie groups to a human action classification problem and a hypothesis testing problem involving random walks on $SO(3)$.

3.1.3 *Notation*

Throughout this paper, we will denote the time parameter for a path $\gamma : [0,1] \to G$ using a subscript t, meaning $\gamma_t := \gamma(t)$. Derivatives are shown using the prime notation, as in $\gamma'_t := \frac{d\gamma}{dt}(t)$. If we have a path in Euclidean space $\alpha : [0,1] \to \mathbb{R}^N$, we will use superscripts to represent the components, such as $\alpha = (\alpha^1, \alpha^2, \ldots, \alpha^N)$. If G is a Lie group, we will use $\mathfrak{g}$ to denote its Lie algebra and use $\bar{\mathfrak{g}}$ to be the underlying vector space of $\mathfrak{g}$ (forgetting the Lie bracket structure).

Continuous paths will often be denoted using the lowercase Greek symbols α, β, γ, and the space of all piecewise regular paths in G is denoted $C^{pr}(G)$. For $T \in \mathbb{N}$, we let $[T] = \{1, \dots, T\}$ denote the finite set of integers up to T. Discrete time series will be distinguished using the hat notation $\hat{\gamma} : [T] \to G$, and the space of all discrete time series in G will be denoted $\mathrm{Seq}(G)$.

There are also several parameters that will be used consistently throughout the paper. Unless otherwise specified, we reserve the following symbols for the given meaning.

- N is the dimension of the Lie group G that paths take values in;
- $T + 1$ is the length of a discrete time series (so that the discrete derivative will be of length T);
- M is the level of the truncated signature.

3.2 LIE GROUPS, PATHS, AND TIME SERIES

We begin this section by recalling several basic facts about Lie groups [AB15], followed by paths on Lie groups and the interpretation of sampled time series on Lie groups, stressing the differences from sampled time series on Euclidean space.

3.2.1 *A Review of Lie Groups*

Recall that a *Lie group* G is a smooth manifold with a group structure such that the multiplication and inversion maps are both smooth. Let $g_0 \in G$. The *left translation map by* g_0, written as $L_{g_0} : G \to G$, is defined to be $L_{g_0}(g) = g_0 g$. The *right translation map* $R_{g_0} : G \to G$ is defined analogously. This induces a mapping on tangent spaces $L_{g_0 *} : T_g G \to T_{g_0 g} G$. A vector field X on G is called *left-invariant* if

$$L_{g_0 *} X(g) = X(L_{g_0} g) = X(g_0 g)$$

for all $g_0, g \in G$. This implies that all left-invariant vector fields X are defined by their value at the identity $e \in G$,

$$X(g) = L_{g *} X(e),$$

and thus, we obtain a one-to-one correspondence between left-invariant vector fields and the tangent space at the identity, which we denote by $\mathfrak{g} := T_e G$. Vector fields act on smooth functions $f : G \to \mathbb{R}$, and we define an operation of left-invariant vector fields X and Y by

$$[X, Y](f) := X(Y(f)) - Y(X(f)),$$

where $[X, Y]$ is also left-invariant. This provides $\mathfrak{g}$ with the structure of a Lie algebra, where the Lie bracket $[\cdot, \cdot] : \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}$ is a bilinear mapping such that for all $X, Y, Z \in \mathfrak{g}$

$$[X, Y] = -[X, Y]$$
$$[X, [Y, Z]] + [Z, [X, Y]] + [Y, [Z, X]] = 0.$$

Similarly, left translation induces a map $L^*_{g_0} : T^*_{g_0 g}G \to T^*_g G$ on cotangent spaces. A 1-form $\omega \in T^*G$ is called *left-invariant* if

$$L^*_{g_0}\omega(g_0 g) = L^*_{g_0}\omega(L_{g_0} g) = \omega(g)$$

for all $g, g_0 \in G$. Again, we obtain a correspondence between left-invariant 1-forms and the cotangent space at the identity via the property

$$\omega(g) = L^*_{g^{-1}}\omega(e).$$

Thus, we may identify the left-invariant 1-forms by the dual of the Lie algebra, $\mathfrak{g}^*$.

Remark 3.2.1. We can think of tangent vectors (derivatives) of a path $\gamma : [0, T] \to G$ as elements of the Lie algebra $\mathfrak{g}$ in two ways. First, for a tangent vector $v \in T_g G$, we can compute the pushforward the tangent vector along the left multiplication map $L_{g^{-1}} v \in T_e G = \mathfrak{g}$. Second, a basis of the Lie algebra provides a global frame for G, meaning, it provides a basis for $T_g G$ for all g. By considering $v \in T_g G$ in terms of this basis, we may also think of v as an element of $\mathfrak{g}$.

In summary, the structure of the Lie group allows us to consider tangent vectors at *any* point on G using a single vector space: a fact repeatedly used throughout this paper.

Given a left-invariant vector field $X \in \mathfrak{g}$, there exists a unique 1-parameter subgroup $\rho_X : \mathbb{R} \to G$ such that $\rho_X(0) = e$ and $\rho'_X(0) = X(e)$. This is defined by the integral curve of X which passes through the identity at $t = 0$.

Definition 3.2.1. The *Lie exponential map* of G is defined as

$$\exp : \mathfrak{g} \to G, \quad \exp(X) \coloneqq \rho_X(1),$$

where ρ_X is the 1-parameter subgroup defined above.

This exponential map provides a way to move between a Lie group and its Lie algebra.

Proposition 3.2.1. *The exponential map* $\exp : \mathfrak{g} \to G$ *is smooth and* $d(\exp)_0 = \mathrm{id}$. *Thus,* $\exp$ *is a diffeomorphism between an open neighborhood of the origin* $0 \in \mathfrak{g}$ *and an open neighborhood of the identity* $e \in G$.

Thus, if elements are near the origin, we can define an inverse map.

Definition 3.2.2. Suppose $U \subset \mathfrak{g}$ is a neighborhood of the origin such that the exponential map is a diffeomorphism. Let $V = \exp(U)$. The *logarithm map* on V is defined to be

$$\log : V \to \mathfrak{g}, \quad \log(g) := \exp^{-1}(g).$$

A homomorphism of Lie groups $F : G \to H$ is a smooth map which is also a group homomorphism, and a homomorphism of Lie algebras $\phi : \mathfrak{g} \to \mathfrak{h}$ is a linear map that preserves the Lie bracket $F([X, Y]) = [F(X), F(Y)]$ for all $X, Y \in \mathfrak{g}$. A Lie group homomorphism $F : G \to H$ induces a Lie algebra homomorphism $F_* : \mathfrak{g} \to \mathfrak{h}$ between the respective Lie algebras by the induced map between the tangent spaces at the identity $F_* : T_eG \to T_eH$.

Example 3.2.1. The *special orthogonal group* $SO(3)$ — orientation-preserving rotations of $\mathbb{R}^3$ — will be the running example used throughout this paper. This is a matrix Lie group and can be explicitly described as the space of all 3×3 orthogonal matrices ($AA^\intercal = A^\intercal A = I$) with determinant $+1$. The Lie algebra of $SO(3)$ is $\mathfrak{so}(3)$, which consists of all 3×3 skew-symmetric matrices ($B = -B^\intercal$). An explicit basis for $\mathfrak{so}(3)$ is

$$e_1 = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad e_2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \quad e_3 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}.$$

We will denote the duals of these basis vectors to be $\omega_i = e_i \in \mathfrak{g}^*$. For all matrix Lie groups, the Lie exponential and logarithm are simply the matrix exponential and logarithm. Suppose $\theta \in \mathbb{R}$. The exponential map in these three basis directions gives us

$$\exp(\theta e_1) = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

$$\exp(\theta e_2) = \begin{pmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{pmatrix},$$

$$\exp(\theta e_3) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{pmatrix}.$$

These are exactly the rotation matrices about the z, y, and x axes respectively. Therefore, we may think of the basis vectors e_i of the Lie algebra as infinitesimal rotations in the respective

directions. In particular, given a path $\gamma \in P(SO(3))$, the value $\omega_i(\gamma_t')$ corresponds to the infinitesimal rotation of γ at time t in the direction of e_i. If we integrate this over the domain of the path,

$$\int_0^1 \omega_i(\gamma_t')dt,$$

we obtain the cumulative rotation of γ in the direction of e_i over the unit interval. This interpretation will be important to keep in mind when we define the path signature in Section 3.3.

Finally, we briefly discuss the Riemannian structure of Lie groups. Recall that a *Riemannian metric* on a smooth manifold M is the assignment of an inner product $\langle \cdot, \cdot \rangle_p$ to the tangent space T_pM for every point $p \in M$, which varies smoothly. Specifically, this means that if X, Y are smooth vector fields defined on a neighborhood of p, then the map $p \mapsto \langle X_p, Y_p \rangle_p$ is smooth. On a Lie group, we often want a Riemannian metric that is compatible with the algebraic structure of G. A Riemannian metric is *left-invariant* if

$$\langle X, Y \rangle_g = \langle L_{h*}X, L_{h*}Y \rangle_{hg}$$

for all $g, h \in G$ and $X, Y \in T_gG$, and a right-invariant Riemannian metric is defined similarly. Such left-invariant metrics can simply be defined on the tangent space at the identity.

Proposition 3.2.2. *There is a one-to-one correspondence between left-invariant metrics on a Lie group G and inner products on its Lie algebra $\mathfrak{g}$.*

Namely, evaluating the inner product $\langle \, , \, \rangle_g$ simply corresponds to viewing the tangent vectors as elements of the identity, and then evaluating the chosen inner product on $\mathfrak{g}$. We will assume that all Riemannian metrics under discussion are left-invariant, and simply call them Riemannian metrics.

3.2.2 *Paths on Lie Groups*

A Riemannian metric $\langle \cdot, \cdot \rangle$, where we now omit the subscript g since it is left-invariant, provides a notion of length for piecewise regular paths on G. Suppose $\gamma \in C^{pr}(G)$. Then the *length* of γ is defined to be

$$\ell(\gamma) := \int_0^1 \sqrt{\langle \gamma_t', \gamma_t' \rangle} dt.$$

This allows us to define a metric on the Lie group. If $g_1, g_2 \in G$, then the distance between g and h is defined to be the infimum length of paths connecting g_1 and g_2,

$$d(g_1, g_2) := \inf \{\ell(\gamma) : \gamma \in C^{pr}(G),\, \gamma_0 = g_1,\, \gamma_1 = g_2\}.$$

Note that since the Riemannian metric is left invariant, this metric is also left invariant,

$$d(hg_1, hg_2) = d(g_1, g_2),$$

for all $h \in G$. The more familiar notion of length in the path signature literature is the 1-variation of a path.

Definition 3.2.3. Suppose (X, d_X) is a metric space and let $\gamma \in PX$. The *1-variation of* γ on $[0,1]$ is defined as

$$|\gamma|_{1-var} = \sup_{(t_i)} \sum_i d_X(\gamma_{t_i}, \gamma_{t_{i+1}}), \tag{3.2.1}$$

where the sum is taken over all partitions $0 = t_1 \leq \ldots \leq t_n = 1$ of $[0,1]$.

Using the metric induced by the Riemannian metric, we may consider the 1-variation length of paths in G. Under the piecewise regular hypothesis, these two lengths are equivalent.

Lemma 3.2.1 ([Bur15])**.** *Let $\gamma \in C^{pr}(G)$. We have $\ell(\gamma) = |\gamma|_{1-var}$.*

At this point, in the case of paths on Euclidean space, we may use the 1-variation to define a metric on $C_0^{pr}(\mathbb{R}^N)$, which are the paths which start at the origin. Given a Lie group G with a left-invariant Riemannian metric, we could follow the same procedure to obtain a metric space structure on $C_e^{pr}(G)$. However, this is not the metric space structure on $C^{pr}(G)$ that is the most compatible with the path signature. We will defer this discussion until Section 3.3.3.

The path space $C^{pr}(\mathbb{R}^N)$ is endowed with a vector space structure since $\mathbb{R}^N$ itself is a vector space. Similarly, we can endow $C^{pr}(G)$ with a group structure by pointwise multiplication, where the identity is the constant path at the identity, and the inverse to a path $\gamma \in C^{pr}(G)$ is the pointwise inverse. However, we are missing a notion of scaling for paths in $C^{pr}(G)$, and such an operation is important to have in machine learning, since algorithms may require normalization of data. Such a scaling is obtained by proving a correspondence between paths in G and paths in $\mathfrak{g}$, and then transferring the scaling operation from $\mathfrak{g}$ to G.

This is done by considering paths on G from the point of view of differential equations. We have the following existence and uniqueness theorem for first order ordinary differential

equations. Let $C^{pc}(\mathfrak{g})$ denote the space of piecewise continuous paths $\gamma : [0,1] \to \mathfrak{g}$ which are right continuous, meaning $\lim_{t \downarrow t_0} \gamma_t = \gamma_{t_0}$.

Theorem 3.2.1. *Let $f \in C^{pc}(\mathfrak{g})$, so that $f : [0,1] \to \mathfrak{g}$ is piecewise continuous and right continuous, where we consider elements of $\mathfrak{g}$ as left-invariant vector fields. Then, the solution of the first order ODE*

$$\gamma_t' = f_t(\gamma_t), \quad \gamma_0 = g \tag{3.2.2}$$

exists and is unique.

Note that in this theorem, we consider a function $\gamma : [0,1] \to G$ to be a solution of this ODE if the differential equation holds at all points except the points of discontinuity of f. This implies that we can represent piecewise regular paths in G as paths in the Lie algebra $\mathfrak{g}$, along with its initial point. Let $C_g^{pr}(G) \subset C^{pr}(G)$ be defined as

$$C_g^{pr}(G) = \{\gamma \in C^{pr}(G) \,:\, \gamma_0 = g\}.$$

Corollary 3.2.1. *Suppose G is a Lie group and $\mathfrak{g}$ its Lie algebra. The map $\Psi_g : C^{pc}(\mathfrak{g}) \to C_g^{pr}(G)$, which takes $f \in C^{pc}(\mathfrak{g})$ to the solution of the ODE in Equation 3.2.2 with initial condition $\gamma_0 = g$, is a bijection.*

Proof. Firstly, the map Ψ_g is well defined by the existence and uniqueness theorem above. The inverse to Ψ_g can be defined by taking the derivative at every differentiable point. Suppose $\gamma \in C_g^{pr}(G)$, and let $d(\gamma) \subset [0,1]$ denote the set of points such that γ is differentiable. Note that $[0,1] - d(\gamma)$ is a finite set since γ is piecewise regular. Now, define $\Psi_g^{-1}(\gamma)(t) = \gamma_t'$ for all $t \in d(\gamma)$, and at the nondifferentiable points by right continuity

$$\Psi_g^{-1}(\gamma)(t) = \lim_{s \downarrow t} \gamma_s'.$$

This map is well defined: $\Psi_g^{-1}(\gamma)(t)$ is continuous for every $t \in [0,1] - d(\gamma)$, and right continuous by definition. □

We can view $C^{pc}(\mathfrak{g})$ as a Lie algebra, with pointwise vector space operations, and pointwise Lie bracket. Because the group structure of $C^{pr}(G)$ and the Lie algebra structure of $C^{pc}(\mathfrak{g})$ are defined pointwise, the map Ψ_g is compatible with Lie algebra morphisms induced by Lie group morphisms. Namely, if $F : G \to H$ is a Lie group morphism, we obtain a group homomorphism $F : C^{pr}(G) \to C^{pr}(H)$ by applying the map pointwise. Analogously, if $F_* : \mathfrak{g} \to \mathfrak{h}$ is the induced Lie algebra morphism, we obtain a Lie algebra morphism $F_* : C^{pc}(\mathfrak{g}) \to C^{pc}(\mathfrak{h})$. The following lemma is immediate since the group structure on $C^{pr}(G)$ and the Lie algebra structure on $C^{pc}(\mathfrak{g})$ are defined pointwise.

Lemma 3.2.2. *Suppose $F : G \to H$ is a morphism of Lie groups, and $F_* : \mathfrak{g} \to \mathfrak{h}$ is the induced morphism of Lie algebras. Then the following diagram commutes*

The map Ψ_g allows us to view paths on Lie groups as paths in a linear space, while retaining all first order differential information. We can use the fact that many operations for paths on $\mathbb{R}^N$ are defined via operations on the Lie algebra, and thus generalize these operations to Lie groups.

For a path $\alpha \in C^{pr}(\mathbb{R}^N)$ and $\lambda \geq 0$, denote the *vector space scaling* operation as

$$(\lambda\alpha)_t := \lambda\alpha_t.$$

However, another way of viewing the scaling operation for paths that begin at the origin is by scaling in the Lie algebra. Suppose $\lambda \geq 0$, and denote the vector space scaling in a Lie algebra $\mathfrak{g}$ by $c_\lambda : \mathfrak{g} \to \mathfrak{g}$.

Lemma 3.2.3. *Let $\alpha \in C_0^{pr}(\mathbb{R}^N)$. Then*

$$\lambda\alpha = \Psi_0 \circ c_\lambda \circ \Psi_0^{-1}(\alpha).$$

Proof. In $\mathbb{R}^N$, the map Ψ_0 is simply integration in $\mathbb{R}^N$, and Ψ_0^{-1} is differentiation. Thus, we have

$$\begin{aligned}\left(\Psi_0 \circ c_\lambda \circ \Psi_0^{-1}(\alpha)\right)_t &= \int_0^t \lambda\alpha_s' ds \\ &= \lambda \int_0^t \alpha_s' ds \\ &= (\lambda\alpha)_t.\end{aligned}$$

□

We use this fact as motivation to define scaling on Lie groups.

Definition 3.2.4. Suppose G is a Lie group and $\mathfrak{g}$ its Lie algebra. Let $\gamma \in C^{pr}(G)$ and $\lambda \geq 0$. We define the *Lie algebra scaling* of γ by λ to be

$$\lambda \cdot \gamma := \Psi_{\gamma_0} \circ c_\lambda \circ \Psi_{\gamma_0}^{-1}(\gamma). \tag{3.2.3}$$

Remark 3.2.2. We highlight three important differences between vector space scaling for paths in $\mathbb{R}^N$ and Lie algebra scaling for paths in an arbitrary Lie group G, and provide a reason for each.

1. Returning to the setting of paths in $\mathbb{R}^N$, the two notions of scaling differ slightly when the path does not start at the origin. If we have $\alpha \in C^{pr}(\mathbb{R}^N)$ such that $\alpha_0 = x$,

then $(\lambda\alpha)_0 = \lambda x$, while $(\lambda \cdot \alpha)_0 = x$. However, if we align the initial points, the paths coincide,

$$(\lambda\alpha) - \lambda x = (\lambda \cdot \alpha) - x.$$

This difference is due to the fact that arbitrary Lie groups do not have a natural scaling operation. However, if our Lie group was equipped with a suitable scaling operation, such as a Carnot group [LD17], then we would be able to do define a scaling operation that coincides with the vector space scaling in $C^{pr}(\mathbb{R}^N)$.

2. We have only defined scaling by a nonnegative number. Definition 3.2.4 could be extended to all real numbers λ without any changes, but the interpretation of negative scaling is more difficult in arbitrary Lie groups. For a path $\alpha \in C_0^{pr}(\mathbb{R}^N)$, scaling by $\lambda = -1$ simply produces the pointwise inverse of a path. However, this is not the case in a general Lie group. For example, let $X, Y \in \mathfrak{g}$ and consider the piecewise path

$$\gamma_t = \begin{cases} e^{2tX} & : t \in [0, \frac{1}{2}) \\ e^X e^{(2t-1)Y} & : t \in [\frac{1}{2}, 1]. \end{cases}$$

Here, we have $\gamma_1 = e^X e^Y$ and $(-1 \cdot \gamma)_1 = e^{-X} e^{-Y}$, which are not inverses in general. Thus, we see that the obstruction to this interpretation is the noncommutativity of arbitrary Lie groups. However, in the setting of abelian Lie groups, such an interpretation would hold.

3. By definition, the vector space scaling in $C^{pr}(\mathbb{R}^N)$ obeys the distributive law: $\lambda(\alpha + \beta) = (\lambda\alpha) + (\lambda\beta)$ for $\alpha, \beta \in C^{pr}(\mathbb{R}^N)$ and $\lambda \in \mathbb{R}$. In other words, the vector space scaling is a pointwise Lie group homomorphism for $\mathbb{R}^N$. However, $c_\lambda : \mathfrak{g} \to \mathfrak{g}$ is not a morphism of Lie algebras in general since $c_\lambda([X, Y]) = \lambda[X, Y] \neq \lambda^2[X, Y] = [c_\lambda X, c_\lambda Y]$. Thus, it cannot be the induced map of an underlying Lie group homomorphism for G, so the Lie algebra scaling for G is not distributive, $\lambda \cdot (\alpha\beta) \neq (\lambda \cdot \alpha)(\lambda \cdot \beta)$, in general. In the case of an abelian Lie group H, the associated Lie algebra $\mathfrak{h}$ is abelian so that $[X, Y] = 0$ for all $X, Y \in \mathfrak{h}$, and thus Lie algebra scaling can be viewed as a pointwise Lie group morphism.

Due to these remarks, we must keep in mind that the scaling operation for paths in Lie groups is not compatible with the algebraic structure of G.

3.2.3 *Discrete Time Series on G*

In this subsection, we will consider the interpretation of discrete time series on an arbitrary Lie group G, and also discuss derivative computations for these discrete time series. We will continue the theme of comparison with the corresponding notions in $\mathbb{R}^N$.

Remark 3.2.3. Here, we will assume that discrete time series are uniformly sampled at integer times. This does not result in any loss of generality due to the reparametrization invariance of the path signature, given in Proposition 3.3.2.

Let $T \in \mathbb{N}$ and $\hat{x} : [T+1] \to \mathbb{R}^N$ be a discrete time series in $\mathbb{R}^N$ of length $T+1$. There is a natural interpretation of $\hat{x}$ as a continuous time series $x : [T+1] \to \mathbb{R}^N$ by linear interpolation between points. Namely, it is the interpolation with a constant derivative between the discrete points defined in $\hat{x}$. This is the interpretation that we implicitly take when we compute derivatives of discrete time series by finite differences $\hat{x}'_i = \hat{x}_{i+1} - \hat{x}_i$ to get the discrete derivative $\hat{x}' : [T] \to \mathbb{R}^N$. Additionally, we can think about the continuous path x as a geodesic interpolation of the discrete path $\hat{x}$.

However, the interpretation is more subtle in the case of arbitrary Lie groups. Suppose we have a discrete time series in G, which we denote by $\hat{\gamma} : [T+1] \to G$. We wish to obtain an interpolation such that the derivative, when viewed in the Lie algebra $\mathfrak{g}$, is constant between adjacent points. This can be achieved by taking the logarithm of the difference between adjacent points. We define the *discrete derivative* $\hat{\gamma}' : [T] \to \mathfrak{g}$ of a discrete Lie group valued path by

$$\hat{\gamma}'_i := \log\left(\hat{\gamma}_i^{-1}\hat{\gamma}_{i+1}\right) \in \mathfrak{g}. \tag{3.2.4}$$

Then, we can define the continuous interpolation $\gamma : [0, T+1] \to G$ using the exponential map such that for $t \in [i, i+1)$, the interpolation is

$$\gamma_t := \hat{\gamma}_t \exp\left((t-i)\hat{\gamma}'_t\right).$$

We note that this construction reduces to linear interpolation in the case of $G = \mathbb{R}^n$. This is due to the fact that for the additive Lie group $\mathbb{R}^N$, the exponential and logarithm map are both the identity and are both globally defined. Additionally, the group operation is addition, so we should interpret all of the products as sums. However, there are two essential differences between the case of arbitrary Lie groups and Euclidean space.

Firstly, for an arbitrary Lie group G, the logarithm map is only defined in a neighborhood of the identity. The two reasons the logarithm may not be defined in a larger neighborhood are the loss of injectivity and the loss of surjectivity of the exponential map. On any compact Lie group, the exponential map will not be injective at any point. In this case,

we can define the logarithm to be the value closest to the origin, but non-injectivity may still occur. For example, the point antipodal to the identity in S^1 has no unique logarithm since there are two paths of equal distance to the identity. However, if we perturb the target point in either direction, there exists a unique shortest path. This implies that by undersampling the underlying time series, we may infer incorrect information. The case of S^1 is exactly the situation encountered in the Nyquist sampling theorem.

The exponential map is not always surjective, with the simplest examples being non-connected Lie groups. However, connected Lie groups such as $SL(2, \mathbb{R})$ can still have non-surjective exponential maps. In these cases, discrete derivatives may not exist, and finer sampling is required so that the difference between adjacent points $\tilde{\gamma}_i^{-1}\tilde{\gamma}_{i+1}$ is closer to the identity and has a well-defined logarithm. However, for compact Lie groups such as $SO(3)$, the Lie exponential map is surjective.

Secondly, the interpolation defined here may not be a geodesic connecting the two points. Suppose h is a Riemannian metric on G. In general, geodesics do not coincide with the one-parameter subgroups of G. In other words, in these cases, the Riemannian exponential map and the Lie exponential map are *not* the same. However, for bi-invariant metrics, they coincide.

Theorem 3.2.2. *The Lie exponential map and the Riemannian exponential map at the identity agree on Lie groups with bi-invariant metrics.*

Thus, for all Lie groups equipped with bi-invariant metrics, we may continue to interpret the interpolation as a geodesic interpolation. In fact, this holds for all compact Lie groups.

Proposition 3.2.3. *Every compact Lie group admits a bi-invariant metric.*

From this discussion, we find that for a compact Lie group G, the interpretation of discrete time series on G is similar to the case of $\mathbb{R}^N$, with the main difference being the non-injectivity of the exponential map.

3.3 PATH SIGNATURES ON LIE GROUPS

This subsection, based on the exposition of path signatures on Euclidean space given in [GL20], begins by defining the path signature for Lie groups. We show several basic properties which are well-known for path signatures on Euclidean space, culminating in the definition of tree-like equivalence for paths and the property that the signature is an injective group homomorphism. This material was originally developed by [Che54; Che57; Che58] and is not novel.

We then prove a signature preserving bijection between paths on an N-dimensional Lie group G and paths on $\mathbb{R}^N$, which highlights the extent to which the theory naturally extends to the case of Lie groups. This result provides a Euclidean representation of Lie

group valued time series, and can thus be used to apply classical Euclidean data analysis techniques to Lie group valued time series.

We then consider the extension of the equivariance property of path signatures. This is followed by an interpretation of the second-level signature terms as indicators of lead-lag behavior between the directions corresponding to our choice of basis vectors for the Lie algebra $\mathfrak{g}$. Finally, we close this section by discussing computational aspects of the path signature for discrete time series, as well as symmetry breaking path transformations which can be used as a preprocessing step.

In this section, we use $(e_1, \ldots, e_N)$ to denote an ordered basis of $\mathfrak{g}$ and use $(\omega_1, \ldots, \omega_N)$ to denote the dual basis of $\mathfrak{g}^*$ such that $\omega_i(e_j) = \delta_{i,j}$, where $\delta_{i,j}$ is the Kronecker delta.

3.3.1 *Path Signature as a Group Homomorphism*

Let G be an N-dimensional Lie group. Recall that $C^{pr}(G)$ denotes the space of piecewise regular paths $\gamma : [0,1] \to G$.

Definition 3.3.1. Let $\gamma \in C^{pr}(G)$. Suppose $\omega_1, \ldots, \omega_N \in \mathfrak{g}^*$ form a basis of $\mathfrak{g}^*$. For $i \in [N]$, define a path $S^i(\gamma)_t : [0,1] \to \mathbb{R}$ as

$$S^i(\gamma)_t \coloneqq \int_0^t \omega_i(\gamma'_s) ds.$$

Next, let $I = (i_1, \ldots, i_m)$ be a multi-index, where $i_j \in [N]$. Higher order paths $S^I(\gamma)_t : [0,1] \to \mathbb{R}$ are inductively defined as

$$S^I(\gamma)_t \coloneqq \int_0^t S^{(i_1,\ldots,i_{m-1})}(\gamma)_s \omega_{i_m}(\gamma'_s) ds. \tag{3.3.1}$$

The *path signature of γ with respect to I* is defined to be $S^I(\gamma) \coloneqq S^I(\gamma)_1$.

We can also present the definition in a non-inductive way. Let Δ^m be the standard m-simplex

$$\Delta^m = \{(t_1, \ldots, t_m) \,:\, 0 \leq t_1 < t_2 < \ldots < t_m \leq 1\}.$$

By collapsing the inductive definition, we can write the path signature of γ with respect to $I = (i_1, \ldots, i_m)$ as

$$S^I(\gamma) = \int_{\Delta^m} \omega_{i_1}(\gamma'_{t_1}) \ldots \omega_{i_m}(\gamma'_{t_m})\, dt_1 \ldots dt_m. \tag{3.3.2}$$

We can amalgamate the path signatures with respect to every multi-index I into an element of a tensor algebra. Let $\bar{\mathfrak{g}}$ be the underlying vector space of the Lie algebra $\mathfrak{g}$. The

tensor algebra $T((\bar{\mathfrak{g}}))$ is defined in Section 1.2.2. Let $e_1, \dots, e_N$ be a basis for $\mathfrak{g}$. We define the *path signature* of $\Gamma \in C^{pr}(G)$ to be

$$S(\gamma) \coloneqq 1 + \sum_{m \geq 1} \sum_{|I|=m} S^I(\gamma) e_{i_1} \otimes \dots \otimes e_{i_m} \in T((\bar{\mathfrak{g}})). \tag{3.3.3}$$

Remark 3.3.1. For path signatures defined on Euclidean space $\mathbb{R}^N$, we often choose the standard 1-forms $(dx_1, \dots, dx_N)$ to be the basis of $\mathfrak{r}$, the Lie algebra of $\mathbb{R}^N$. Suppose $\alpha \in C^{pr}(\mathbb{R}^N)$. We can also write our path component-wise as $\alpha = (\alpha^1, \dots, \alpha^N)$, where each $\alpha^i : [0,1] \to \mathbb{R}$. Then, evaluation of the standard 1-forms is simply $dx_i(\alpha'_t) = (\alpha^i)'_t$. Thus, in the Euclidean case, the definition of the path signature reduces to

$$S^I(\alpha) = \int_{\Delta^m} (\alpha^{i_1})'_{t_1} \dots (\alpha^{i_m})'_{t_m} \, dt_1 \dots dt_m. \tag{3.3.4}$$

Let $\gamma \in C^{pr}(G)$ and $g \in G$. The *left translation* of γ by g is defined to be the path $(g\gamma)_t \coloneqq g(\gamma_t)$, where we left translate the path γ by g pointwise (one can analogously define the *right translation* of a path). Similar to the case of Euclidean space, path signatures are left translation invariant and reparametrization invariant.

Proposition 3.3.1 (Left translation invariance)**.** *Let $\gamma \in C^{pr}(G)$ and $g \in G$. Then $S(g\gamma) = S(\gamma)$.*

Proof. It suffices to show that $S^I(g\gamma) = S^I(\gamma)$ for all multi-indices I. Note that we have

$$(g\gamma)'_t = L_{g*}\gamma'_t.$$

Specifically, this implies that γ'_t and $g\gamma'_t$ are represented by the same element in the Lie algebra $\mathfrak{g}$. Therefore for any $\omega \in \mathfrak{g}^*$, we have $\omega(g\gamma'_t) = \omega(\gamma'_t)$. Thus, $S^I(g\gamma) = S^I(\gamma)$ for all I. □

Proposition 3.3.2 (Reparametrization invariance)**.** *Let $\gamma : [a,b] \to G$ be a piecewise regular path, and let $\phi : [c,d] \to [a,b]$ be a strictly increasing function. Then $S(\gamma \circ \phi) = S(\gamma)$.*

Proof. This is the Change of Variables Theorem. Reparametrization invariance of the first level of the signature is given as

$$S^i(\gamma \circ \phi) = \int_c^d \omega_i((\gamma \circ \phi)'_t) dt = \int_c^d \omega_i(\gamma'_{\phi_t}) \phi'_t dt = \int_a^b \omega_i(\gamma'_\tau) d\tau = S^i(\gamma).$$

Invariance for higher order terms is shown by induction using the same argument. □

In particular this proposition justifies our choice of only considering paths parametrized by $[0,1]$, as any other path can be reparametrized into this domain. Next, we would like to

understand how scaling of paths in G given in Definition 3.2.4 affects the path signature. Note that the vector space scaling in $\bar{\mathfrak{g}}$ induces a dilation map $\sigma_\lambda : T((\bar{\mathfrak{g}})) \to T((\bar{\mathfrak{g}}))$, defined in Equation (1.2.17).

Proposition 3.3.3. *Let $\gamma \in C^{pr}(G)$ and $\lambda \geq 0$. Then $S(\lambda \cdot \gamma) = \delta_\lambda S(\gamma)$.*

Proof. Consider the multi-index $I = (i_1, \ldots, i_k)$. Then,

$$\begin{aligned} S^I(\lambda \cdot \gamma) &= \int_{\Delta^k} \omega_{i_1}(\lambda \gamma'_{t_1}) \ldots \omega_{i_k}(\lambda \gamma'_{t_k})\, dt_1 \ldots dt_k \\ &= \lambda^k \int_{\Delta^k} \omega_{i_1}(\gamma'_{t_1}) \ldots \omega_{i_k}(\gamma'_{t_k})\, dt_1 \ldots dt_k \\ &= \lambda^k S^I(\gamma). \end{aligned}$$

□

We have seen that the group structure on G allows us to define a group structure on $C^{pr}(G)$ by pointwise multiplication. The group structure on G allows us to obtain another group structure on a quotient of $C^{pr}(G)$ where the group operation is given by concatenation. Let $\alpha, \beta \in C^{pr}(G)$. The *concatenation* of α and β is defined to be

$$(\alpha * \beta)_t = \begin{cases} \alpha_{2t} & : t \in [0, \frac{1}{2}) \\ \alpha_1 (\beta_0)^{-1} \beta_{2t-1} & : t \in [\frac{1}{2}, 1]. \end{cases}$$

The inverse of a path γ is defined to be the same path, but in the reverse direction

$$(\gamma^{-1})_t = \gamma_{1-t}.$$

Concatenation or inversion of piecewise regular paths is still piecewise regular. In order to obtain an identity element, we must quotient out by an equivalence relation.

Definition 3.3.2. A path $\gamma \in C^{pr}(G)$ is called *reducible* if there exist paths $\alpha, \beta, \zeta \in C^{pr}(G)$ such that $\gamma = \alpha * \zeta * \zeta^{-1} * \beta$, up to reparametrization. The path $\alpha * \beta$ is called a *reduction* of γ. We define the reduction of $\zeta * \zeta^{-1}$ to be c_e, the constant path at the identity $e \in G$. A path γ is *irreducible* if no reduction exists. An irreducible path $\tilde{\gamma}$ obtained by finitely many iterative reductions of a path γ is called an *irreducible reduction* of γ.

Theorem 3.3.1 ([Che58])**.** *Every piecewise regular path $\gamma \in C^{pr}(G)$ has a unique irreducible reduction up to reparametrization.*

This result allows us to define the notion of tree-like equivalence.

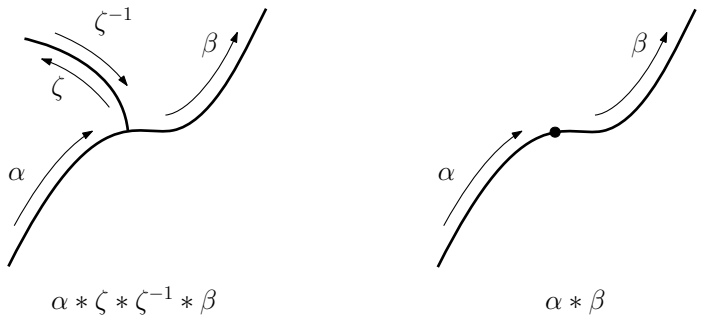


Figure 3.1: (Left) An example of a reducible path $\alpha * \zeta * \zeta^{-1} * \beta$. (Right) The irreducible reduction of the path on the left.

Definition 3.3.3. A path $\gamma \in C^{pr}(G)$ is a *tree-like path* if its irreducible reduction is c_e, the constant path at the identity. Two paths α, β are *tree-like equivalent*, $\alpha \sim_t \beta$, if $\alpha * \beta^{-1}$ is a tree-like path.

Remark 3.3.2. The definition of tree-like equivalence includes translations. Indeed, suppose $\gamma \in C^{pr}(G)$ and $g \in G$. Define $g\gamma$ and γg to be the left and right translations of the path γ by g. Then $\gamma \sim_t g\gamma$ since $\gamma * (g\gamma)^{-1} = \gamma * \gamma^{-1}$ by the definition of the concatenation operator. The same holds for right translations.

Additionally, tree-like equivalence also includes reparametrization since the definition of reductions are reparametrization invariant.

Proposition 3.3.4. *Tree-like equivalence is an equivalence relation.*

Proof. Let $\gamma, \gamma_1, \gamma_2, \gamma_3 \in C^{pr}(G)$. Note that the subscript here denotes distinct paths, and does not denote the time parameter. By definition the reduction of $\gamma * \gamma^{-1}$ is the constant path, so $\gamma \sim_t \gamma$.

Next, if $\gamma = \alpha * \zeta * \zeta^{-1} * \beta$, for paths $\alpha, \beta, \zeta \in C^{pr}(G)$, then $\gamma^{-1} = \beta^{-1} * \zeta * \zeta^{-1} * \alpha^{-1}$. Thus, a path is reducible if and only if its inverse is reducible. Additionally, the reduction $\beta^{-1} * \alpha^{-1}$ of γ^{-1} is the inverse of the reduction $\alpha * \beta$ of γ. Now, suppose $\gamma_1 \sim_t \gamma_2$ so that $\gamma_1 * \gamma_2^{-1}$ is tree-like. By the above argument, $\gamma_2 * \gamma_1^{-1}$ is also tree-like, so $\gamma_2 \sim_t \gamma_1$.

Finally, the concatentation $\alpha * \beta$ of two tree-like paths is also tree-like, by performing all the reductions of α and then performing all the reductions on β. Suppose $\gamma_1 \sim_t \gamma_2$ and $\gamma_2 \sim_t \gamma_3$. Then, $\gamma_1 * \gamma_3^{-1}$ is a reduction of $(\gamma_1 * \gamma_2^{-1}) * (\gamma_2 * \gamma_3^{-1})$, and the latter path is tree-like since it is a concatenation of two tree-like paths. By the uniqueness of irreducible reductions, $\gamma_1 * \gamma_3^{-1}$ is tree-like. Thus, $\gamma_1 \sim_t \gamma_3$. □

We can now define $\widetilde{C}^{pr}(G) := C^{pr}(G)/\sim_t$ to be the space of tree-like equivalence classes of piecewise regular paths in G. We define the identity element to be $[c_e] \in \widetilde{C}^{pr}(G)$, the equivalence class of the constant path at the identity. Compatibility of concatenation and

inversion are implicit in the above proof, and the group axioms are easily checked. Thus, we have shown the following.

Proposition 3.3.5. *The quotient $\widetilde{C}^{pr}(G)$ is a group.*

We can now state Chen's injectivity theorem.

Theorem 3.3.2 ([Che58])**.** *Suppose G is a real Lie group. Let $\alpha, \beta \in C^{pr}(G)$. Then $S(\alpha) = S(\beta)$ if and only if α and β are tree-like equivalent.*

Chen also showed that the signature is a group homomorphism. Namely, suppose $\alpha, \beta \in C^{pr}(G)$. Chen's identity [Che54] states that

$$S(\alpha * \beta) = S(\alpha) \otimes S(\beta). \tag{3.3.5}$$

Putting the previous results together, we obtain the following characterization.

Proposition 3.3.6. *The path signature map $S : \widetilde{C}^{pr}(G) \to T((\overline{\mathfrak{g}}))$ is an injective group homomorphism.*

We will also require an internal multiplicative structure on the path signature coefficients which is an immediate generalization of the Euclidean path signature. The shuffle product $\sqcup\!\sqcup$ of two multi-indices is defined in Definition 1.2.7.

Theorem 3.3.3. *Let I and J be multi-indices in $[N]$, of lengths k and l respectively, and suppose $\gamma \in C^{pr}(G)$. Then*

$$S^I(\gamma)S^J(\gamma) = \sum_{K \in I \sqcup\!\sqcup J} S^K(\gamma). \tag{3.3.6}$$

Proof. Let $R = (r_1, \dots, r_k, r_{k+1}, \dots r_{k+l}) = (i_1, \dots, i_k, j_1, \dots, j_l)$. Writing out the signature on the left side of the equation using Equation 3.3.2, we get

$$\begin{aligned}\int_{\Delta^k} \omega_{i_1}(\gamma'_{t_1}) \dots \omega_{i_k}(\gamma'_{t_k}) dt_1 \dots dt_k \int_{\Delta^l} \omega_{j_1}(\gamma'_{t_1}) \dots \omega_{j_l}(\gamma'_{t_l}) dt_1 \dots dt_l \\ = \int_{\Delta^k \times \Delta^l} \omega_{r_1}(\gamma'_{t_1}) \dots \omega_{r_{k+l}}(\gamma'_{t_{k+l}})\, dt_1 \dots dt_{k+l},\end{aligned}$$

and the sum on the right side is

$$\sum_{\sigma \in Sh(k,l)} \int_{\Delta^{k+l}} \omega_{\sigma(r_1)}(\gamma'_{t_1}) \dots \omega_{\sigma(r_{k+l})}(\gamma'_{t_{k+l}}) dt_1 \dots dt_{k+l}.$$

The equivalence of the two formulas is given by the standard decomposition of $\Delta^k \times \Delta^l$ into $(k+l)$-simplices,

$$\begin{aligned}\Delta^k \times \Delta^l &= \{(t_1, \ldots, t_{k+l}) : 0 < t_1 < \ldots < t_k < 1, 0 < t_{k+1} < \ldots < t_{k+l} < 1\} \\ &= \bigsqcup_{\sigma \in Sh(k,l)} \left\{(t_{\sigma(1)}, \ldots, t_{\sigma(k+l)}) : 0 < t_1 < \ldots < t_{k+l} < 1\right\}.\end{aligned}$$

□

3.3.2 *Relationship between Euclidean and Lie Group Valued Paths*

In this subsection, we define a signature-preserving bijection between piecewise regular paths in G and piecewise regular paths in $\mathbb{R}^N$ which start at the identity and origin respectively.

The idea behind the following proposition is that the path signature computation only requires the first derivative of paths. The Lie bracket is unused in the computation of path signatures, so we can simply consider the Lie algebras of Lie group as vector spaces. Thus, we can identify the underlying vector space of the Lie algebra $\mathfrak{g}$ with the underlying vector space of the Lie algebra $\mathfrak{r}$ of $\mathbb{R}^N$. We then use the correspondence $\Psi_g : C^{pc}(\mathfrak{g}) \to C_g^{pr}(G)$ between piecewise continuous paths $C^{pc}(\mathfrak{g})$ and piecewise regular paths $C^{pr}(G)$ given in Corollary 3.2.1, to map paths on G to paths on $\mathbb{R}^N$.

In the following proposition, we abuse notation and consider elements of the Lie algebras $\mathfrak{r}$ of $\mathbb{R}^N$ and $\mathfrak{g}$ of G as both the tangent space at the identity, and the vector space of left-invariant vector fields. Similarly, we consider elements of the dual of the Lie algebra $\mathfrak{r}^*$ and $\mathfrak{g}^*$ as both the cotangent space at the identity, and the vector space of left-invariant 1-forms.

Because we will be using two different path signature functions, we will denote by $S_{\mathbb{R}} : C^{pr}(\mathbb{R}^N) \to T((\mathbb{R}^N))$ the path signature for $\mathbb{R}^N$ with respect to the ordered basis of standard 1-forms $(dx_1, \ldots, dx_N)$ of $\mathfrak{r}^*$. We denote $S_G : C^{pr}(G) \to T((\mathbb{R}^N))$ to be the path signature for G with respect to a given ordered basis $(\omega_1, \ldots, \omega_N)$ of $\mathfrak{g}^*$.

Proposition 3.3.7. *Suppose G is an N-dimensional Lie group with Lie algebra $\mathfrak{g}$. Let $\phi : \bar{\mathfrak{r}} \to \bar{\mathfrak{g}}$ be an isomorphism of vector spaces, and $\phi^* : \mathfrak{g}^* \to \mathfrak{r}^*$ be its dual. Let $(dx_1, \ldots, dx_N)$ denote the standard 1-forms of $\mathbb{R}^N$, and define $\omega_i = (\phi^*)^{-1}(dx_i)$. Let $S_G : C^{pr}(G) \to T((\mathbb{R}^N))$ be the path signature map for G with respect to the ordered basis $(\omega_1, \ldots, \omega_N)$ of $\mathfrak{g}^*$. Then, there exists a bijection $\Phi : C_0^{pr}(\mathbb{R}^N) \to C_e^{pr}(G)$ such that $S_{\mathbb{R}}(\gamma) = S_G(\Phi(\gamma))$ for all $\gamma \in C^{pr}(\mathbb{R}^N)$.*

Proof. The construction of the map Φ is derived from Corollary 3.2.1. Let $\Psi_{\mathbb{R}} : C^{pc}(\mathfrak{r}) \to C_0^{pr}(\mathbb{R}^N)$ and $\Psi_G : C^{pc}(\mathfrak{g}) \to C_e^{pr}(G)$ be the bijections from Corollary 3.2.1 for $\mathbb{R}^N$ and G respectively. Now, define Φ by

$$\Phi : C_0^{pr}(\mathbb{R}^N) \xrightarrow{\Psi_{\mathbb{R}}^{-1}} C^{pc}(\mathfrak{r}) \xrightarrow{\phi} C^{pc}(\mathfrak{g}) \xrightarrow{\Psi_G} C_e^{pr}(G).$$

The idea is that we start with a path $\gamma \in C_0^{pr}(\mathbb{R}^N)$, and apply the following maps:

1. $\Psi_{\mathbb{R}}^{-1}$: take the derivative γ' to obtain a path in $\mathfrak{r}$
2. ϕ : identify the underlying vector space of $\mathfrak{r}$ with $\mathfrak{g}$
3. Ψ_G : solve the differential equation (Equation 3.2.2) with the identity initial condition to obtain a path in G.

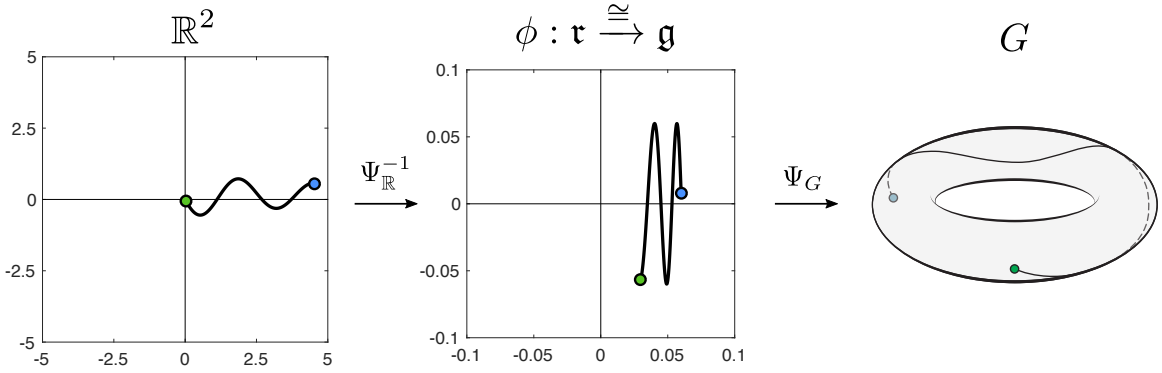


Figure 3.2: An example of the function Φ when we take $G = S^1 \times S^1$. (Left) A path $\gamma \in P\mathbb{R}_0^2$. (Middle) The derivative of γ as a path in $\mathfrak{r}$ or $\mathfrak{g}$. (Right) The corresponding path $\Phi(\gamma)$.

Because all three maps are bijective, Φ is also bijective. To show that the signatures are invariant under this mapping, let $I = (i_1, \ldots, i_k)$ and $\gamma \in C_0^{pr}(\mathbb{R}^N)$. The path signature of γ with respect to I is

$$S_{\mathbb{R}}^I(\gamma) = \int_{\Delta^k} dx_{i_1}(\gamma_{t_1}) \ldots dx_{i_k}(\gamma_{t_k}) dt_1 \ldots dt_k.$$

Note that the derivative of $\Phi(\gamma)$ is given by $\Phi(\gamma)'_t = \phi(\gamma'_t)$ and thus, the path signature of $\Phi(\gamma)$ with respect to I is

$$\begin{aligned} S^I_G(\Phi(\gamma)) &= \int_{\Delta^k} \omega_{i_1}(\phi(\gamma'_{t_1})) \dots \omega_{i_k}(\phi(\gamma'_{t_k}))dt_1 \dots dt_k \\ &= \int_{\Delta^k} \phi^*(\omega_{i_1})(\gamma'_{t_1}) \dots \phi^*(\omega_{i_k})(\gamma'_{t_k})dt_1 \dots dt_k \\ &= S^I_{\mathbb{R}}(\gamma). \end{aligned}$$

The final equality holds because the dual isomorphism ϕ^* takes ω_i to dx_i. Thus, $S^I_{\mathbb{R}}(\gamma) = S^I_G(\Phi(\gamma))$ for all $\Gamma \in C^{pr}(\mathbb{R}^N)_e$ and all multi-indices I. □

3.3.3 *Stability of the Path Signature*

In this section, we will discuss the stability of path signatures, which is of crucial importance in machine learning applications. We begin by considering a metric structure on $C^{pr}(G)$. We mentioned in Section 3.2.2 that a metric on $C^{pr}_0(\mathbb{R}^N)$ can be obtained by using the 1-variation of paths. Namely, given $\alpha, \beta \in C^{pr}_0(\mathbb{R}^N)$, we may consider the distance between these two paths as $|\alpha - \beta|_{1-var}$, where the difference is performed pointwise. Such an approach could be used for $C^{pr}(G)$ in theory. Now suppose $\alpha, \beta \in C^{pr}_e(G)$, we can use the metric on G to define the distance to be $|\beta^{-1}\alpha|_{1-var}$, where the inversion and multiplication are both performed pointwise. However, this notion of distance is not well-suited for the path signature.

The main reason for this is that the computation of $|\beta^{-1}\alpha|_{1-var}$ depends fundamentally on the adjoint action of the Lie group on the Lie algebra, which is governed by the Lie bracket. Namely, the adjoint action is trivial if and only if the Lie bracket is zero. However, the path signature ignores the Lie bracket structure, so the prospect of Lipschitz continuity of the signature with respect to this metric is problematic.

We therefore consider a different metric. Note that our path signature computations have consistently been performed on the underlying vector space of the Lie algebra $\bar{\mathfrak{g}}$, so it seems natural to directly define a metric using the derivatives $\alpha', \beta' \in \mathrm{Seq}(\mathfrak{g})$. One such notion of a distance would be the L^1 distance between these derivatives

$$\|\alpha' - \beta'\|_{L^1} = \int_0^1 \|\alpha'_t - \beta'_t\|_{\mathfrak{g}}\, dt$$

which in particular does not use the Lie bracket structure. In fact this L^1 distance is exactly the 1-variation of the corresponding paths $\Phi^{-1}(\alpha), \Phi^{-1}(\beta) \in C_0^{pr}(\mathbb{R}^N)$, given by the bijection in Proposition 3.3.7. Thus, we can define the metric on $C_e^{pr}(G)$ to be

$$d_{\mathbb{R}}(\alpha, \beta) := |\Phi^{-1}(\alpha) - \Phi^{-1}(\beta)|_{1-var}.$$

Note that equipped with this metric, the map Φ is trivially an isometry.

Lemma 3.3.1. *Suppose $\Phi : C_0^{pr}(\mathbb{R}^N) \to C_e^{pr}(G)$ is the map defined in Proposition 3.3.7. Suppose $C_0^{pr}(\mathbb{R}^N)$ is equipped with the 1-variation metric, and $C_e^{pr}(G)$ is equipped with the metric $d_{\mathbb{R}}$. Then, Φ is an isometry.*

Using this isometry, stability for Lie group path signatures is a direct corollary of stability for Euclidean path signatures from Corollary 3.3.1.

Corollary 3.3.1. *Let $\ell > 0$ be a constant such that $\ell \neq 1$ and $\alpha, \beta \in C_e^{pr}(G)$ be bounded variation paths such that $|\alpha|_{1-var}, |\beta|_{1-var} \leq \ell$. For all $m \geq 1$,*

$$\|S_{\leq m}(\alpha) - S_{\leq m}(\beta)\| \leq \exp(1) \left(\frac{1 - \ell^m}{1 - \ell} \right)^{1/2} d_{\mathbb{R}}(\alpha, \beta). \tag{3.3.7}$$

3.3.4 *Equivariance of the Path Signature*

At this point, a natural question to consider is how do path signatures behave under Lie group morphisms? Let G_1 and G_2 be Lie groups of dimensions N_1 and N_2 respectively. Given a Lie group morphism $F : G_1 \to G_2$, we have an induced Lie algebra morphism $F_* : \mathfrak{g}_1 \to \mathfrak{g}_2$ between the corresponding Lie algebras. In particular, all Lie algebra morphisms are linear transformations, so if we forget the Lie bracket, this results in a map $F_* : \bar{\mathfrak{g}}_1 \to \bar{\mathfrak{g}}_2$ between the underlying vector spaces. Because linear transformations induce maps on tensor products of the space $F_*^{\otimes m} : \bar{\mathfrak{g}}_1^{\otimes m} \to \bar{\mathfrak{g}}_2^{\otimes m}$, we also get an induced map of algebras between tensor algebras

$$F_* : T((\bar{\mathfrak{g}}_1)) \to T((\bar{\mathfrak{g}}_2)).$$

If $(e_1, \ldots, e_{N_1})$ is an ordered basis for $\mathfrak{g}_1$ and $(f_1, \ldots, f_{N_2})$ is an ordered basis for $\mathfrak{g}_2$, then we can write $F_* : \mathfrak{g}_1 \to \mathfrak{g}_2$ as an $N_2 \times N_1$ matrix in terms of these bases, which we call M. We can describe the action of F_* in the tensor algebra using this matrix. Let $\mathbf{t} \in T((\bar{\mathfrak{g}}_1))$. In general, the action on the order m elements $\mathbf{t}_m \in \bar{\mathfrak{g}}^{\otimes m}$ is a tensor-matrix multiplication, as

described in [PSS19], in which all m sides of the tensor $\mathbf{t}_m$ are multiplied by the matrix M. This can be written out as

$$F_*\mathbf{t} = \sum_{m=0}^{\infty} \sum_{|I|=m} \mathbf{t}^I (Me_{i_1}) \otimes (Me_{i_2}) \otimes \ldots \otimes (Me_{i_k}).$$

The low order tensors can be written out in usual matrix notation. Consider $\mathbf{t}_1$ as a column vector. The action on first order elements is matrix multiplication,

$$(F_*\mathbf{t})_1 = M\mathbf{t}_1.$$

Considering $\mathbf{t}_2$ as a matrix, the action on the second order elements is conjugation by M,

$$(F_*\mathbf{t})_2 = M\mathbf{t}_2 M^\intercal.$$

For higher orders, we can no longer use matrix notation, so we explicitly define the action for a given index. Let $J = (j_1, \ldots, j_n)$ be a multi-index where $j_k \in [N_2]$. Then, the element of $F_*\mathbf{t}$ corresponding to the multi-index J is

$$(F_*\mathbf{t})^J = \sum_{i_1=1}^{N_1} \sum_{i_2=1}^{N_1} \ldots \sum_{i_n=1}^{N_1} \mathbf{t}^{(i_1,\ldots,i_n)} M_{j_1,i_1} M_{j_2,i_2}, \ldots, M_{j_n,i_n}.$$

The following is a generalization of the equivariance of the path signature in Euclidean space, which is discussed in [FV10] and [PSS19]. Here, suppose $(\omega_1, \ldots, \omega_{N_1})$ is the dual basis to $(e_1, \ldots, e_{N_1})$ and $(\nu_1, \ldots, \nu_{N_2})$ is the dual basis to $(f_1, \ldots, f_{N_2})$.

Proposition 3.3.8. *Let G_1 and G_2 be Lie groups, with Lie algebras $\mathfrak{g}_1$ and $\mathfrak{g}_2$ respectively. Suppose $F : G_1 \to G_2$ is a Lie group morphism and $\gamma \in C^{pr}(G_1)$. Then*

$$S(F\gamma) = F_* S(\gamma).$$

Proof. The proof of this claim is simply due to the linearity of integrals and 1-forms. Consider the multi-index $J = (j_1, \ldots, j_m)$. Then,

$$S^J(F\gamma) = \int_{\Delta^m} \nu_{j_1}(F_*\gamma'_{t_1}) \ldots \nu_{j_m}(F_*\gamma'_{t_m}) dt_1 \ldots dt_m.$$

Consider a single factor in the integrand. Using the basis $(e_1, \ldots, e_{N_1})$ for $\mathfrak{g}$, write the derivative γ' as

$$\gamma'_t = \sum_{i=1}^{N_1} c_t^i e_i$$

where $c^i : [0,1] \to \mathbb{R}$ are the component paths. Then, since $\nu_j(F_*\gamma'_t)$ denotes the j^{th} component of $F_*\gamma'_t$, we can write this as

$$\nu_j(F_*\gamma'_t) = \sum_{i=1}^{N_1} M_{j,i}\omega_i(\gamma'_t).$$

Substituting this back into the formula for $S^J(F\gamma)$, we get

$$\begin{aligned} S^J(F\gamma) &= \sum_{i_1=1}^{N_1} \dots \sum_{i_n=1}^{N_1} \left(M_{j_1,i_1} \dots M_{j_m,i_m}\right) \int_{\Delta^m} \omega_{i_1}(\gamma'_t) \dots \omega_{i_m}(\gamma'_t) dt_1 \dots dt_m \\ &= \sum_{i_1=1}^{N_1} \dots \sum_{i_n=1}^{N_1} \left(M_{j_1,i_1} \dots M_{j_m,i_m}\right) S^{(i_1,\dots,i_m)}(\gamma) \\ &= (F_*S(\gamma))^J. \end{aligned}$$

□

3.3.5 *Lead-Lag Relationships*

For path signatures defined on Euclidean space, a certain linear combination of second degree signature terms provides a reparametrization invariant indicator of lead-lag behavior in cyclic real-valued time series, as initially introduced in [BS16] and discussed in Section 1.1. In this subsection, we will extend this interpretation to time series valued in Lie groups.

In the setting of Lie groups, we can define the lead matrix of a path $\gamma \in C^{pr}(G)$ in the same manner as the Euclidean case in Definition 1.1.1, but the interpretation must be slightly modified. Writing out the integral for $S^{i,j}(\gamma)$ given a basis $(\omega_1, \dots, \omega_N)$ of $\mathfrak{g}^*$, we get

$$S^{i,j}(\gamma) = \int_0^1 \left(\int_0^t \omega_i(\gamma'_s) ds \right) \omega_j(\gamma'_t) dt.$$

The inner integral is simply $S^i(\gamma)_t$ and represents the cumulative variation of the path in the direction of e_i (the dual of ω_i), which is the analogue of the displacement in Euclidean space. Thus, for a cyclic time series $\gamma \in C^{pr}(G)$, we say that the e^i direction exhibits cyclic leading behavior with respect to the e^j direction if the following holds:

1. **(positive influence)** when $S^i(\gamma)_t$ is positive (negative), then $\omega_i(\gamma'_t)$ is positive (negative), and

2. **(negative influence)** when $S^j(\gamma)_t$ is positive (negative), then $\omega_j(\gamma'_t)$ is negative (positive).

Thus the lead matrix, as defined in Equation 1.1.4, can be interpreted as a measure of this cyclic leading behavior for Lie group time series. An example of this interpretation is given in Section 3.5.1.

However, the geometric interpretation in terms of signed area is no longer available. This is because any area computation on Lie groups will require second-order differential information about the paths, but path signatures are only defined using first order differential information. This suggests that an interpretation in terms of areas on the Lie group will not be possible. However, by using Proposition 3.3.7, the value $A^{i,j}(\gamma)$ can still be interpreted as the signed area of the corresponding path $\Phi^{-1}(\gamma)$, where Φ is the bijection given in Proposition 3.3.7.

3.3.6 *Topological Considerations*

In this section, we will consider the topological interpretation of the first level signature terms for some Lie groups. Namely, the first level signature term S^i is homotopy invariant if the differential 1-form corresponding to the basis vector $\omega_i \in \mathfrak{g}^*$ is a closed form. For simplicity, we will assume that all paths are piecewise smooth in this section. Note in particular that the continuous interpretation of discrete time series is piecewise smooth, so the discussion in this section holds for analysis of these discrete time series.

Recall the following definition of a homotopy between paths.

Definition 3.3.4. Suppose $\alpha, \beta : [0,1] \to G$ are *homotopic relative to endpoints* if $\alpha_0 = \beta_0$, $\alpha_1 = \beta_1$ and there exists a continuous function $h : [0,1]^2 \to G$, called a *homotopy*, such that

$$h(0,t) = \alpha_t, \quad h(1,t) = \beta_t, \quad h(s,0) = \alpha_0 = \beta_0, \quad h(s,1) = \alpha_1 = \beta_1.$$

We use the notation $\alpha \simeq \beta$ if the paths α and β are homotopic relative to endpoints.

Loosely speaking, two paths are homotopic relative to endpoints if their endpoints coincide, and there exists a continuous deformation from one path to the other. Namely, homotopy relative to endpoints forms an equivalence relation in $C^{pr}(G)$. We say that a map $f : C^{pr}(G) \to \mathbb{R}$ is *homotopy invariant* if $f(\alpha) = f(\beta)$ whenever $\alpha \simeq \beta$.

Recall that a differential form ω is *closed* if its exterior derivative is trivial, $d\omega = 0$. By Stokes' theorem, the first level signature terms for closed forms are homotopy invariant.

Indeed, let $\alpha, \beta \in C^{pr}(G)$, and let $h : [0,1]^2 \to G$ be a homotopy between α and β. By Stokes' theorem, we have

$$\int_{\partial h} \omega = \int_h d\omega = 0,$$

but the boundary of h is exactly $\alpha * \beta^{-1}$. Thus, we have

$$\int_\alpha \omega - \int_\beta \omega = \int_{\partial h} \omega = 0.$$

For left-invariant forms on Lie groups, there is a simple way to determine whether the form is closed. We begin with the invariant formula for the exterior derivative [Lee03]. Let ω be a 1-form on G, and X, Y are vector fields on G, then

$$d\omega(X,Y) = X(\omega(Y)) - Y(\omega(X)) - \omega([X,Y]).$$

If in particular, if $\omega \in \mathfrak{g}^*$ is a left-invariant 1-form and $X, Y \in \mathfrak{g}$ are left-invariant vector fields, then this formula reduces to

$$d\omega(X,Y) = \omega([X,Y])$$

since $\omega(X)$ and $\omega(Y)$ are constant functions. Thus, the left invariant form ω is closed if and only if $\omega([X,Y]) = 0$ for all $X, Y \in \mathfrak{g}$. In particular, this implies that all left-invariant 1-forms are closed on abelian Lie groups such as $\mathbb{R}^N$ and T^N, since $[X,Y] = 0$ for all $X, Y \in \mathfrak{g}$. However, there are no closed left invariant 1-forms on $SO(3)$ since a nontrivial $\omega \in \mathfrak{so}(3)^*$ must be nonzero for at least some $Z \in \mathfrak{so}(3)$. However, for all $Z \in \mathfrak{so}(3)$, there exist $X, Y \in \mathfrak{so}(3)$ such that $Z = [X,Y]$. In fact, this argument extends to all semisimple Lie groups, and thus there are no closed left-invariant 1-forms on any semisimple Lie group.

3.3.7 *Discretization of the Path Signature*

We have focused our discussion of the path signature so far on the continuous setting in order to discuss the theoretical framework. However, applications are studied in the discrete setting. In this section, we provide the explicit computation of path signatures for discrete time series, and discuss a useful discrete approximation.

Let $v = (v^1, \ldots, v^N) \in \mathfrak{g}$, where we have written out the components of v in terms of some choice of basis on $\mathfrak{g}$. Consider the continuous time series $\gamma_t = \exp(vt)$, where exp

is the Lie exponential. Note that the derivative $\gamma'_t = v$ is constant. In this case, the path signature of γ is straightforward to compute. Given $I = (i_1, \ldots, i_m)$, the path signature is

$$\begin{aligned} S^I(\gamma) &= \int_{\Delta^m} \omega_{i_1}(\gamma'_{t_1}) \ldots \omega_{i_m}(\gamma'_{t_m}) dt_1 \ldots dt_m \\ &= \int_{\Delta^m} v_{i_1} \ldots v_{i_m} dt_1 \ldots dt_m \\ &= \frac{v_{i_1} \ldots v_{i_m}}{m!}. \end{aligned}$$

The entire path signature can be written concisely as the tensor exponential.

Definition 3.3.5. Let V be a real vector space. The tensor exponential $\exp_\otimes : V \to T((V))$ is defined to be

$$(\exp_\otimes(v))_m = \frac{v^{\otimes m}}{m!}.$$

Then, we may write the path signature of $\gamma = \exp(vt)$ to be

$$S(\gamma) = \exp_\otimes(v).$$

Now, suppose we have a discrete time series $\hat{\gamma} : [T+1] \to G$. Recall from Equation 3.2.4 that we may compute the discrete derivative by

$$\hat{\gamma}'_t = \log\left(\hat{\gamma}_t^{-1} \hat{\gamma}_{t+1}\right)$$

to obtain a discrete time series $\hat{\gamma}' : [T] \to \mathfrak{g}$. As discussed in Section 3.2.3, we interpret these discrete time series as continuous paths by interpolating using the exponential. Thus, the continuous interpretation of the discrete time series can be thought of as a concatenation of several exponential paths

$$\gamma = \exp(\hat{\gamma}'_1 t) * \exp(\hat{\gamma}'_2 t) * \ldots * \exp(\hat{\gamma}'_T t).$$

Therefore, by the above computation of the path signature of an exponential path and Chen's identity, we define the continuous path signature of the discrete time series to be

$$S(\hat{\gamma}) = \exp_\otimes(\hat{\gamma}'_1) \otimes \exp_\otimes(\hat{\gamma}'_2) \otimes \ldots \otimes \exp_\otimes(\hat{\gamma}'_T). \tag{3.3.8}$$

By using tensor operations, this formula provides an effective implementation for the computation of the path signature.

An alternative approach is to compute an approximation of the path signature for discrete time series.

Definition 3.3.6. Let $\hat{\gamma} : [T+1] \to G$. Suppose $\omega_1, \dots, \omega_N \in \mathfrak{g}^*$ form a basis of $\mathfrak{g}^*$. Let $I = (i_1, \dots, i_m)$ be a multi-index, where $i_j \in [N]$. Let $\hat{\gamma}' : [T] \to \mathfrak{g}$ be the discrete derivative of $\hat{\gamma}$. We define the discrete m-simplex with length T to be

$$\hat{\Delta}_T^m = \{(t_1, \dots, t_m) \in [T]^m \quad : \quad 0 \leq t_1 < t_2 < \dots < t_m \leq 1\}.$$

The discrete path signature of $\hat{\gamma}$ with respect to I is defined to be

$$\hat{S}^I(\hat{\gamma}) := \sum_{(t_1,\dots,t_m) \in \hat{\Delta}_T^m} \omega_{i_1}(\hat{\gamma}'_{t_1}) \dots \omega_{i_m}(\hat{\gamma}'_{t_m}),$$

where $t_1, \dots, t_m \in [T]$. The discrete path signature can be viewed as a map

$$\hat{S} : \mathrm{Seq}(G) \to T((\overline{\mathfrak{g}})).$$

We consider the discrete path signature since it is more amenable to computation. The approximations from [KO19] and discused in Section 1.3.5 extend immediately to the Lie group setting. In particular, [KO19] derived efficient algorithms to compute the discrete path signature kernel, and we extend these algorithms to Lie groups in Section 3.4.2.

3.3.8 *Path Transformations*

We now define the Lie group formulations of the path transformations from Section 1.3.6.

1. The *time transformation* is

$$\begin{aligned} T_{\mathrm{Time}} : C_e^{pr}(G) &\to \widetilde{P}(G \times \mathbb{R}) \\ \gamma_t &\mapsto (\gamma_t, t). \end{aligned} \tag{3.3.9}$$

2. The *identity initialized (IdInit) transformation* is

$$\begin{aligned} T_{\mathrm{IdInit}} : C^{pr}(G) &\to C_e^{pr}(G) \\ \gamma &\mapsto \ell_{\gamma_0} * \gamma, \end{aligned} \tag{3.3.10}$$

   where ℓ_g is the exponential path from the identity e to the point $g \in G$.

3. The *sliding window transformation* is

$$\begin{aligned} T_{\mathrm{SWin},m} : C^{pr}(G) &\to C^{pr}(G)^{m+1} \\ \gamma_t &\mapsto (\gamma_t, \gamma_{t-\tau}, \gamma_{t-2\tau}, \dots, \gamma_{t-m\tau}), \end{aligned} \tag{3.3.11}$$

where we define $\gamma_t = e$ if $t < 0$.

3.4 THE PATH SIGNATURE KERNEL FOR LIE GROUPS

In this section, we show that a normalized variant of the path signature can be used to define a universal and characteristic kernel for Lie group valued time series. Then, we show that the algorithms introduced in [KO19] for efficient computation of the signature kernel can also be used for Lie group valued paths.

3.4.1 *The Path Signature Kernel*

In this subsection, we will define the path signature kernel, and show that it is universal and characteristic. This was shown for the Euclidean case in [CO18], which states that these properties hold when studying paths evolving in a Hilbert space. Through our definition of the path signature for Lie groups, we provide a clarification: the space itself need not be a Hilbert space, but rather the space of tangent vectors must be a Hilbert space. In our case of Lie groups, this is the Lie algebra with a Riemannian metric. With this setup, the path signature kernel for Lie groups is universal and characteristic.

Theorem 3.4.1. *Let $\Lambda : \overline{T}_1(\mathbb{R}^n) \to \overline{T}_1(\mathbb{R}^n)$ be a tensor normalization. The normalized signature*

$$\Phi : \widetilde{C}^{pr}(G) \to T_1((\mathfrak{g})), \quad \Phi = \Lambda \circ S,$$

1. *is a continuous injection from $\widetilde{C}^{pr}(G)$ into a bounded subset of $\overline{T}_1(\mathbb{R}^n)$,*
2. *is universal to $\mathcal{F} := C_b(\widetilde{C}^{pr}(G), \mathbb{R})$ equipped with the strict topology, and*
3. *is characteristic to the space of finite regular Borel measures on $\widetilde{C}^{pr}(G)$.*

Proof. The fact that Φ is an injection follows from the injectivity of the path signature from Theorem 3.3.2 and the definition of the tensor normalization. Continuity follows from the stability property of Corollary 1.2.4. Next, we move on to universality. Define

$$L = 1 + \bigoplus_{m=1}^{\infty} (\mathfrak{g})^{\otimes m}$$

to be a dense subspace of $T_1((\mathfrak{g}))$ (note that L only contains finite linear combinations of tensors, whereas $T_1((\mathfrak{g}))$ contains all power series of tensors) and define

$$\mathcal{F}_0 = \{\langle \ell, \Phi(\cdot) \rangle : \widetilde{C}^{pr}(G) \to \mathbb{R} \ : \ \ell \in L\}.$$

We aim to show that $\mathcal{F}_0$ satisfies the hypotheses of the second point of Theorem 1.3.3. By the injectivity of Φ, the class of functions $\mathcal{F}_0$ separates points, and because the path signature is defined with constant term 1, the path signature is nonzero for all paths $\gamma \in \widetilde{C}^{pr}(G)$. Finally, by the shuffle product identity from Theorem 3.3.3, the class of functions $\mathcal{F}_0$ is closed under shuffle multiplication and is therefore a subalgebra of $C_b(\widetilde{C}^{pr}(G), \mathbb{R})$. Namely, let I and J be multi-indices, and e_I and e_J be the corresponding basis vectors in L. Then, we may define multiplication in $\mathcal{F}_0$ by

$$\langle e_I, \Phi(\cdot)\rangle\langle e_J, \Phi(\cdot)\rangle = \left\langle \sum_{K \in I \sqcup\!\sqcup J} e_K, \Phi(\cdot) \right\rangle,$$

which is closed. Thus, Φ is universal with respect to $\mathcal{F}$. Finally, by the duality in Theorem 1.3.2 and the third point of Theorem 1.3.3, the function class $\mathcal{F}$ is characteristic with respect to finite regular Borel measures on $\widetilde{C}^{pr}(G)$. □

Remark 3.4.1. Although this theorem is stated for tree-like equivalence classes of paths in G, by precomposing with the time transformation or identity start transformation discussed in Section 1.3.6, we can also obtain universal and characteristic feature maps that are not reparametrization or translation invariant.

3.4.2 *The Kernel Trick*

The *kernel trick* refers to an efficient method to compute kernels without having to compute explicit representations of elements in the feature space. Several efficient algorithms to compute the Euclidean path signature kernel are provided in [KO19], who state that their algorithms hold for path signatures computed for Hilbert space-valued data (their methods hold in more generality). The algorithms depend only on an inner product structure in the space where the integrals are being computed, namely in $\bar{\mathfrak{g}}$, and thus also hold in our present context of Lie group-valued data. In this section, we will provide an explicit generalization of their main algorithm.

As in Section 3.3.3, we will restrict ourselves to the truncated signature

$$S_M : C^{pr}(G) \to T^{(M)}(\bar{\mathfrak{g}}).$$

The *signature kernel truncated at level M* is defined to be

$$K_M : C^{pr}(G) \times C^{pr}(G) \to \mathbb{R}, \quad (\alpha, \beta) \mapsto \langle S_M(\alpha), S_M(\beta)\rangle \tag{3.4.1}$$

We will begin by simplifying the computation of the kernel for continuous paths.

Proposition 3.4.1. *The signature kernel can be computed as*

$$K_M(\alpha, \beta) = \sum_{m=0}^{M} \int_{(s,t)\in\Delta^m\times\Delta^m} \prod_{i=1}^{m} \langle \alpha'_{s_i}, \beta'_{t_i} \rangle_{\mathfrak{g}} \, ds \, dt, \tag{3.4.2}$$

where we view α'_t and β'_t as elements of the Lie algebra $\mathfrak{g}$, and the inner product in the integrand $\langle \cdot, \cdot \rangle_{\mathfrak{g}}$ is computed in the Lie algebra. Also, we denote $s = (s_1, \ldots, s_m)$ and $t = (t_1, \ldots, t_m)$ as elements of Δ^m, and write $ds := ds_1 \ldots ds_m$ and $dt := dt_1 \ldots dt_m$.

Proof. Let's consider the inner product at a single level m. Recall that $\pi_m : T((\mathfrak{g})) \to \mathfrak{g}^{\otimes m}$ is the projection on to the level m tensors, and $\langle \cdot, \cdot \rangle_m$ refers to the inner product on $\mathfrak{g}^{\otimes m}$. Then,

$$\begin{aligned}
&\langle \pi_m S(\alpha), \pi_m S(\beta) \rangle_m \\
&\qquad = \sum_{I:|I|=m} S^I(\alpha) \cdot S^I(\beta) \\
&\qquad = \sum_{I=(i_1,\ldots,i_m)} \int_{\Delta^m} \omega_{i_1}(\alpha'_{s_1}) \ldots \omega_{i_m}(\alpha'_{s_m}) ds \int_{\Delta^m} \omega_{i_1}(\beta'_{t_1}) \ldots \omega_{i_m}(\beta'_{t_m}) dt \\
&\qquad = \int_{(s,t)\in\Delta^m\times\Delta^m} \sum_{I=(i_1,\ldots,i_m)} \left[\omega_{i_1}(\alpha'_{s_1})\omega_{i_1}(\beta'_{t_1})\right] \ldots \left[\omega_{i_m}(\alpha'_{s_m})\omega_{i_m}(\beta'_{t_m})\right] \, ds \, dt \\
&\qquad = \int_{(s,t)\in\Delta^m\times\Delta^m} \prod_{i=1}^{m} \langle \alpha'_{s_i}, \beta'_{t_i} \rangle_{\mathfrak{g}} \, ds \, dt.
\end{aligned}$$

Then, adding up all of the levels, we get our desired result. □

As noted by [KO19], the expression in Equation 3.4.2 can be efficiently computed by a method that is similar to Horner's scheme for computing polynomial expressions. Suppose we wish to compute the expression $p(x) = \sum_{i=0}^{M} x^i$. By expanding this polynomial as

$$p(x) = 1 + x(1 + x(1 + \ldots + x(1 + x)))$$

where the recursion occurs M times and computing the brackets from the inside to the outside, we can evaluate the expression using M additions and M multiplications. In contrast, the naive computation of $p(x)$ would require M additions and $\frac{M^2+M}{2}$ multiplications. Note that we may write out this recursion explicitly as follows. Let

$$q_1 = 1 + x, \quad q_m = 1 + x q_{m-1}.$$

Then, we may write $p(x) = q_M$. We can significantly reduce the number of operations required to compute the integrals in Equation 3.4.2 by adapting this procedure.

Corollary 3.4.1. *Let*

$$Q_1(s,t) = 1 + \int_{\substack{s' \in [0,s] \\ t' \in [0,t]}} \langle \alpha'_{s'}, \beta'_{t'} \rangle_{\mathfrak{g}} ds' dt'. \tag{3.4.3}$$

Then, we recursively define

$$Q_m(s,t) = 1 + \int_{\substack{s' \in [0,s] \\ t' \in [0,t]}} Q_{m-1}(s',t') \langle \alpha'_{s'}, \beta'_{t'} \rangle_{\mathfrak{g}} ds' dt'. \tag{3.4.4}$$

The signature kernel can be computed as

$$K_M(\alpha, \beta) = Q_M(1,1). \tag{3.4.5}$$

Proof. We will consider the case of $M = 2$. Here, we have

$$\begin{aligned}
Q_2(1,1) &= 1 + \int_{\substack{s_1 \in [0,1] \\ t_1 \in [0,1]}} Q_1(s_1,t_1) \langle \alpha'_{s_1}, \beta'_{t_1} \rangle_{\mathfrak{g}} ds_1 dt_1 \\
&= 1 + \int_{\substack{s_1 \in [0,1] \\ t_1 \in [0,1]}} \left(1 + \int_{\substack{s_2 \in [0,s_1] \\ t_2 \in [0,t_1]}} \langle \alpha'_{s_2}, \beta'_{t_2} \rangle_{\mathfrak{g}} ds_2 dt_2 \right) \langle \alpha'_{s_1}, \beta'_{t_1} \rangle_{\mathfrak{g}} ds_1 dt_1 \\
&= 1 + \int_{\substack{s_1 \in [0,1] \\ t_1 \in [0,1]}} \langle \alpha'_{s_1}, \beta'_{t_1} \rangle_{\mathfrak{g}} ds_1 dt_1 + \int_{\substack{s_1 \in [0,1] \\ t_1 \in [0,1]}} \int_{\substack{s_2 \in [0,s_1] \\ t_2 \in [0,t_1]}} \prod_{i=1}^{2} \langle \alpha'_{s_i}, \beta'_{t_i} \rangle_{\mathfrak{g}} \, ds_1 ds_2 dt_1 dt_2 \\
&= K_2(\alpha, \beta).
\end{aligned}$$

The general proof proceeds in the same manner. By each successive unfolding of the definition of Q_m, we recover an additional summand in Equation 3.4.2. □

Next, we will consider the discrete formulation of this expression. For simplicity, we will consider discrete time series of the same length, though the following results also hold when the two discrete time series are of different lengths. Suppose $\hat{\alpha}, \hat{\beta} : [T+1] \to G$, and let $\hat{\alpha}', \hat{\beta}' : [T] \to \mathfrak{g}$ be their corresponding discrete derivatives. Following the notation in Section 3.3.7 for the discrete signature, we define the *discrete signature kernel truncated at level M* to be

$$\hat{K}_M(\hat{\alpha}, \hat{\beta}) := \langle \hat{S}_M(\hat{\alpha}), \hat{S}_M(\hat{\beta}) \rangle_M. \tag{3.4.6}$$

Notice we are using the discrete signature given in Definition 3.3.6. Then, the discrete analogues of Proposition 3.4.1 and Corollary 3.4.1 are as follows.

Proposition 3.4.2. *The discrete signature kernel can be computed as*

$$\hat{K}_M(\hat{\alpha}, \hat{\beta}) = \sum_{m=0}^{M} \sum_{(s,t)\in \hat{\Delta}_T^m \times \hat{\Delta}_T^m} \prod_{i=1}^{m} \langle \alpha'_{s_i}, \beta'_{t_i} \rangle_{\mathfrak{g}}. \tag{3.4.7}$$

Corollary 3.4.2. *Let*

$$\hat{Q}_1(s,t) = 1 + \sum_{s'\in[s],\, t'\in[t]} \langle \alpha'_{s'}, \beta'_{t'} \rangle_{\mathfrak{g}} \tag{3.4.8}$$

and recursively define

$$\hat{Q}_m(s,t) = 1 + \sum_{s'\in[s],\, t'\in[t]} \hat{Q}_{m-1}(s',t') \langle \alpha'_{s'}, \beta'_{t'} \rangle_{\mathfrak{g}}. \tag{3.4.9}$$

The discrete signature kernel can be computed as

$$\hat{K}_M(\hat{\alpha}, \hat{\beta}) = \hat{Q}_M(T,T). \tag{3.4.10}$$

The proofs for these discrete formulas proceed in exactly the same manner as their continuous counterparts. This final recursive formula for the kernel provides an efficient recursive algorithm for the computation of the discrete signature, which can be found in [KO19].

3.5 EXPERIMENTS

In this section, we provide two detailed experiments to demonstrate the universal and characteristic properties of the path signature. First, we consider the human action recognition problem from computer vision, using a Lie group representation of the data. We find that the path signature method is simple to implement, achieves comparable classification performance to shallow learning methods, and provides an interpretable feature set. Second, we perform a kernel two-sample hypothesis test aiming to distinguish between two different random walks on $SO(3)$. Here, we find that the path signature for $SO(3)$ significantly outperforms the same hypothesis test done using the Euclidean representation of $SO(3)$.

3.5.1 *Human Action Recognition*

In this subsection, we aim to utilize the universal property of the signature kernel to study a classification problem. Namely, we will be working in the domain of *human action recognition*, in which several recent works, such as [Hua+17], [RWF18] and [Li+19],

have used Lie group representations and deep learning to achieve state of the art results. Euclidean path signatures have also been used to study the real valued and joint-based representation of this problem in [Yan+19], and preliminary experiments using Lie group path signatures have been performed in [CLT19], though detailed quantative comparisons with larger datasets were not provided.

Our goal in this section is to demonstrate the utility, simplicity and interpretability of the Lie group path signature using the G3D-Gaming dataset from [BMA12]. Thus, our focus will be on establishing baseline results using support vector machines and random forests, and our main comparison will be with the "shallow" learning methods described in [VAC14] and [VC16]. We achieve comparable classification results using a random forest algorithm and the second level signature as a feature set. A major advantage of the second level signature is that we can interpret this feature set in terms of the underlying movement, and we provide a brief discussion of this analysis.

3.5.1.1 *Lie group representation*

In this experiment, we aim to classify actions based on human skeletal motion data, which we represent as a $SO(3)^k$ valued time series. This Lie group representation of human motion was introduced in [VAC14] and [VC16], which describes a pose as an element of $SO(3)^k$. The relative rotation of two body parts is an element of $SO(3)$, and by recording the relative rotation of k pairs of body parts, we can describe the full pose of a human.

A skeleton $S = (V, E)$ can be described using a set of joints $V = \{v^i\}_{i=1}^{n_V}$, where $v^i \in \mathbb{R}^3$, and a set of body parts $E = \{e^i\}_{i=1}^{n_E}$, where each body part is a function $e^i : \{0, 1\} \to V$. We may consider $e^i(0)$ and $e^i(1)$ to respectively be the start and end point of the body part e^i. Additionally, we will denote the unit vector describing the direction from the start to the end of the body part by

$$\hat{e}^i = \frac{e^i(1) - e^i(0)}{\|e^i(1) - e^i(0)\|}.$$

We will consider all pairs of body parts (e^i, e^j) that share a joint, such that $e^i(c_1) = e^j(c_2)$ for some $c_1, c_2 \in \{0, 1\}$, so k is the number of adjacent pairs of body parts. To obtain the rotation matrix for a chosen pair (e^i, e^j), we rotate the global coordinate system (with minimum rotation) such that $\hat{e}^i$ is the x-axis. Then, the rotation matrix $R_{i,j} \in SO(3)$ is the minimum rotation from $\hat{e}^i$ to $\hat{e}^j$ in this coordinate system. By repeating this for all adjacent pairs, we obtain an element of $SO(3)^k$, and further repeating this for all time steps, we can represent this motion as a time series in $SO(3)^k$.

3.5.1.2 *Data and preprocessing*

The dataset we use is the G3D-Gaming dataset from [BMA12]. This contains 663 sequences of 20 different gaming motions performed by 10 subjects, and each subject performed each action at least two times. The 3D locations of 20 joints are provided for every frame. However, the number of frames for each recorded action varies widely from 3 frames to 330 frames. We use the code provided in [VC16] to generate the $SO(3)$ representation.

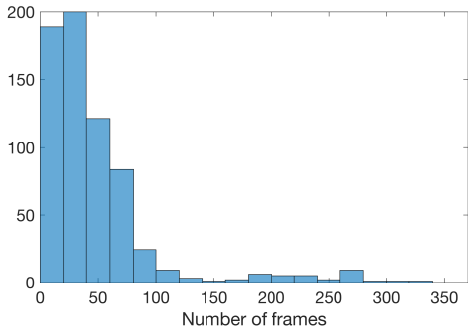


Figure 3.3: Histogram of number of frames per recorded action in G3D dataset.

Note that in [VAC14] and [VC16], all pairs of body parts are used in the Lie group representation, which results in $k = 342$ pairs for this dataset. In contrast, we only use all adjacent pairs of body parts, which we call the *primary pairs*, resulting in $k = 18$ pairs for this dataset. We use significantly less data because the path signatures take into account the relationships between all input pairs, so we have information regarding non-primary pairs through the higher order signature terms. The numbering of the primary pairs are given in the following figure.

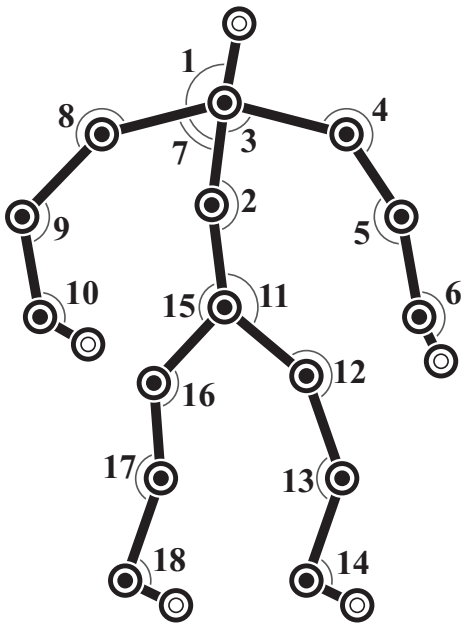


Figure 3.4: Numbering of the primary pairs of body parts.

We categorize these 18 primary pairs into 5 classes.

Body	1-2
L Arm	3-6
R Arm	7-10
L Leg	11-14
R Leg	15-18

Table 3.1: The body regions corresponding to groups of primary pairs.

Extensive preprocessing of this data is performed in [VAC14] and [VC16] in order to deal with several difficulties. During the training stage, the following steps were taken.

1. Each time series is resampled via interpolation so that all time series have a fixed length.
2. A nominal curve in $SO(3)^k$ is generated for each action class.
3. To handle issues of rate variation and temporal misalignment, dynamic time warping (DTW) is used to warp each time series to its corresponding nominal curve.
4. A rolling and unwrapping procedure is computed with respect to its nominal curve to obtain a curve in the Lie algebra $\mathfrak{so}(3)^k$.

5. (Optional) A Fourier temporal pyramid (FTP) representation of the Lie algebra curve may be computed to further deal with temporal misalignment.

A classifier such as a support vector machine (SVM) is then trained for every action class (one vs. rest). In the testing stage, the data is preprocessed with respect to all nominal curves and the corresponding SVM is used for prediction. This amounts to a large preprocessing cost, especially for test samples, which must be preprocessed with respect to all action classes.

In contrast, we perform minimal preprocessing since we can compare time series with varying numbers of frames using path signatures, and the issues of rate variation and temporal misalignment are handled by reparametrization invariance.

3.5.1.3 *Results*

Following [VAC14; VC16], we use a cross-subject test setting, where we use half of the subjects for training, and the other half for testing. All of the reported classification results are averaged over ten different combinations of the train/test split. We perform the classification using a kernel SVM, as well as a random forest. For the kernel SVM, we report the results using the signature kernel truncated at level 6. We use the Julia implementation of `scikit-learn`, with the `SVC` implementation for support vector machines, which uses the one-against-one approach [KPD90] for multi-class classifcation. For the random forest, we compute the level 2 signature of the time series and treat it as a feature set. We use the random forest implementation in the Julia `DecisionTrees` package. We follow the suggested default random forest hyperparameters in [PWB19], and use 1,000 trees, $\sqrt{n_f}$ features for each tree, where n_f is the total number of input features, a maximum depth of 100, and use 70% of the data to train each tree. A tensor renormalization is used for all path signature computations as described in Proposition 1.3.1, and using the function

$$\psi(x) = \begin{cases} x^2 & \text{if } x \leq \sqrt{M} \\ M + M^{1+a}(M^{-a} - x^{-a})/a & \text{if } x > \sqrt{M} \end{cases}$$

where $M = 4$ and $a = 1$ as in [CO18].

We apply each algorithm to several embeddings of the discrete time series. In addition to using the raw time series, we also use the identity initialized (IdInit) embedding to remove translation invariance, and the sliding window (SWin) embedding with 1 to 6 lags.

Based on our results, a random forest trained using level 2 signatures outperforms an SVM trained using level 6 signatures for all embeddings. This may be due to the fact that random forests are in general better suited for multi-class classification tasks, since the SVM approach is to split the multi-class problem into $\binom{20}{2}$ binary classification tasks.

We note by using the random forest classifier, and introducing lags, we are able to achieve results which are comparable to the 87.95% accuracy of [VC16], given the fact that we are

	Random Forest	SVM
Raw	69.99%	65.20%
IdInit	80.48%	77.86%
SW, 1 lag	80.32%	79.24%
SW, 2 lags	82.71%	80.27%
SW, 3 lags	83.77%	80.84%
SW, 4 lags	84.47%	81.08%
SW, 5 lags	84.86%	81.44%
SW, 6 lags	85.56%	81.44%

Table 3.2: Classification results using a random forest and SVM for different embeddings.

using significantly less input data, minimal preprocessing, and default hyperparameters on the random forest.

Remark 3.5.1. In Section 1.3.6.3, we mentioned that one possible explanation for the strong empirical performance of the sliding window embedding is the breaking of translation invariance. In these results, we can isolate the effect of breaking translation invariance by using the IdInit embedding. We see that for both the random forest and the SVM, the performance improves significantly after using the IdInit embedding when compared to the raw time series. The sliding window embedding with 1 lag also significantly improves the performance when compared to the raw time series, but we see that the increase is comparable to that of the IdInit embedding. When we increase the number of lags, the path signature is able to capture more information by integrating with respect to past values, and thus the performance continues to improve. [Fer21] suggested the problem of explaining why the sliding window embedding performs well in practice. This empirical evidence indicates that one such reason is the breaking of the translation invariance of the path signature.

3.5.1.4 *Interpretation of the second level signature matrix*

In addition to comparable classification accuracy, the second level signature matrix affords an interpretable feature set. This is one of the highlights of this method, as none of the cited classification methods are easily interpretable. Namely, in both the narrow and deep learning approaches, the input to the classification algorithm is the stacked (and in some cases preprocessed) Lie group valued time series, resulting in an extremely high dimensional vector. In this section, we will discuss the interpretation of the raw second level signature matrix, which was used for the random forest classification.

The second level path signature of a path $\gamma \in C^{pr}(G)$, where G is N-dimensional, can be viewed as a $(N \times N)$ matrix where the (i, j) entry is simply $S^{(i,j)}(\gamma)$. For these applications, our Lie group is $G = SO(3)$[18], which has $N = 54$ dimensions. We will consider the absolute value of the averaged second level signature matrices for each action class of the G3D data set. Specifically, because we are focused only on the interpretation in this section, we will use all of the data when computing averages. Also, we take the absolute value to simplify the interpretation of these matrices. We begin by considering a single action class (walk) in order to describe how to read the matrix.

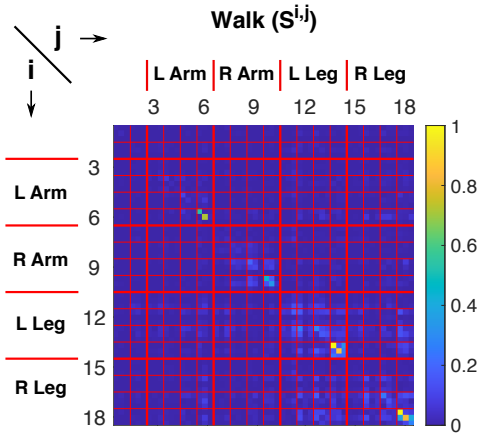


Figure 3.5: Averaged absolute second level signature matrix for the action class "walk."

Here we have visualized the (54×54) matrix as a heatmap, where the entries are arranged in accordance with standard matrix notation. The thin red grid lines separate the matrix into (3×3) blocks, where each triplet of entries corresponds to a single primary pair. The numerical labels on the axes enumerate the blocks (or primary pairs), which correspond to the primary pairs as shown in Figure 3.4. The thicker red grid lines are used to separate the different categories of primary pairs, as outlined in Table 3.1, and labelled in the figure. The diagonal entries of the matrix measure the square of the cumulative rotation of a given basis direction for a given primary pair. The off diagonal entries measure the positive or negative influence of the basis direction i on the basis direction j, as defined in Section 3.3.5. For a closer look at this example, we isolate the blocks corresponding to the left and right legs, and reduce the color threshold.

The blocks with the largest magnitude are the diagonal blocks 14 and 18, which correspond to the joints of the left and right foot (see Figure 3.4. In addition, many of the off-diagonal entries are nonzero, which corresponds to the action of walking. While walking, we alternate moving our left and right legs, and the signature matrix measures this as an influence of rotations about joints in one leg on the rotations about joints in the other leg.

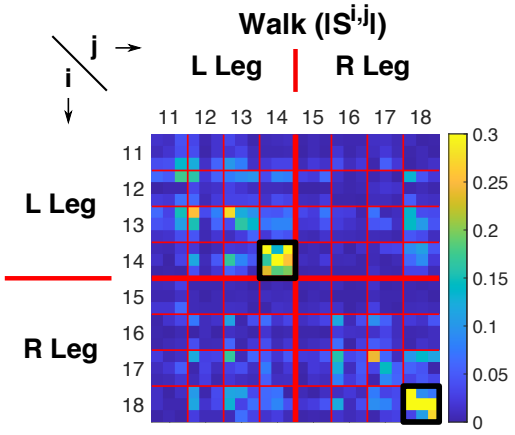


Figure 3.6: Averaged absolute second level signature matrix for the action class "walk."

In Figure 3.7, we have plotted the absolute value of the averaged second level signature matrix for all action classes in the data set. We omit the labelling to simplify the figure, but all labelling is the same as Figure 3.5. In particular, the colors range from 0 to 1.

3.5.2 *Random Walks on Lie Groups*

In this experiment, we aim to take advantage of the characteristic property of the signature kernel and perform a kernel two-sample hypothesis testing problem, as described in [Gre+12]. We aim to distinguish between two distinct discrete random walks on $SO(3)$, or in other words, between two distinct probability distribution on $P(SO(3))$. In addition, by using the matrix representation of $SO(3)$, we can also view these as probability distributions on $P\mathbb{R}^9$, and we also perform the hypothesis testing in this context. We find that for this problem, the path signature defined using left-invariant forms on $SO(3)$ significantly outperforms the Euclidean signature.

3.5.2.1 *Hypothesis testing with MMD*

Suppose μ and ν are two probability measures on $C^{pr}(G)$ and we have i.i.d. samples $\{X^i\}_{i=1}^n$ of μ and $\{Y^i\}_{i=1}^n$ of ν, each with n samples. Note that we use the superscript to index the separate samples, since the subscript is reserved for the time parameter. We wish to distinguish between the null hypothesis $H_0 : \mu = \nu$ against the alternative hypothesis $H_1 : \mu \neq \nu$. This is done by computing the unbiased empirical MMD (Equation 1.3.4) between our two collections of samples, and comparing this against a chosen threshold. A Type I error occurs when the test falsely rejects the null hypothesis, and a Type II error

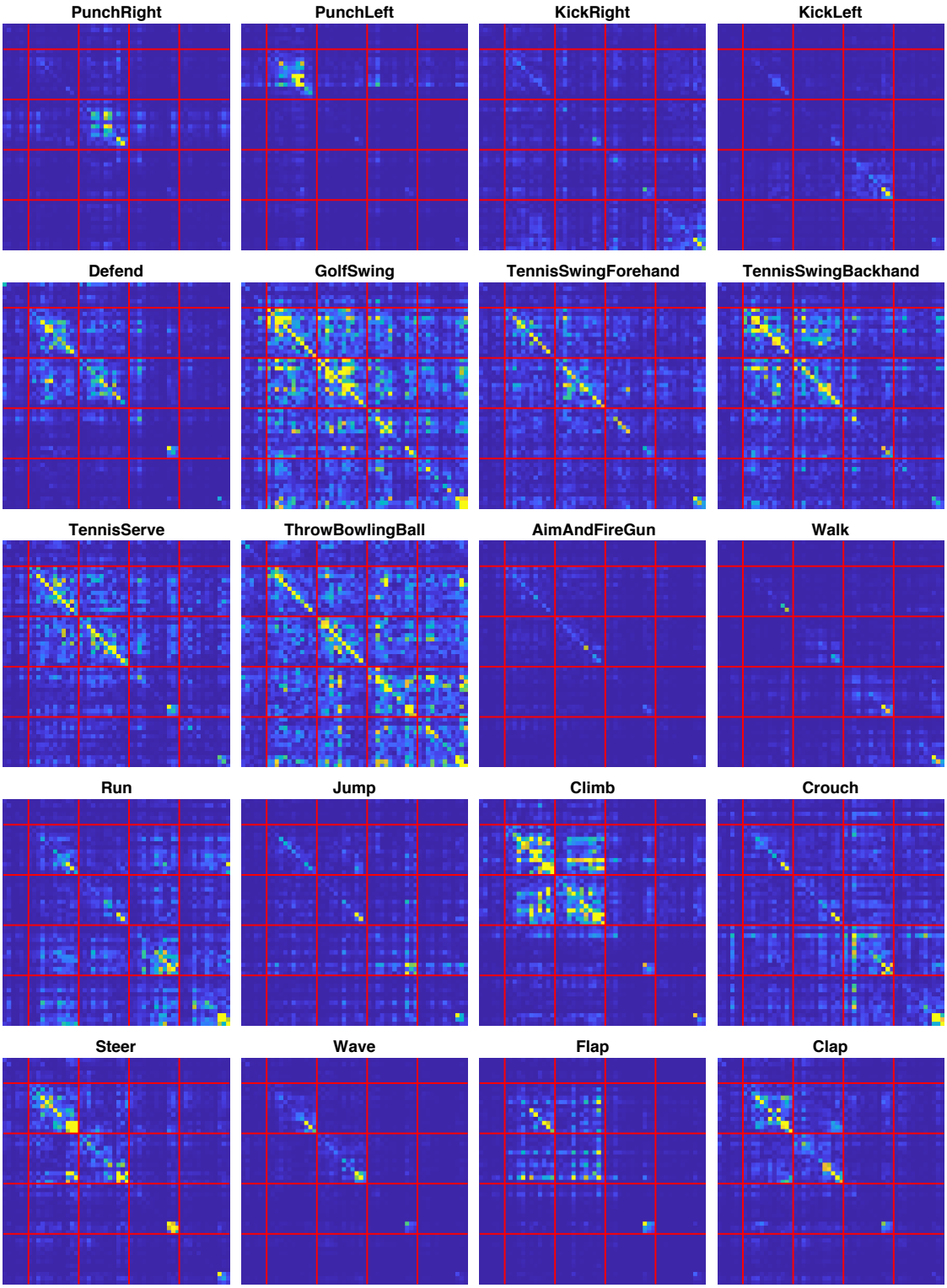


Figure 3.7: Averaged absolute S^2 matrices for all actions in G3D dataset.

occurs when the test falsely accepts the null hypothesis. The *level* α of a test is a chosen upper bound on the probability of a Type I error.

The threshold for the MMD is chosen such that the test is *consistent*, meaning the test achieves a level α and a Type II error of zero in the asymptotic large sample limit. One method for choosing the threshold is by using large deviation bounds for the MMD under the assumption that the null hypothesis is true, which are derived in [Gre+12]. However, because such bounds must hold for all measures, the threshold chosen in this way is conservative and may not yield optimal results with a finite number of samples. An alternative method is to use a data-dependent threshold. Our method here will be to generate a null distribution for the MMD by performing a permutation test, and computing the $1 - \alpha$ quantile for a level α test. Alternative methods for deriving data-dependent thresholds are explored in [Gre+09; Gre+12].

3.5.2.2 *Random walk generation*

For this experiment, we will consider discrete random walks on $SO(3)$. Each walk is initialized by sampling a point uniformly at random from $SO(3)$. A method to sample uniformly from $SO(3)$ is given in [DS87]. Next, we sample T i.i.d. points $\{v_i\}_{i=1}^{T}$ from the unit sphere in the Lie algebra $\mathfrak{so}(3)$. We use a unimodal probability distribution on S^2, called the von Mises-Fisher distribution. This is parametrized by the mean direction $x \in S^2$, and the concentration parameter $\kappa \geq 0$, which describes the concentration of the distribution about the mean.

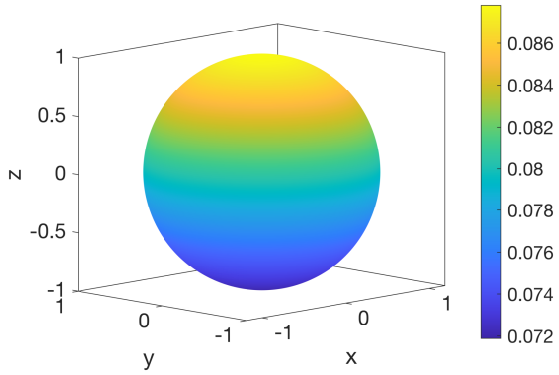


Figure 3.8: The von-Mises Fisher density on S^2 with mean direction $x = (0, 0, 1)$ and $\kappa = 0.1$.

We may think of this as a random walk with drift in the mean direction. A numerically stable method of sampling from this distribution is discussed in [Jak12]. The random walk $\hat{\gamma} : [T+1] \to SO(3)$ is defined recursively as

$$\hat{\gamma}_{i+1} = \hat{\gamma}_i \exp(cv_i),$$

where $c > 0$ is the step size and $\hat{\gamma}_1$ is a randomly sampled point on $SO(3)$.

3.5.2.3 *Results*

We perform two classes of tests.

1. H_0 is false. Here, we attempt to distinguish between distributions with different mean directions. We take $x_\mu = (1, 0, 0)$ and $x_\nu = (0, 1, 0)$.
2. H_0 is true. Here we take the mean directions of both distributions to be $x_\mu = x_\nu = (1, 0, 0)$.

All distributions will have a concentration parameter of $\kappa = 0.1$. For each test, we will sample $n = 50$ random walks from each distribution, and each random walk will have $L = 100$ steps. To generate the null distribution of the MMD, we perform a permutation test with 2,000 permutations for a given set of samples. A level of $\alpha = 0.05$ is chosen, so we compute the 0.95 quantile of the null distribution as the threshold used for the MMD. The path signature is truncated at level 4 in the MMD computation. We perform 1,000 tests for each class, and each test is done using both Lie group and Euclidean path signatures.

The following table provides the error rates (false positive/negative) for the two classes of tests using the two methods.

	Lie Group	Euclidean
H_0 false	6.1%	83.7%
H_0 true	5.7%	4.9%

Table 3.3: Error rates for hypothesis testing. Each test was run 1,000 times.

Additionally, we provide histograms that summarize the test results. The test distributions show the distribution of MMD_u over the 1,000 independent trials. The null distribution shown is generated from a permutation test for a single trial. The red line shows the 0.95 quantile, and represents the threshold for that trial. The histograms for H_0 being false are shown in Figure 3.9 and the histograms for H_0 being true are shown in Figure 3.10.

We find that the Lie group path signatures significantly outperform Euclidean path signatures. This is due to the fact that the Euclidean representation of the data is ill-suited for this problem. We are aiming to detect a slight drift in the direction of the rotation, which is a translation invariant feature in $SO(3)$. However, this is not a translation invariant feature in the Euclidean representation of the problem, so the effect of the drift is confounded in the Euclidean path signature.

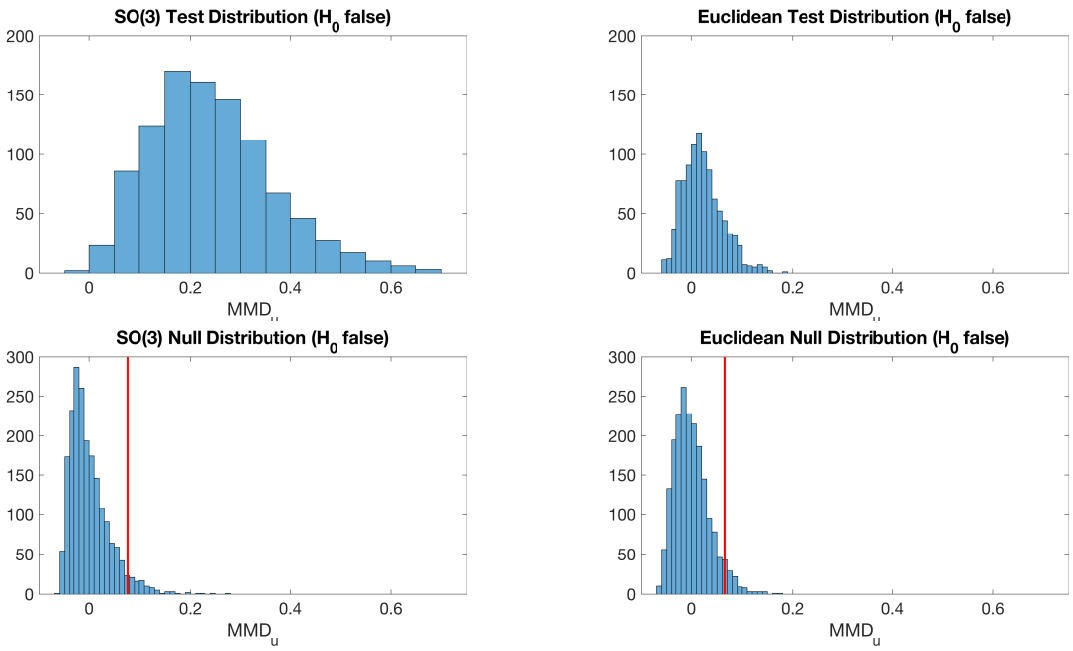


Figure 3.9: Test (top) and null (bottom) distributions of MMD_u when H_0 is false.

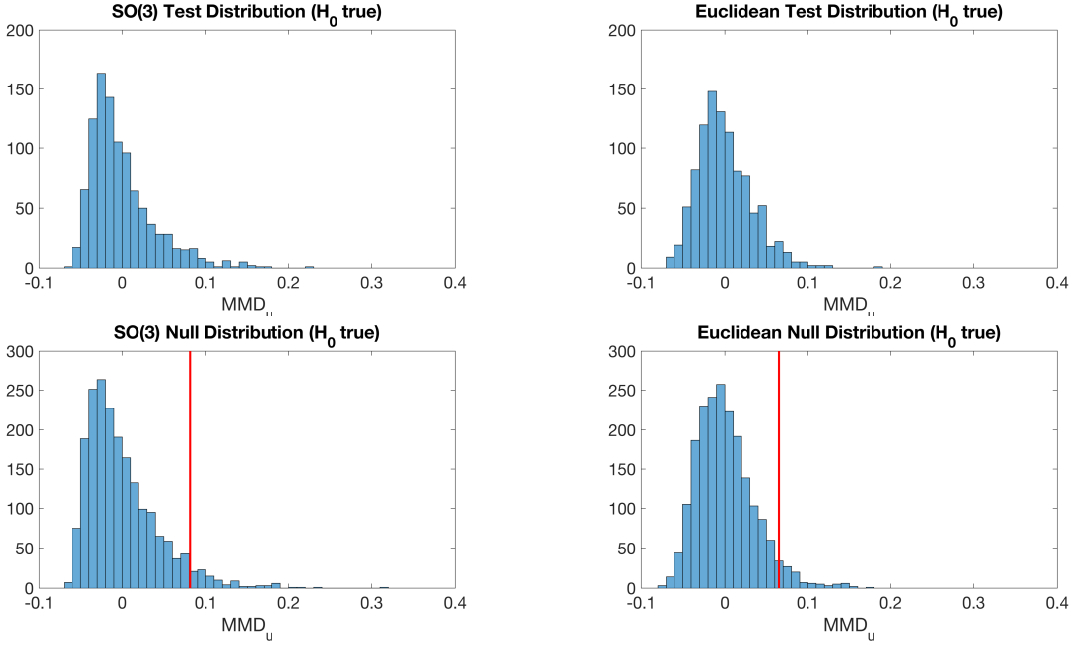


Figure 3.10: Test (top) and null (bottom) distributions of MMD_u when H_0 is true.

3.6 CONCLUSION

We have defined path signature for Lie group valued time series, and studied several of its properties, the main result being the universal and characteristic properties of the signature kernel. By defining signature using only the derivative of the path, computational techniques from Euclidean valued paths to Lie group valued paths can be exported cleanly. Our theory is validated using two detailed experiments highlighting both the universal and characteristic properties, showing that the path signature has strong empirical performance, while providing an interpretable feature set which can be used to better understand the underlying phenomena.

Lie group valued data is ubiquitous; however, in previous studies analyzing such data, the analysis pipeline can be complicated due to the ostensible complexities when dealing with Lie groups (as described in Section 3.5.1). Our derivations show that Lie group valued data can be treated in a manner nearly identical to standard Euclidean valued data.

The work in this paper provides the foundations for further studying properties and applications of path signatures for Lie groups and more. We highlight two directions for possible future research.

1. Although the path signature is a complete characterization of paths in Euclidean space, the inversion problem of recovering a path given its path signature is difficult [LX17; Cha+17]. [PSS19] studies a restricted and approximate inversion problem from an algebro-geometric perspective of the path signatures, which was initiated in [AFS19].

   One may consider a similar question for Lie groups. Suppose we fix a path $\gamma \in C^{pr}(G)$, and we are given the path signature of a path $\alpha \in C^{pr}(H)$, where $\alpha = F\gamma$, where $F : G \to H$ is a Lie group homomorphism. Is it possible to recover the Lie group homomorphism F?

2. In [Che58], the path signature is defined for a manifold M by choosing a collection of 1-forms $\{\omega_i\}_{i=1}^N$ (as no natural Lie algebra basis is available). Choose $x \in M$ to be a basepoint, and let $S : C_x^{pr}(M) \to T((\mathbb{R}^N))$ be the path signature defined with respect to these 1-forms and let $\alpha, \beta \in C_x^{pr}(M)$. If the 1-forms span the cotangent bundle of M at every point $x \in M$, Chen's injectivity result states that $S(\alpha) = S(\beta)$ if and only if α and β are tree-like equivalent (in a slightly modified sense). The injectivity theorem given in Theorem 3.3.2 is a customization of Chen's result to the case $M = G$ since a basis of $\mathfrak{g}^*$ spans the cotangent bundle at every point.

   Recently, there has been interest in studying time series evolving on manifolds, and the development of a path signature kernel for paths on manifolds would provide a powerful tool for geometric time series analysis. One of the difficulties

of this definition would be the representation of data on manifolds and the choice of 1-forms used. One could begin with path signatures on *parallelizable* manifolds, which by definition admit a smooth basis of vector fields and 1-forms. This would encompass all orientable 3-dimensional manifolds.

4

SIGNATURES, LIPSCHITZ-FREE SPACES, AND PATHS OF PERSISTENCE DIAGRAMS

4.1 INTRODUCTION

In complex systems, collective behavior emerges from the interactions between individual elements [D'O+06; NJG12], but the state of an individual element often contains very little information. It is therefore reasonable, and even desirable, to describe the state of the system by using abstract or qualitative organizational features. The negligible contribution of individual agents suggests that it should be possible to infer such features from a relatively small representative subsample of the constituent agents, and this is often the approach applied in practice.

One tool for characterizing these organizational structures is persistent homology [Cha+16; EH08; Ghr08; Ott+17; ZC05], a common tool from topological data analysis (TDA). It is applied by constructing a nested family of spaces X_ϵ which model the structure of the underlying system at a scale parameter $\epsilon > 0$; such a family satisfies $X_{\epsilon'} \hookrightarrow X_\epsilon$ if $\epsilon' < \epsilon$ and is called a *filtration*. By applying the k^{th} homology functor, H_k, to this filtration, we can track how homology classes *persist* as the scale parameter ϵ varies. In particular, we record the scale parameters b and d at which an individual homology class appears or disappears respectively; these scale parameters are called the *birth* and *death* times of a homology class. The multiset of the birth and death times for all of the k^{th} homology classes in a data set is called a *persistence diagram*. Persistence diagrams provide a multi-scale summary of the system's organization, while discarding precise information about, and even the identity of, individual elements in a system. Thus, they are well-suited to characterizing the state of a multi-agent complex system.

While the details of individual agents can often be abstracted away when describing a static snapshot of a complex system, the same cannot be said for the temporal structure when characterizing dynamics. Even if the ensemble of states the system attains is similar, their ordering of these states must be carefully recorded to provide understanding. For example, a system in which agents move into consensus versus one in which they move out of consensus are operating under very different dynamic principles, even though

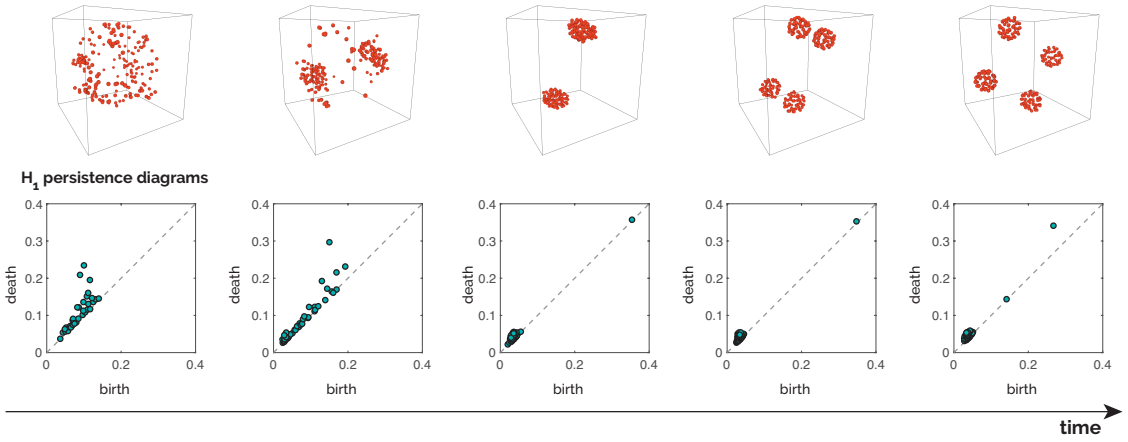


Figure 4.1: Snapshots of a path of H_1 persistence diagrams (bottom row) generated from a time-varying point cloud in $\mathbb{R}^3$ (top row).

reordering the time dimension so one runs backward will result in a similar collection of coarse system organizations, and thus similar topological signatures. *Paths of persistence diagrams*[1] [CEM06; Mun13; PH15; Rie+20; TZH15], which is shown in Figure 4.1, provide a hybrid form of measurement that retains this temporal[2] information while discarding the relatively uninformative data about the individual agents.

However, as mathematical objects, paths of persistence diagrams are themselves not necessarily straightforward to understand and measure. Neither spaces of persistence diagrams nor spaces of paths are easy to vectorize, nor to develop standard statistical methods for. To address this problem, in this paper we to develop a collection of informative, stable, and computable features for the space of paths of persistence diagrams, and to apply those features to investigate and to demonstrate how to classify the dynamic behaviors of complex systems.

The study of path spaces is a classical topic in algebraic topology [Che77]. Recent work has leveraged and extended these ideas to provide sophisticated feature sets in the form of *path signatures* for time series analysis in machine learning [CK16; KO19; Ada+17; Mor+21; LG20; Kid+19]. The path signature is a reparametrization-invariant characterization of paths valued in a Banach space, and has been shown to be *universal*

1 Here, we specifically mean paths in the space of persistence diagrams, as compared to the concept of a *persistence vineyard* [CEM06], which retains information about how individual homology classes evolve over time. Such granular information is often difficult to obtain in the context of real data, particularly when individual elements of the system cannot be tracked.

2 The methods we will develop work without change for any one-dimensional parameterized family of persistence diagrams, however this terminology collides with that of the scale parameter so we will favor the dynamic terminology throughout.

and *characteristic* [CO18], two properties which provide strong theoretical guarantees in the context of kernel methods.

In order to define an intrinsic path signature for time-varying persistence diagrams, we first develop an isometric embedding of the space of persistence diagrams into a Banach space.[3] The classical space of finite persistence diagrams has the structure of a commutative monoid, and thus we must enrich this space with additional algebraic and analytic properties. In particular, we enrich this space with additive inverses to obtain an abelian group, with a notion of scaling to obtain a vector space, and a norm which induces a complete metric to obtain the desired Banach space. We survey existing enrichments [BE20b; DL21; MMH11] which provide some of these additional properties while lacking others. Then, we draw on the notion of the Lipschitz-free space (also called the Arens-Eells space) from functional analysis [Wea18; CDW16], a universal construction on metric spaces which provides a Banach space with the desired properties; in particular, we show that there is an isometric embedding of finite persistence diagrams into the corresponding Lipschitz-free space. While explicit representations of the relevant Lipschitz-free spaces are not available, we suggest a construction by extending the method of group completion introduced in [BE20b].

Having developed the theoretical foundations necessary to apply the theory of path signatures to time-varying persistence diagrams, we describe how this framework can be used to define explicit feature maps for studying time-varying data. Motivated by the perspective of persistence diagrams as measures [Cha+16; DL21], we introduce the *moment map*, a new feature map for static persistence diagrams which is injective and Lipschitz continuous on bounded persistence diagrams. Composing this feature map with the path signature, we obtain a computable representation for paths of persistence diagrams which can be used in data science applications.

Finally, we demonstrate the efficacy of this approach by performing a detailed numerical experiment to recover estimates of the parameters for the 3D D'Orsogna model of swarm dynamics [D'O+06]. In particular, we highlight the ability of persistent homology to capture global features of a system's organization from sparse observations by estimating these parameters using subsets of the positions of the agents, randomly selecting which agents' positions to use at each time point.

4.1.1 *Contributions*

The aim of this paper is to combine the tools of persistent homology and path signatures to study the dynamic topological structure of filtrations of topological spaces (from a

3 A Hilbert space embedding would provide even more structure, though this has been shown to be impossible without significant metric distortion [BW19; Wag21; TS20].

theoretical perspective) and of data sets (from a computational perspective). In particular, our main contributions are as follows:

- We connect the the theory of persistence diagrams with the existing theory of Lipschitz-free spaces [Wea18] in order to obtain an isometric embedding of the space of persistence diagrams into a Banach space. To the authors' knowledge, this connection is new, and so we provide a detailed discussion of Lipschitz-free spaces in relation to existing generalizations of the space of persistence diagrams.
- We introduce the space of bounded variation paths of persistence diagrams and show that the path signature provides a universal and characteristic feature map for this space. In particular, the expected signature characterizes measures on this path space, and can thus be used for statistical applications. In addition, this characterization provides a foundation for the investigation of stochastic dynamics using persistent homology via analogy to the theory of rough paths.
- We introduce a new feature map for persistence diagrams motivated by the perspective of persistence diagrams as measures. Specifically, by computing the moments of bounded persistence diagrams, we obtain an injective stable feature map for the space of persistence diagrams.
- Combining the moment map with the discrete path signature, we obtain a computable feature map for paths of persistence diagrams. Furthermore, this feature map is stable with respect to the 1-variation metric on paths, and is also universal and characteristic.
- We develop code to perform signature kernel computations for paths of persistence diagrams, which can be found at:

  `https://github.com/ldarrick/paths-of-persistence-diagrams`.

4.1.2 *Previous and Related Work*

The question of how to study dynamic data using persistent homology reaches back to the early development of the subject. Perhaps the earliest initial approach is through persistence vineyards, introduced by Cohen-Steiner et al. [CEM06]. While persistence vineyards capture substantial information about the evolution of topological summaries, they require a more complex representation and their use poses substantial computational and theoretical challenges. Later, Munch et al. [Mun+15] studied the paths of persistence diagrams through the lens of statistical and probabilistic features, such such as Fréchet means. An alternative approach to the study of dynamic metric spaces was considered by

Kim and Mémoli [KM20] by building a 3-parameter filtration, and using methods from multiparameter persistence.

Because their features are often understood qualitatively, complex time-varying systems are a popular potential application of persistent homology. One application of broad interest is the dynamic topological structure of swarming/flocking behavior, which was studied by Topaz et al. [TZH15] and Bhaskar et al. [Bha+19]. For purposes of interpretability and comparison with existing methods, we use simulations of swarm data our computational experiments. Other examples of time-varying persistence diagrams arise, for example, in neuroscience, where they have been used to study time-varying fMRI [Rie+20] and as EEG data [Yoo+16]. In another direction, persistent homology has been used to detect periodicity in time series through sliding window embeddings [PH15; Per+15]. This method has been used to detect chatter in physical cutting processes such as turning and milling [KM16] and biphonation in videos of vibrating vocal folds [TP18].

Path signatures were originally developed as the *iterated integral model for path spaces* by Chen [Che58] to characterize paths on manifolds, which was subsequently generalized to a de Rham-type cochain algebra for loop spaces [Che77]. Path signatures were later used as the foundation for the theory of rough paths by Lyons [Lyo98], and has been transformed into a powerful feature set for time series data [CK16]. Chevyrev and Oberhauser [CO18] showed the signature is universal and characteristic, providing a theoretical justification for its use in the context of kernel methods, facilitated by Kiraly and Oberhauser's [KO19] development of efficient algorithms to compute truncated signature kernels. A more recent method to compute untruncated signature kernels using PDEs was developed by Cass et al. [Cas+21]. The connection between path signatures and persistent homology was initiated by Chevyrev, Nanda and Oberhauser [CNO20] to develop feature maps for static persistence diagrams.

Interest in applying statistical and machine learning tools in the context of topological data analysis has also driven recent work in on generalized spaces of persistence diagrams. Bubenik and Elchesen [BE20a; BE20b] study the space of persistence diagrams from an algebraic perspective, and develop spaces of persistence diagrams with additional structure through the use of universal properties. The connection between persistence diagrams and the theory of optimal transport was made explicit by Divol and Lacombe [DL21]. In particular, they view persistence diagrams as Radon measures, and develop a complete metric space for such persistence measures.

Recent work shows the space of finite persistence diagrams does not admit an isometry into a Hilbert space [TS20] for any p-Wasserstein metric with $p \in [1, \infty]$. Indeed, when $p \in (2, \infty]$, there does not even exist a *coarse embedding* into a Hilbert space [BW19; Wag21]. Finally, the limitations of bi-Lipschitz embeddings of finite bounded persistence diagrams into a Hilbert space are discussed in [CB19]. Given these limitations, our construction of

isometric embeddings of spaces of persistence diagrams into Banach spaces are among the strongest positive results one could reasonably hope for.

4.2 A BANACH SPACE FOR PERSISTENCE DIAGRAMS

Formally, a *persistence module* is a functor $F : (\mathbb{R}, \leq) \to \mathsf{Vect}$, where we view the poset $(\mathbb{R}, \leq)$ as a category. The *decomposition theorem* [Cha+16] tells us that, up to a reasonable notion of equivalence, the information in this functor can be summarized by a *persistence diagram*, a multiset of points supported on a half plane of $\mathbb{R}^2$. In the context of real data, it usually suffices to consider the space $\mathcal{D}^+_{\text{fin}}$ of *finite* persistence diagrams: finite multisets of points supported on a half plane. The space of finite persistence diagrams is a commutative monoid with a metric structure given by p-Wasserstein distances.

However, there are several other algebraic and analytic properties which are desirable for machine learning and statistics that $\mathcal{D}^+_{\text{fin}}$ lacks. First, a notion of scaling by positive or real constants is required to normalize outputs and to compute averages for finite samples. Furthermore, in order to compute expectations of measures of persistence diagrams, we require a notion of Lebesgue integration. Second, completeness is necessary to study notions of convergence. Third, the ability to represent differences between persistence diagrams is required to define p-variation metrics for paths of persistence diagrams. Finally, in combination with the three previous properties, the existence of a norm connects the algebraic structure to the metric structure of the space of persistence diagrams. Our goal in this section is to enrich the space of finite persistence diagrams to include all four of these properties, which will yield a Banach space of persistence diagrams.

We consider various *enrichments* of $\mathcal{D}^+_{\text{fin}}$, which satisfy some of the desirable properties described above. Several of the enrichments of the space of persistence diagrams we will describe have been previously considered, and their properties are summarized in Figure 4.2, where the arrows denote inclusion. The spaces $\mathcal{D}^+_{\text{fin}}, \mathcal{D}^+_1, \mathcal{M}^+_{\text{fin}}, \mathcal{M}_{\text{fin}}$, and $\mathcal{M}^+_1$ have appeared previously in the literature, while $\mathcal{F}$, in the center of the Venn diagram, is a new addition: the *Lipschitz-free space* of a metric space [Wea18]. Before we describe these constructions in detail, we briefly survey the appearances of these spaces in the literature:

1. **Scaling:** The perspective of persistence diagrams as measures was introduced in [Cha+16], though these measures were still restricted to take values in $\mathbb{N}$. The extension of this perspective to Radon measures $\mathcal{M}$ was initiated in [DL21], which allows for a notion of scaling. The simplest generalization is $\mathcal{M}^+_{\text{fin},1}$, the space of finite Radon measures with finite 1-Wasserstein distance to the empty diagram. In this generalized setting, Divol and Lacombe were able to characterize convergence in the Wasserstein metric, define expectations of probability measures valued in persistence diagrams, and consider a law of large numbers.

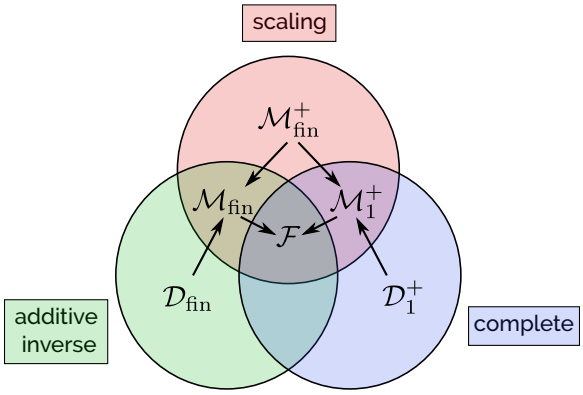


Figure 4.2: Enrichments of the space $\mathcal{D}_{\text{fin}}^{+}$ of finite persistence diagrams and the properties of each; arrows indicate inclusion. Note that $\mathcal{D}_{\text{fin}}^{+}$ includes into all of these enriched spaces.

2. **Completeness:** The finiteness assumption on persistence diagrams introduces an obstruction to being a complete metric space. A resolution to this problem was developed in [MMH11] by considering diagrams with countably many points, but with a finite 1-Wasserstein distance to the empty diagram, denoted by $\mathcal{D}_{1}^{+}$. A similar condition was studied for Radon measures with infinite mass in [DL21], and this extended space of measures, denoted $\mathcal{M}_{1}^{+}$, was also shown to be complete.

3. **Additive Inverse:** The addition of an additive inverse was considered in [BE20b] through the use of group completion of commutative monoids, which provides formal differences to objects in a monoid. The group completion of $\mathcal{D}_{\text{fin}}^{+}$ is the space of diagrams with possibly negative coefficients on points, called *virtual persistence diagrams* and denoted $\mathcal{D}_{\text{fin}}$. Furthermore, the group completion of finite Radon measures is simply the space of signed finite Radon measures, denoted $\mathcal{M}_{\text{fin},1}$.

4.2.1 *Generalized Spaces of Persistence Diagrams*

We will describe each of these enrichments in the more general context of *metric pairs* (X, d, A), where (X, d) is a metric space and $A \subset X$ is an arbitrary subset. By working with metric pairs, our definitions and results immediately extend to enriched spaces of *generalized persistence diagrams* [BdS15], which are summaries of *generalized persistence modules*, functors $F : \mathsf{P} \to \mathsf{A}$, where P is a preordered set viewed as a category, and A is an abelian category. Discussion on applications to generalized persistence diagrams is in [BE20a].

We introduce several enriched spaces of persistence diagrams from an algebraic perspective, following [BE20a; BE20b]. We say that a metric pair (X, d, A) is *admissible* if X is a locally compact Polish metric space and A is closed, implying that $X - A$ is open. These

are the hypotheses required to state the results of [DL21]. We will consider the following constructions on a metric space X:

1. $D^+(X)$, the free commutative monoid
2. $D(X)$, the free abelian group
3. $V(X)$, the free vector space
4. $M^+(X)$, the space of finite Radon measures
5. $\overline{M}^+(X)$, the space of Radon measures
6. $M(X)$, be the space of finite signed Radon measures

In particular, we treat $M^+(X)$ as a commutative monoid and $\overline{M}^+(X)$, $M(X)$ as vector spaces. Now, suppose (X, d, A) is an admissible metric pair. We define the following *spaces of enriched persistence diagrams on X relative to A*:

1. $\mathcal{D}^+_{\text{fin}}(X, A) \coloneqq D^+(X)/D^+(A)$,
2. $\mathcal{D}_{\text{fin}}(X, A) \coloneqq D(X)/D(A)$,
3. $\mathcal{V}_{\text{fin}}(X, A) \coloneqq V(X)/V(A)$,
4. $\mathcal{M}^+_{\text{fin}}(X, A) \coloneqq M^+(X)/M^+(A)$,
5. $\mathcal{M}^+(X, A) \coloneqq \overline{M}^+(X)/\overline{M}^+(A)$,
6. $\mathcal{M}_{\text{fin}}(X, A) \coloneqq M(X)/M(A)$.

We begin by considering the unsigned enrichments, which are related by

$$\mathcal{D}^+_{\text{fin}}(X, A) \hookrightarrow \mathcal{M}^+_{\text{fin}}(X, A) \hookrightarrow \mathcal{M}^+(X, A)$$

and define an extended metric on these spaces by defining an extended metric on $\mathcal{M}^+(X, A)$.

Definition 4.2.1. Let (X, d, A) be an admissible metric pair. Let $\pi_1, \pi_2 : X \times X \to X$ be the projection map on to the first and second coordinates respectively, and $(\pi_1)_*, (\pi_2)_* : \overline{M}^+(X \times X) \to \overline{M}^+(X)$ be the induced pushforward maps of measures. Let $\mu, \nu \in \mathcal{M}^+(X, A)$. An *admissible coupling between μ and ν* is an element $\sigma \in \overline{M}^+(X \times X)$ such that

$$(\pi_1)_*\sigma = \mu \pmod{\overline{M}^+(A)}$$
$$(\pi_2)_*\sigma = \nu \pmod{\overline{M}^+(A)}.$$

Let the space of all admissible couplings between μ and ν be denoted by $\text{ACpl}(\mu, \nu)$. The *p-cost of σ with respect to d* is defined by

$$\text{Cost}_p[d](\sigma) \coloneqq \left(\int_{X \times X} d(x, y)^p d\sigma(x, y) \right)^{1/p}. \tag{4.2.1}$$

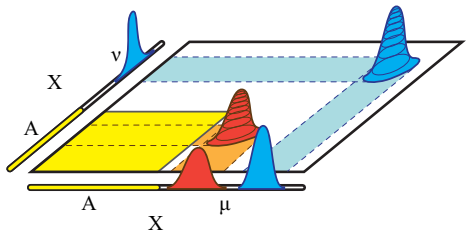


Figure 4.3: A visualization of an admissible coupling between the measures μ and ν, shown on the two coordinate axes. The plane represents the product space $X \times X$, with $A \times A$ in yellow. The colors of the coupling measure correspond to the colors of the marginalization to μ and ν.

Then, the *p-Wasserstein distance* $W_p^\partial[d] : \mathcal{M}^+(X,A) \times \mathcal{M}^+(X,A) \to [0,\infty]$ is defined as follows. For $\mu, \nu \in \mathcal{M}^+(X,A)$,

$$W_p^\partial[d](\mu,\nu) := \inf_{\sigma \in \mathrm{ACpl}(\mu,\nu)} \mathrm{Cost}_p[d](\mu,\nu). \tag{4.2.2}$$

Finally, the *p-persistence* $\mathrm{Pers}_p : \mathcal{M}^+(X,A) \to \mathbb{R}$ is

$$\mathrm{Pers}_p(\mu) := W_p^\partial[d](\mu, 0),$$

where $0 \in \mathcal{M}^+(X,A)$ is the zero measure.

While this definition does not use the standard notation for admissible couplings used in [FG10; DL21], it is equivalent as shown in [BE20b]. Moreover, this definition specializes to extended distance functions and p-persistence functions on $\mathcal{D}^+_{\mathrm{fin}}(X,A)$ and $\mathcal{M}^+_{\mathrm{fin}}(X,A)$. By restricting to measures of finite p-persistence

$$\mathcal{M}^+_{\mathrm{fin},p}(X,A) := \{\mu \in \mathcal{M}_{\mathrm{fin}}(X,A) \,:\, \mathrm{Pers}_p(\mu) < \infty\},$$
$$\mathcal{M}^+_p(X,A) := \{\mu \in \mathcal{M}(X,A) \,:\, \mathrm{Pers}_p(\mu) < \infty\},$$

the p-Wasserstein distance function $W_p^\partial[d]$ becomes an honest metric. The space $\mathcal{M}^+_p(X,A)$ is shown to be complete in [DL21], and the completion of $(\mathcal{D}^+_{\mathrm{fin}}(X,A), W_p^\partial[d])$ includes diagrams with countably many points with finite p-persistence [MMH11; BE20b], which is denoted $(\mathcal{D}^+_p, W_p^\partial[d])$.

Remark 4.2.1. In order to obtain the standard definition of persistence diagrams, we set

$$X = \Omega := \{(x_1,x_2) \in \mathbb{R}^2 \,:\, x_1 \le x_2\} \tag{4.2.3}$$
$$A = \partial\Omega := \{(x,x) \in \mathbb{R}^2\}, \tag{4.2.4}$$

A finite persistence diagram $\alpha \in \mathcal{D}^+_{\text{fin}}(\Omega, \partial\Omega)$ is then a coset in $D^+(\Omega)/D^+(\partial\Omega)$, consisting of a finite number of points in the interior of Ω, and an arbitrary number of points on the boundary $\partial\Omega$. A matching $\sigma \in D^+(\Omega \times \Omega)$ between two diagrams $\alpha, \beta \in \mathcal{D}^+_{\text{fin}}(\Omega, \partial\Omega)$ and its associated cost $\text{Cost}_p[d](\alpha, \beta)$ corresponds to the ordinary definitions for persistence diagrams, since σ is permitted to match points in α and β with any point on the diagonal $\partial\Omega$. Therefore, we recover the ordinary p-Wasserstein distance $W^\partial_p[d]$ for persistence diagrams.

For a metric space (X, d) where X is a monoid, we say that the metric is *translation-invariant* if $d(x + z, y + z) = d(x, y)$ for all $x, y, z, \in X$.

Lemma 4.2.1. *Let (X, d, A) be an admissible metric pair. Then $W^\partial_1[d]$ is a translation-invariant metric on $\mathcal{M}^+_{fin,1}(X, A)$.*

Proof. Let $\mu, \nu, \tau \in \mathcal{M}^+_{\text{fin},1}(X, A)$, and let $(\widetilde{X}, \tilde{d}, *)$ be the quotient of (X, d, A). From 4.2.3, there exist measures $\tilde{\mu}, \tilde{\nu} \in M^+(\widetilde{X})$ with equal mass such that $W^\partial_1[d](\mu, \nu) = W_1[d](\tilde{\mu}, \tilde{\nu})$. Furthermore, let $\tilde{\tau} \in M^+(\widetilde{X})$ be the measure such that $\tilde{\tau}(*) = 0$, and agrees with τ on $X - A$. Then, we also have $W^\partial_1[d](\mu + \tau, \nu + \tau) = W_1[d](\tilde{\mu} + \tilde{\tau}, \tilde{\nu} + \tilde{\tau})$.

Next, the Kantorovich-Rubinstein duality for the ordinary 1-Wasserstein metric $W_1[d]$ [Vil09] states that

$$W_1[d](\tilde{\mu}, \tilde{\nu}) = \sup_{\|\psi\|_{\text{Lip}} \leq 1} \left[\int_{\widetilde{X}} \psi d\tilde{\mu} - \int_{\widetilde{X}} \psi d\tilde{\nu} \right].$$

Therefore, we have

$$\begin{aligned} W_1[d](\tilde{\mu} + \tilde{\tau}, \tilde{\nu} + \tilde{\tau}) &= \sup_{\|\psi\|_{\text{Lip}} \leq 1} \left[\int_{\widetilde{X}} \psi d(\tilde{\mu} + \tilde{\tau}) - \int_{\widetilde{X}} \psi d(\tilde{\nu} + \tilde{\tau}) \right] \\ &= \sup_{\|\psi\|_{\text{Lip}} \leq 1} \left[\int_{\widetilde{X}} \psi d\tilde{\mu} - \int_{\widetilde{X}} \psi d\tilde{\nu} \right] \\ &= W_1[d](\tilde{\mu}, \tilde{\nu}). \end{aligned}$$

□

We can extend this translation invariant metric to the signed enrichments using the *group completion of monoids*, a technique introduced in [BE20b].

Definition 4.2.2. Let $(M, +)$ be a commutative monoid. The *Grothendieck group of M* (or the *group completion of M*, denoted $K(M)$ is defined as follows. Let $\sim$ denote the equivalence relation on $M \times M$ given by $(a, b) \sim (c, d)$ if there exists some $k \in M$ such that

$$a + d + k = b + c + k.$$

As a set, we define $K(M) \coloneqq (M \times M)/\sim$, and we denote the equivalence class of (a, b) by $a - b$. The binary operation on $K(M)$ is defined to be

$$(a - b) + (c - d) \coloneqq (a + c) - (b + d).$$

Lemma 4.2.2. *Suppose M is a commutative monoid. The group completion $(K(M), +)$ is an abelian group.*

The main result we will use from Bubenik and Elchesen is the combination of Proposition 3.1, 3.8 and 3.11 in [BE20b].

Proposition 4.2.1. *Let $K(M)$ be the Grothendieck group of a commutative monoid $(M, +)$ equipped with a nontrivial ($d \neq 0$) and translation-invariant metric d. Then, the function $\rho : K(M) \times K(M) \to [0, \infty)$ defined by*

$$\rho(a - b, c - d) \coloneqq d(a + d, b + c) \tag{4.2.5}$$

is the unique translation invariant metric extending d to $K(M)$.

This leads to the extension of the translation invariant metric to signed enrichments.

Proposition 4.2.2. *Let (X, d, A) be an admissible metric pair. Suppose $\mu \in \mathcal{M}_{fin}(X, A)$, and let $\mu = \mu^+ - \mu^-$ be its Jordan decomposition into two positive finite Radon measures. Let*

$$\mathcal{M}_{fin,1}(X, A) \coloneqq \{\mu \in \mathcal{M}_{fin}(X, A) \,:\, \mathrm{Pers}_1(\mu^+), \mathrm{Pers}_1(\mu^-) < \infty\}.$$

The extended function $W_1^\partial[d] : \mathcal{M}_{fin,1}(X, A) \times \mathcal{M}_{fin,1}(X, A) \to \mathbb{R}$ defined by

$$W_1^\partial[d](\mu, \nu) \coloneqq W_1^\partial[d](\mu^+ + \nu^-, \nu^+ + \mu^-), \tag{4.2.6}$$

where the right hand side is the 1-Wasserstein metric for positive measures, is a translation invariant metric on $\mathcal{M}_{fin,1}(X, A)$. Furthermore, we extend the 1-persistence to $\mathcal{M}_{fin,1}(X, A)$ by defining

$$\mathrm{Pers}_1(\mu) \coloneqq W_1^\partial[d](\mu, 0),$$

where 0 is the empty measure on $\mathcal{M}_{fin,1}(X, A)$.

Proof. First, we note that $\mathcal{M}_{\mathrm{fin},1}(X, A)$ is the group completion

$$K(\mathcal{M}^+_{\mathrm{fin},1}(X, A)).$$

Thus, by Proposition 4.2.1, the function defined in Equation (4.2.6) is a translation-invariant metric on $\mathcal{M}_{\mathrm{fin},1}(X, A)$. □

The metric and 1-persistence can then be restricted to the the other signed enrichments, which are related by

$$\mathcal{D}_{\text{fin}}(X, A) \hookrightarrow \mathcal{V}_{\text{fin}}(X, A) \hookrightarrow \mathcal{M}_{\text{fin},1}(X, A).$$

As discussed in [BE20b], the translation-invariance of $W_1^\partial[d]$ on positive measures is required to define the translation-invariant extension. Furthermore, the translation-invariant metric can be used to define a norm.

Lemma 4.2.3. *The space $\mathcal{M}_{fin,1}(X, A)$ is a normed vector space with norm $\|\mu\|_1 := \text{Pers}_1(\mu)$ such that $\|\mu - \nu\|_1 = W_1^\partial[d](\mu, \nu)$.*

Proof. By the translation invariance of $W_1^\partial[d]$, we have

$$\|\mu - \nu\|_1 = W_1^\partial[d](\mu - \nu, 0) = W_1^\partial[d](\mu, \nu).$$

The properties of the norm are then immediate from the properties of the metric. □

The space $\mathcal{M}_{\text{fin},1}(X, A)$ has most of the properties we require: a notion of scaling, additive inverses, and the existence of a norm which corresponds to the 1-Wasserstein metric. However, the following example shows that the completion of $\mathcal{M}_{\text{fin},1}(X, A)$ cannot be directly formulated as a space of signed measures.

Example 4.2.1. The following example is adapted from [PRT19], which also considered Wasserstein norms, though it uses a different notion of generalized Wasserstein distances.

Let $X = \{(x, y) \in \mathbb{R}^2 : y \geq 0\}$, d be the standard Euclidean metric, and $A = \{(x, 0) : x \in \mathbb{R}\}$. Next, consider the following sequence of elements in $\mathcal{M}_{\text{fin},1}(X, A)$:

$$\mu_n = \sum_{i=1}^{n} \delta_{(i+2^{-i},1)} - \delta_{(i-2^{-i},1)}.$$

Then, we have

$$W_1^\partial[d](\mu_n, \mu_{n+k}) = \sum_{i=n+1}^{n+k} 2^{-(i-1)} \leq \sum_{i=n+1}^{\infty} 2^{-(i-1)} = 2^{-(n-1)},$$

so the sequence is Cauchy. However, this sequence does not converge to any object in $K(\mathcal{M}_1^+(X, A))$. The limiting object can be formally described as a difference of two measures

$$\mu^+ = \sum_{i=1}^{\infty} \delta_{(i+2^{-i},1)}, \quad \mu^- = \sum_{i=1}^{\infty} \delta_{(i-2^{-i},1)}.$$

However, both of these measures have infinite 1-norm, so $\mu^+ - \mu^- \notin K(\mathcal{M}_1^+(X,A))$, where $K(\cdot)$ denotes the group completion of monoids.

Due to this obstruction, we will now turn to an alternative description of the completion in terms of Lipschitz-free spaces.

4.2.2 *Quotient Metric Spaces*

In order to relate these enriched spaces of persistence diagrams for metric pairs to Lipschitz-free spaces, we must reframe these constructions in the context of pointed metric spaces by taking quotients.

Definition 4.2.3. Let (X,d,A) be a metric pair. We define the *quotient metric space* $(\widetilde{X},\tilde{d},*)$ as a pointed metric space as follows. Let $\widetilde{X} \coloneqq (X-A)\cup\{*\}$ denote the quotient set obtained by collapsing A to a point denoted by $*$, which we treat as the basepoint. The metric $\tilde{d}$ on $\widetilde{X}$ is defined by

$$\tilde{d} \coloneqq \min\left\{d(x,y), d(x,A)+d(y,A)\right\},$$

where $d(x,A) \coloneqq \inf_{y\in A}\{d(x,y)\}$. Furthermore, we will abuse notation and view the metric d as a function on the $\widetilde{X}\times\widetilde{X}$ given by $(x,y)\mapsto d(x,y)$ for $x,y\in X-A$, $(x,*)\mapsto d(x,A)$, and $(*,*)\mapsto 0$.

This is a special case of a general construction for quotient metric spaces [Wea18], and is shown to be a metric in this particular case in [BE20a]. Note that pointed metric spaces $(\widetilde{X},\tilde{d},*)$ are special cases of metric pairs, and all the definitions above for enriched spaces still hold. When the basepoint is clear, we will suppress it from the notation for the enriched spaces of diagrams. For instance, persistence diagrams of pointed metric spaces $(\widetilde{X},\tilde{d},*)$ will be denoted by

$$\mathcal{D}_{\text{fin}}^+(\widetilde{X}) \coloneqq \mathcal{D}_{\text{fin}}^+(\widetilde{X},*).$$

Some of the enriched spaces of the metric pair can be related to those of its quotient through the ordinary Wasserstein metric. Here, we focus on $\mathcal{M}_{\text{fin},1}^+(X,A)$, since infinite mass Radon measures defined on (X,d,A) may no longer be Radon measures when projected onto its quotient.

Example 4.2.2. We have avoided the discussion of infinite mass persistence measures from the perspective of quotient metric spaces due to the fact that some of these measures in $\mathcal{M}_1(X,A)$ may not be Radon measures on the quotient $(\widetilde{X},\tilde{d},*)$. In particular, Radon measures $\tilde{\mu}\in\mathcal{M}_1(\widetilde{X},*)$ must be locally finite, meaning for every $x\in\widetilde{X}$, there exists some

open neighborhood U of x such that $\mu(U) < \infty$. However, there exist Radon measures $\mu \in \mathcal{M}_1(X, A)$ which have infinite mass close to A. Indeed, let

$$X = \{(x, y) \in \mathbb{R}^2 : y \geq 0\}$$
$$A = \{(x, 0) \in \mathbb{R}^2\},$$

and consider the measure

$$\mu = \sum_{i=1}^{\infty} \delta_{(i, 2^{-i})},$$

where $\delta_{(x,y)}$ denotes the Dirac measure at $(x, y) \in X$. Note that

$$\|\mu\|_1 = \sum_{i=1}^{\infty} 2^{-i} = 1,$$

so that $\mu \in \mathcal{M}_1(X, A)$. However, for any open neighborhood U of the subset A, we have $\mu(U) = \infty$. Therefore, the pushforward $q_*\mu$ under the quotient map $q : (X, A) \to (\widetilde{X}, *)$ is not a Radon measure.

Definition 4.2.4. Let (X, d) be a metric space and suppose $\mu, \nu \in M^+(X)$ such that $\mu(X) = \nu(X)$. Let $\pi_1, \pi_2 : X \times X \to X$ be the projections to the first and second factor. The set of *couplings* between μ and ν is defined to be

$$\mathrm{Cpl}(\mu, \nu) := \left\{\sigma \in \mathcal{M}^+_{\mathrm{fin}}(X \times X) : (\pi_1)_*\sigma = \mu, (\pi_2)_*\sigma = \nu\right\}.$$

The *p-Wasserstein distance* between μ and ν is

$$W_p[d](\mu, \nu) := \inf_{\sigma \in \mathrm{Cpl}(\mu,\nu)} \left(\int_{X \times X} d(x, y)^p d\sigma(x, y) \right)^{1/p}.$$

The following proposition in [DL21] shows that we can compute partial Wasserstein distances using the ordinary Wasserstein distance for measures.

Proposition 4.2.3 ([DL21])**.** *Let (X, d, A) be a metric pair and let $(\widetilde{X}, \tilde{d}, *)$ be its quotient. Let $\mu, \nu \in \mathcal{M}^+_{fin,p}(X)$ and $r_\mu = \mu(X - A), r_\nu = \nu(X - A)$, which is well defined since it does not depend on the equivalence class. Let $r \geq r_\mu + r_\nu$. Define the modifications $\tilde{\mu}, \tilde{\nu} \in M^+(\widetilde{X})$ by*

$$\tilde{\mu} = \mu + (r - r_\mu)\delta_*$$
$$\tilde{\nu} = \nu + (r - r_\nu)\delta_*,$$

where δ_* is the Dirac measure at the basepoint $*$. Then,

$$W_p^\partial[d](\mu,\nu) = W_p[d](\tilde{\mu},\tilde{\nu}).$$

Corollary 4.2.1. *Let (X,d,A) be a metric pair and $(\widetilde{X},\tilde{d},*)$ be its quotient metric space. Let $q : \mathcal{M}_{fin,p}^+(X,A) \to \mathcal{M}_{fin,p}^+(\widetilde{X})$ be the bijective map induced by the quotient map. Suppose $\mu,\nu \in \mathcal{M}_{fin,p}^+(X,A)$, then*

$$W_p^\partial[d](\mu,\nu) = W_p^\partial[\tilde{d}](q(\mu),q(\nu)).$$

Proof. Let $\tilde{\mu},\tilde{\nu} \in M^+(\widetilde{X})$ be the modifications of μ and ν as in the previous proposition. Then,

$$W_p^\partial[d](\mu,\nu) = W_p[d](\tilde{\mu},\tilde{\nu}) = W_p^\partial[\tilde{d}](q(\mu),q(\nu)),$$

since $\tilde{\mu},\tilde{\nu}$ are also modifications of $q(\mu)$ and $q(\nu)$. □

4.2.3 *Lipschitz-Free Spaces*

We will now discuss the construction of a Lipschitz-free space of a metric space [Wea18], which is a Banach space whose linear structure coincides with the metric structure of the underlying space. We will view this as an enriched space of persistence diagrams which contains the enrichments previously discussed. We have kept the notation in this section consistent with the previous section, but this differs slightly from the standard notation for Lipschitz-free spaces. The material in this section is primarily adapted from [Wea18; CDW16]. Note that Lipschitz-free spaces are called Arens-Eells spaces in [Wea18].

Definition 4.2.5. Let $(X,d,*)$ be a pointed metric space. The *Lipschitz-free space* of X, denoted $\mathcal{F}(X)$, is the completion of $\mathcal{V}_{\text{fin}}(X,*)$ with respect to the 1-Wasserstein norm.

As the name implies, the Lipschitz-free space is related to the space of Lipschitz functions. Given a function $f : X \to \mathbb{R}$, its *Lipschitz constant* is defined to be

$$\|f\|_{Lip} := \sup_{x\neq y\in X}\left\{\frac{|f(x)-f(y)|}{d(x,y)}\right\} \in [0,\infty]. \tag{4.2.7}$$

We define the space of Lipschitz functions on $(X,d,*)$ to be

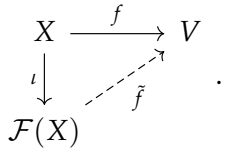
$$\text{Lip}_0(X) := \{f : X \to \mathbb{R} \,:\, \|f\|_{Lip} < \infty, f(*) = 0\} \tag{4.2.8}$$

where we treat the Lipschitz constant as a norm on $\mathrm{Lip}_0(X)$. Furthermore, $\mathrm{Lip}_0(X)$ is a Banach space with respect to this Lipschitz norm. The Lipschitz-free space can be characterized by the following universal property.

Theorem 4.2.1 ([Wea18]). *Let $(X, d, *)$ be a pointed metric space. Then the map*

$$\begin{aligned} \iota : X &\to \mathcal{F}(X) \\ x &\mapsto \delta_x \end{aligned}$$

isometrically embeds X in $\mathcal{F}(X)$. If V is any Banach space, and $f : X \to V$ is a Lipschitz map which preserves the basepoint, there exists a unique bounded linear map $\tilde{f} : \mathcal{F}(X) \to V$ such that $\tilde{f} \circ \iota = f$; in other words, the following diagram commutes



Furthermore, $\|\tilde{f}\|_{Lip} = \|f\|_{Lip}$.

The unique extension is defined by first extending f to $\overline{f} : \mathcal{V}_{\mathrm{fin}}(X) \to V$, which is defined as follows. Any $\alpha \in \mathcal{V}_{\mathrm{fin}}(X)$ can be written as

$$\alpha = \sum_{i=1}^{n} a_i \delta_{x_i},$$

where all the $x_i \in X$ are distinct from each other and the basepoint. Then, we define

$$f(\alpha) := \sum_{i=1}^{n} a_i f(x_i).$$

This can be seen as an extension to the universal properties discussed for $\mathcal{D}^{+}_{\mathrm{fin}}(X)$ and $\mathcal{D}_{\mathrm{fin}}(X)$ discussed in [BE20a; BE20b]. Furthermore, this provides a simple method to prove the stability of maps from $\mathcal{F}(X)$ into a Banach space, extending the characterization of Lipschitz maps on $\mathcal{M}_1^{+}(X, A)$ discussed in [DL21].

There is an equivalent definition of the 1-Wasserstein norm by viewing $\mathcal{V}_{\mathrm{fin}}(X)$ as a subspace of $\mathrm{Lip}_0(X)^*$ and using the dual norm. Given $f \in \mathrm{Lip}_0(X)$, and $\alpha \in \mathcal{V}_{\mathrm{fin}}(X)$, the pairing is defined to be

$$\langle \alpha, f \rangle := \sum_{x \in X} f(x).$$

Note that this pairing is well defined since for any $f \in \mathrm{Lip}_0(X)$, we have $f(*) = 0$, so the pairing does not depend on the representative of the equivalence class of α. The following equivalence is often called the *Kantorovich-Rubinstein duality* for the 1-Wasserstein distance, which is extended to the Lipschitz-free space.

Lemma 4.2.4. *For any $\alpha \in \mathcal{V}_{fin}(X)$, we have*

$$\|\alpha\|_1 = \sup_{f \in \mathrm{Lip}_0(X)} \frac{\langle \alpha, f \rangle}{\|f\|_{Lip}}.$$

Furthermore, the dual norm uniquely extends to a norm on $\mathcal{F}(X)$ which coincides with $\|\cdot\|_1$.

Now that $\mathcal{F}(X)$ is realized as a Banach space situated within $\mathrm{Lip}_0(X)^*$, a natural question is whether $\mathcal{F}(X)$ is simply the entire dual space. In fact, $\mathcal{F}(X)$ is the *predual* of $\mathrm{Lip}_0(X)$.

Proposition 4.2.4 ([Wea18])**.** *Let (X, d, x_0) be a pointed metric space. Then $\mathcal{F}(X)^* \cong \mathrm{Lip}_0(X)$.*

Proposition 4.2.5 ([Wea18])**.** *Let (X, d, x_0) be an infinite pointed metric space. Then $\mathcal{F}(X)$ is not a reflexive Banach space.*

The fact that $\mathcal{F}(X)$ is not reflexive for an infinite metric space implies that $\mathcal{F}(X)$ is in fact smaller than the entire dual space $\mathrm{Lip}_0(X)$. Our main result in this section relates the Lipschitz-free space back to persistence diagrams, providing an isometric embedding of $\mathcal{D}^+_{\mathrm{fin}}(X)$ into a Banach space.

Theorem 4.2.2. *Let $(X, d, *)$ be a pointed metric space. The inclusion $\mathcal{D}^+_{fin}(X) \hookrightarrow \mathcal{F}(X)$ is an isometric embedding.*

Proof. The space of finite persistence diagrams on $(X, d, *)$, $\mathcal{D}^+_{\mathrm{fin}}(X)$, isometrically embeds into $\mathcal{V}_{\mathrm{fin}}(X)$. Then, since $\mathcal{F}(X)$ is the completion of $\mathcal{V}_{\mathrm{fin}}(X)$, the inclusion $\mathcal{D}^+_{\mathrm{fin}}(X) \hookrightarrow \mathcal{F}(X)$ is an isometric embedding. □

Furthermore, the following proposition clarifies the fact that $\mathcal{F}(X)$ is also the completion of $\mathcal{M}_{\mathrm{fin},1}(X)$.

Proposition 4.2.6 ([OO19])**.** *Suppose $(X, d, *)$ is a Polish pointed metric space. The space $(\mathcal{V}_{fin}(X), \|\cdot\|_1)$ is dense in $(\mathcal{M}_{fin,1}(X), \|\cdot\|_1)$. Thus, $\mathcal{F}(X)$ can be regarded as the completion of $(\mathcal{M}_{fin,1}(X), \|\cdot\|_1)$.*

Therefore, the Lipschitz-free space $\mathcal{F}(X)$ subsumes all of the enrichments previously discussed, and provides the desired Banach space for persistence diagrams.

One of the main limitations of the Lipschitz-free space is the lack of an explicit representation of objects. Indeed, there are only a few explicit constructions of such spaces. It is well

known that the Lipschitz-free space of $\mathbb{R}$ is isomorphic to $L^1(\mathbb{R})$ [Wea18], and the Lipschitz-free space for convex subsets $C \subset \mathbb{R}^N$ is isometric to a quotient of $L^1(C, \mathbb{R}^N)$ [CKK17]. As we showed in Example 4.2.1, the Lipschitz-free space cannot be described as a space of signed measures in general.

However, while the limiting positive and negative measures in Example 4.2.1 each have infinite 1-norm, the 1-Wasserstein distance between them is

$$\|\mu^+ - \mu^-\|_1 = W_1^\partial[d](\mu^+, \mu^-) = \sum_{i=1}^{\infty} 2^{-i+1} = 2.$$

This suggests that the formal object $\mu^+ - \mu^-$, which we view as an element of the group completion $K(\mathcal{M}^+(X, A))$ of the monoid of all positive Radon measures, should have finite 1-norm given an appropriate extended metric on $K(\mathcal{M}(X, A))$. Indeed, let

$$\mathcal{K}_1(X, A) = \{\mu^+ - \mu^- \in K(\mathcal{M}(X, A)) \,:\, W_1^\partial[d](\mu^+, \mu^-) < \infty\}, \tag{4.2.9}$$

where we use the extended partial 1-Wasserstein distance from Definition 4.2.1. Following the setting of signed measures, we define a distance function on $\mathcal{K}_1(X, A)$ to be

$$W_1^\partial[d](\mu^+ - \mu^-, \nu^+ - \nu^-) := W_1^\partial[d](\mu^+ + \nu^-, \mu^- + \nu^+), \tag{4.2.10}$$

where on the right, we are once again using the extended partial 1-Wasserstein metric for positive measures.

Lemma 4.2.5. *The function $W_1^\partial[d]$ defined in Equation (4.2.10) is a metric for $\mathcal{K}_1(X, A)$.*

Proof. We begin by showing that $W_1^\partial[d]$ is well defined and finite on $\mathcal{K}_1(X, A)$. First, note that given measures $\mu, \nu, \zeta \in \mathcal{M}^+(X, A)$, we have

$$W_1^\partial[d](\mu + \zeta, \nu + \zeta) \leq W_1^\partial[d](\mu, \nu),$$

for the extended partial 1-Wasserstein metric on $\mathcal{M}^+(X, A)$. Indeed, any admissible coupling $\sigma \in \mathrm{ACpl}(\mu, \nu)$ induces an admissible coupling $\sigma + \Delta_*\zeta \in \mathrm{ACpl}(\mu + \zeta, \nu + \zeta)$, where $\Delta_*\zeta$ denotes the pushfoward of ζ along the diagonal map $\Delta : X \to X \times X$. Furthermore, we have $\mathrm{Cost}_1[d](\sigma) = \mathrm{Cost}_1[d](\sigma + \Delta_*\zeta)$, since the support of $\Delta_*\zeta$ lies on the diagonal and does not contribute to the cost, providing the desired inequality.

Suppose $\mu^+ - \mu^-, \nu^+ - \nu^- \in \mathcal{K}_1(X, A)$. Then, using the fact that $W_1^\partial[d]$ is an extended metric on $\mathcal{M}^+(X, A)$,

$$\begin{aligned} W_1^\partial[d](\mu^+ - \mu^-, \nu^+ - \nu^-) &= W_1^\partial[d](\mu^+ + \nu^-, \mu^- + \nu^+) \\ &\leq W_1^\partial[d](\mu^+ + \nu^-, \mu^- + \nu^-) + W_1^\partial[d](\mu^- + \nu^+, \mu^- + \nu^-) \\ &\leq W_1^\partial[d](\mu^+, \mu^-) + W_1^\partial[d](\nu^+, \nu^-) \\ &< \infty. \end{aligned}$$

Thus, $W_1^\partial[d]$ is finite on $\mathcal{K}_1(X, A)$. The fact that $W_1^\partial[d]$ is an extended metric on $\mathcal{M}^+(X, A)$ implies that $W_1^\partial[d]$ is a metric on $\mathcal{K}_1(X, A)$. □

This leads to the following question.

Question 4.2.1. *Is $(\mathcal{K}_1(X, A), W_1^\partial[d])$ isomorphic to the Lipschitz-free space $\mathcal{F}(\widetilde{X})$?*

An affirmative answer to Question 4.2.1 would provide an explicit representation for the Lipschitz-free space of a quotient metric space of an admissible metric pair, in terms of formal differences of Radon measures. This would provide a natural extension of the enriched spaces of persistence diagrams considered in this section. However, we will leave the consideration of this equivalence to future work. In the next section, we use the Lipschitz-free space to define the Banach space of bounded variation paths of persistence diagrams, as well as the path signature for these objects.

4.3 SIGNATURES FOR PATHS OF PERSISTENCE DIAGRAMS

Equipped with an isometric embedding of the space of finite persistence diagrams $\mathcal{D}_{\text{fin}}^+(X)$ into a Banach space $\mathcal{F}(X)$, we turn to the primary objects of interest in this paper: paths of persistence diagrams. Such paths naturally arise in topological data analysis by applying persistent homology to dynamic data sets. One notion of a time-varying persistence diagram is the *persistence vineyard*, introduced in [CEM06]. While persistence vineyards capture information about how individual persistent homology classes evolve over time, our methods treat persistence diagrams independently at each time step. Thus, we distinguish these two notions by referring to our objects of interest as *paths of persistence diagrams* or *time-varying persistence diagrams*, which have strictly less information that persistence vineyards.

Because we can treat paths of persistence diagrams as paths in the Banach space $\mathcal{F}(\widetilde{\Omega})$, all of the definitions from Section 1.2.3 hold in this setting. In particular, we apply Theorem 1.3.4 to obtain a universal and characteristic feature map for paths of persistence diagrams.

Theorem 4.3.1. *Let $\Lambda : T((\mathcal{F}(\widetilde{\Omega}))) \to T((\mathcal{F}(\widetilde{\Omega})))$ be a tensor normalization. The normalized signature $\Lambda \circ S : \widetilde{C}^{1-var}(\mathcal{F}(\widetilde{\Omega})) \to T((\mathcal{F}(\widetilde{\Omega})))$ is injective, universal with respect to $C_b(\widetilde{C}^{1-var}(V))$ with respect to the strict topology, and characteristic with respect to the space of finite Borel measures on $\widetilde{C}^{1-var}(V)$.*

However, there is an obstruction to using this signature as a feature map in practice: it is not readily computable. Indeed, the Banach space $\mathcal{F}(\widetilde{\Omega})$ is infinite dimensional, and thus even the first level of the signature will be infinite dimensional. Furthermore, there is no clear basis for $\mathcal{F}(\widetilde{\Omega})$ which can be effectively used for computation; in fact, this is the same problem as vectorizing persistence diagrams. We resolve this in the next section by approximating the signature of persistence diagrams by using a feature map for $\mathcal{F}(\widetilde{\Omega})$ as well as truncation of the signature.

While the signature from this section may not be directly amenable to computation, it forms a foundation for the investigation of stochastic time-varying systems using persistent homology. In particular, the results of this section hold in the much more general setting of *rough paths*, an approach to stochastic analysis which has led to influential developments in stochastic differential equations [FH20]. When studying time-varying systems in the presence of noise, the bounded variation hypothesis often does not hold. For example, Brownian motion has unbounded p variation for $p \leq 2$. Rough paths allow us to work with bounded p- variation paths for arbitrary $p \geq 1$ by enriching the path with information from higher iterated integrals. Thus, the theory of rough paths may allow us to describe the persistent homology of stochastic time-varying systems, which we leave to future work.

4.4 METHODS FOR APPLICATION TO DATA

The previous section provided the framework to theoretically study bounded variation paths of persistence diagrams by defining the signature of these paths. In this section, we focus on adapting these tools to provide computable methods for data analysis. Motivated by the observation that persistence diagrams can be viewed as measures, we introduce the *moment map*, a stable and injective feature map into a Hilbert space which computes all the moments of a persistence diagram. Here we make two modifications to the space of persistence diagrams:

1. We use birth/lifetime coordinates rather than birth/death coordinates. Lifetime coordinates provide a more interpretable feature, and simplify the description of partial Wasserstein metrics, as the distance to the boundary is simply the lifetime coordinate. Birth/lifetime coordinates have previously been used in the description of persistence images [Ada+17].

2. We use bounded persistence diagrams, supported in the space

$$\Omega_T := \{(x, y) \in \mathbb{R}^2; 0 \leq x, y \leq T\}. \tag{4.4.1}$$

In practical applications, this is a mild assumption since we often work with finite metric spaces built from data. The boundedness assumption allows us to provide stability results.

We then recall the discrete path signature, introduced in [KO19] and also studied in [DET20]. We discuss computational aspects of the path signature, and discuss its application to time-varying persistence diagrams.

4.4.1 *The Moment Map*

Equipped with the perspective that persistence diagrams are simply Radon measures on Ω_T, we study the use of moments as a feature map to vectorize persistence diagrams. The following is a restatement of a classical result from probability. Let $\mathbb{R}[[e_1, \ldots, e_N]]$ denote the power series algebra in N variables, and treat it as a Hilbert space, where the set of monomials form an orthonormal basis.

Proposition 4.4.1. *Suppose $K \subset \mathbb{R}^N$ is a bounded subset. We define the polynomial map*

$$\begin{aligned} P_0 : K &\to \mathbb{R}[[e_1, \ldots, e_N]] \\ x = (x_1, \ldots, x_N) &\mapsto \sum_{m=0}^{\infty} \sum_{|\alpha|=m} x^\alpha e^\alpha, \end{aligned}$$

where $\alpha = (\alpha_1, \ldots, \alpha_N) \in \mathbb{N}^N$ is an exponent vector, and $x^\alpha = x_1^{\alpha_1} \ldots x_N^{\alpha_N}$. The polynomial map is universal with respect to $\mathcal{G} = C(K)$, the space of continuous real-valued functions on K (equipped with the uniform topology), and characteristic with respect to $\mathcal{G}'$, which is the space of regular Borel measures on K. In particular, the expected exponential map $\mathbb{E}[P_0(\cdot)]$ is injective on the space of finite Borel measures on K.

Proof. First, we note that the elements of the dual, $\langle \ell, P_0(\cdot) \rangle : K \to \mathbb{R}$, are exactly the polynomial functions on K. Then, by the classical Stone-Weierstrass theorem, P_0 is universal with respect to $\mathcal{G}$. By Theorem 1.3.2, the map P_0 is characteristic with respect to $\mathcal{G}'$. □

In our case, we wish to work with the quotient of the metric pair $(\Omega_T, d, \partial\Omega_T)$, where we define

$$\partial\Omega_T := \{(x, 0) : x \in [0, T]\}.$$

We will define $(\widetilde{\Omega}_T, \tilde{d}, *)$ to be the quotient of $(\Omega_T, d, \partial\Omega_T)$, where $*$ is the point representing the quotient of $\partial\Omega_T$. We will also require a modification of the polynomial map such that it will be continuous on this quotient. In particular, let $\mathcal{R} \subset \mathbb{R}[[e_1, e_2]]$ be the power series subalgebra in 2 variables such that the coefficients of e_1^k are 0 for all $k \geq 1$. Note that $\mathcal{R}$ inherits the Hilbert space structure from $\mathbb{R}[[e_1, e_2]]$. Then, we define the modified polynomial map $P : \widetilde{\Omega}_T \to \mathcal{R}$ to be

$$P(x) = 1 + \sum_{m=1}^{\infty} \sum_{\substack{|\alpha|=m \\ \alpha_2 \geq 1}} \sqrt{\frac{1}{\alpha_1!\alpha_2!}} x^\alpha e^\alpha, \tag{4.4.2}$$

for $x \neq *$ and $P(*) = 1$. The coefficients should be thought of as normalization, which is required to show that the map is Lipschitz, and also results in a simple kernel computation. We can use a similar proof to show the universal and characteristic properties of this modified map.

Proposition 4.4.2. *The modified polynomial map $P : \widetilde{\Omega}_T \to \mathcal{R}$ is universal with respect to $\mathcal{G} = C_0(\widetilde{\Omega}_T)$, the continuous functions which vanish at infinity, and characteristic with respect to finite Borel measures.*

Proof. Here the elements of the dual, $\langle \ell, P(\cdot) \rangle : \widetilde{\Omega}_T \to \mathbb{R}$, where $\ell \in \mathcal{R}'$, are the polynomial functions on Ω_T° with no e_1^k monomials. Note that $\widetilde{\mathbb{R}}[[e_1, e_2]]$ is still an algebra; in fact it is a subalgebra of $C_0(\widetilde{\Omega}_T)$. Furthermore, this collection of polynomial functions vanishes nowhere on $\widetilde{\Omega}_T$ and the space $\widetilde{\Omega}_T$ is locally compact. Thus, the locally compact Stone-Weierstrass theorem holds, and thus P is universal. Then, by Theorem 1.3.2, it is characteristic with respect to $\mathcal{G}'$, which contains the space of finite Borel measures. □

Proposition 4.4.3. *There exists an extension $\widetilde{P} : \mathcal{F}(\widetilde{\Omega}_T) \to \mathcal{R}$ of P to the Lipschitz-free space. Furthermore, $\widetilde{P}$ is Lipschitz continuous.*

Proof. From the universal property of Lipschitz-free spaces in Theorem 4.2.1, it suffices to show that P is Lipschitz continuous. Let $x, y \in \widetilde{\Omega}_T$. First, consider the monomial map $m_{a,b} : \Omega_T \to \mathcal{R}$ to be $m_{a,b}(x) = x_1^a x_2^b$ and $m_{a,b}(*) = 0$. We assume that $b \geq 1$ so that $m_{a,b}(\partial\Omega_T) = 0$, so we can define the map on the quotient $\tilde{m}_{a,b} : \widetilde{\Omega}_T \to \mathcal{R}$. First, suppose that $\tilde{d}(x, y) = d(x, y)$. Then, we have

$$|\tilde{m}_{a,b}(x) - \tilde{m}_{a,b}(y)| \leq (a + b)T^{a+b} d(x, y).$$

Next, suppose $\tilde{d}(x,y) = d(x,\partial\Omega_T) + d(y,\partial\Omega_T)$. Let x', y' be the projection of x, y to $\partial\Omega_T$ (the x-axis). Then, we have

$$\begin{aligned}|\tilde{m}_{a,b}(x) - \tilde{m}_{a,b}(y)| &\leq |\tilde{m}_{a,b}(x) - \tilde{m}_{a,b}(*)| + |\tilde{m}_{a,b}(*) - \tilde{m}_{a,b}(y)| \\ &= (a+b)T^{a+b}d(x,x') + (a+b)T^{a+b}d(y,y') \\ &= (a+b)T^{a+b}\big(d(x,\partial\Omega_T) + d(y,\partial\Omega_T)\big).\end{aligned}$$

Therefore, $\tilde{m}_{a,b}$ is Lipschitz continuous with Lipschitz constant $(a+b)T^{a+b}$.

Now, considering the map P, we have

$$\begin{aligned}\|P(x) - P(y)\|^2 &\leq \sum_{m=1}^{\infty} \sum_{\substack{a+b=m,\\ b\geq 1}} \frac{1}{a!b!}|\tilde{m}_{a,b}(x) - \tilde{m}_{a,b}(y)|^2 \\ &\leq \sum_{m=1}^{\infty} \sum_{\substack{a+b=m,\\ b\geq 1}} \frac{1}{m!}\binom{m}{a} m^2 T^{2m}\tilde{d}(x,y)^2 \\ &\leq \sum_{m=1}^{\infty} \frac{m^2(2T^2)^m}{m!}\tilde{d}(x,y)^2,\end{aligned}$$

In order to bound the sum, we note that $m^2 \leq 3^m$ for all $m \geq 1$. Thus, we have

$$C = \left(\sum_{m=1}^{\infty} \frac{m^2(2T^2)^m}{m!}\right)^{1/2} \leq \left(\sum_{m=1}^{\infty} \frac{(6T^2)^m}{m!}\right)^{1/2} \leq \exp(3T^2).$$

Therefore, the map P is Lipschitz continuous with the Lipschitz constant C as defined above. □

Remark 4.4.1. Note that we also lower bound the Lipschitz constant in the above proposition as follows.

$$C = \left(\sum_{m=1}^{\infty} \frac{m^2(2T^2)^m}{m!}\right)^{1/2} \geq \left(\sum_{m=1}^{\infty} \frac{(2T^2)^m}{m!}\right)^{1/2} = \exp(T^2) - 1$$

Thus, the Lipschitz constant is bounded by

$$\exp(T^2) - 1 \leq C \leq \exp(3T^2).$$

Combining the previous two results, we we find that the restriction of the moment map to $\mathcal{M}_{\text{fin}}(\widetilde{\Omega}_T)$ is a stable and injective feature map for static bounded persistence diagrams, which is sufficient for applications to data. We note that while $\widetilde{P}$ is stable on

the Lipschitz-free space $\mathcal{F}(\widetilde{\Omega}_T)$ by the previous proposition, it is not immediate that the extension to the Lipschitz-free space is still injective.

Theorem 4.4.1. *The restriction $\widetilde{P}_{\mathcal{M}} := \widetilde{P}|_{\mathcal{M}_{fin}(\widetilde{\Omega}_T)} : \mathcal{M}_{fin}(\widetilde{\Omega}_T) \to \mathcal{R}$ is Lipschitz continuous and injective.*

Proof. By the definition of the extension in the universal property, the restriction of $\widetilde{P}$ to $\mathcal{M}_{\text{fin}}(\widetilde{\Omega}_T)$ is equivalent to the expected polynomial map,

$$\widetilde{P}_{\mathcal{M}} = \mathbb{E}[P].$$

The result follows from Proposition 4.4.2 and Proposition 4.4.3. □

The continuity and injectivity of $\widetilde{P}_{\mathcal{M}}$ allows us to treat static persistence diagrams as elements of the Hilbert space $\mathcal{R}$. Then by Theorem 1.3.4, we have a universal and characteristic feature map for paths of persistence diagrams, viewed as elements of $\widetilde{C}^{1-var}(\mathcal{R})$.

Corollary 4.4.1 (to Theorem 1.3.4)**.** *Let $\Lambda : T((\mathcal{R})) \to T((\mathcal{R}))$ be a tensor normalization. The normalized signature $\Lambda \circ S : \widetilde{C}^{1-var}(\mathcal{R}) \to T((\mathcal{R}))$ is injective, universal with respect to bounded continuous functions $C_b(\widetilde{C}^{1-var}(\mathcal{R}))$ with respect to the strict topology, and characteristic with respect to the space of finite Borel measures on $\widetilde{C}^{1-var}(\mathcal{R})$.*

Furthermore, the composition of the truncated path signature with the moment map is Lipschitz continuous for paths with a fixed bounded length.

Theorem 4.4.2. *Suppose $L > 0$ and suppose*

$$C_L^{1-var}(V) := \{\gamma \in C^{1-var}(V) \,:\, |\gamma|_{1-var} \leq L\}.$$

Then, the composition $S_{\leq M} \circ \widetilde{P}_{\mathcal{M}} : C_L^{1-var}(\mathcal{M}_{fin}(\widetilde{\Omega}_T)) \to T((\mathcal{R}))$ is Lipschitz continuous.

Proof. By Theorem 4.4.1, $\widetilde{P}_{\mathcal{M}}$ is Lipschitz continuous, and let C be its Lipschitz constant. Then, given two paths $\alpha, \beta \in C^{1-var}(\mathcal{M}_{fin}(\widetilde{\Omega}_T))$, we have

$$\begin{aligned}
|\widetilde{P}_{\mathcal{M}}(\alpha - \beta)|_{1-var} &= \sup_{\Pi} \sum_{i=1}^{n} \|\widetilde{P}_{\mathcal{M}}(\alpha(t_i) - \beta(t_i)) - \widetilde{P}_{\mathcal{M}}(\alpha(t_{i-1}) - \beta(t_{i-1})\| \\
&\leq \sup_{\Pi} \sum_{i=1}^{n} C\|(\alpha(t_i) - \beta(t_i)) - (\alpha(t_{i-1}) - \beta(t_{i-1})\| \\
&\leq C|\alpha - \beta|_{1-var}.
\end{aligned}$$

Then, by the stability result for the truncated path signature in Corollary 1.2.3, we have

$$\begin{aligned}\|S_{\leq M} \circ \widetilde{P}_{\mathcal{M}}(\alpha) - S_{\leq M} \circ \widetilde{P}_{\mathcal{M}}(\beta)\| &\leq \exp(1) \left(\frac{1-L^M}{1-L}\right)^{1/2} |\widetilde{P}_{\mathcal{M}}(\alpha-\beta)|_{1-var} \\ &\leq C \exp(1) \left(\frac{1-L^M}{1-L}\right)^{1/2} |\alpha-\beta|_{1-var}.\end{aligned}$$

□

Now that theoretical properties have been established, we turn to computational concerns. In computational settings, we mainly need to work with persistence diagrams with finitely many points. Given two persistence diagrams $X, Y \in \mathcal{D}^+_{\text{fin}}(\widetilde{\Omega}_T)$ with points

$$X = \{\mathbf{x}_i = (x_{i,1}, x_{i,2})\}_{i=1}^{N_X}, \quad Y = \{\mathbf{y}_i = (y_{i,1}, y_{i,2})\}_{i=1}^{N_Y},$$

we provide an exact formula for the computation of

$$\kappa(X,Y) := \langle \widetilde{P}_{\mathcal{M}}(X), \widetilde{P}_{\mathcal{M}}(Y)\rangle_{\mathcal{R}}.$$

Proposition 4.4.4. *Given two persistence diagrams $X, Y \in \mathcal{D}^+_{fin}(\widetilde{\Omega}_T)$ as defined above, the moment kernel can be computed as*

$$\kappa(X,Y) = \sum_{i=1}^{N_X}\sum_{j=1}^{N_Y} \exp \mathbf{x}_i \cdot \mathbf{y}_j - \exp x_{i,1} \cdot y_{j,1} + 1. \tag{4.4.3}$$

Proof. As persistence diagrams are simply sums of Dirac measures, the kernel can be written as a sum of inner products of the polynomial map

$$\kappa(X,Y) = \sum_{i=1}^{N_X}\sum_{j=1}^{N_Y} \langle P(\mathbf{x}_i), P(\mathbf{y}_j)\rangle_{\mathcal{R}}.$$

Expanding the definition of the monomial map, we find

$$\begin{aligned}\langle P(\mathbf{x}_i), P(\mathbf{y}_j)\rangle &= 1 + \sum_{p=1}^{\infty} \sum_{\substack{a+b=p \\ a\geq 0, b\geq 1}} \binom{p}{a} \frac{(x_{i,1}y_{j,1})^a (x_{i,2}y_{j,2})^b}{p!} \\ &= \exp \mathbf{x}_i \cdot \mathbf{y}_j - \exp x_{i,1} \cdot y_{j,1} + 1\end{aligned}$$

□

In some cases, it may be more required to use a truncation of the moment map, denoted

$$\widetilde{P}_{\mathcal{M},n} : \mathcal{M}_{\text{fin}}(\widetilde{\Omega}_T) \to \mathcal{R}_n,$$

where $\mathcal{R}_n$ is the restriction of $\mathcal{R}$ to the set of polynomials of degree at most n. In these situations, we show that the truncation error can be bounded simply using the Taylor series remainder.

Lemma 4.4.1. *Given* $X = \{\mathbf{x}_i = (x_{i,1}, x_{i,2})\}_{i=1}^{N_X} \in \mathcal{D}_{fin}^{+}(\widetilde{\Omega})$*, the truncation error of* $\widetilde{P}_{\mathcal{M},n}$ *is bounded by*

$$\|\widetilde{P}_{\mathcal{M}}(X) - \widetilde{P}_{\mathcal{M},n}(X)\| \leq \sum_{i=1}^{N_X} \frac{e^{(x_{i,1}+x_{i,2})}(x_{i,1}+x_{i,2})^{n+1}}{(n+1)!}.$$

4.4.2 *Signature Kernels for Paths of Persistence Diagrams*

Combining the results in this section with the discrete path signature from Section 1.3.5, we obtain a feature map for discrete paths of persistence diagrams into the Hilbert space $T^{(m)}(\mathcal{R})$

$$\hat{S}_m \circ \widetilde{P}_{\mathcal{M}} : \text{Seq}(\mathcal{M}_{\text{fin}}(\widetilde{\Omega}_T)) \to T^{(m)}(\mathcal{R}).$$

From the previous section, there are two ways to compute the corresponding signature kernel:

1. Use the truncated form of the moment map $\widetilde{P}_{\mathcal{M},n}$, and explicitly compute the features in $\mathcal{R}_n$. Then, compute the expression from Proposition 1.3.3. The advantage of this method is the fact that we only need to compute the features for each time series once, although this method will result in some truncation error.

2. Use the kernel for the moment map defined in Proposition 4.4.4, and then compute the expression from Proposition 1.3.4. With this method, we are able to use the entire set of moment features, but we must perform the moment kernel computation for every pair of paths, which can be prohibitively expensive.

In practice, the first method is often much more efficient. Indeed, suppose we wish to compute the kernel matrix for a collection of N_{samples} elements of $\text{Seq}(\mathcal{D}_{\text{fin}}(\widetilde{\Omega}_T))$, each of length L. Using the first method, we must compute the moment map $O(N_{\text{samples}}L)$ times. However, in the second method, for each pair of paths, we must compute the moment kernel matrix, which requires $O(L^2)$ computations of the moment kernel. Furthermore, this is repeated for every pair of paths, and thus we perform the moment kernel computation $O(N_{\text{samples}}^2 L^2)$ times.

This must be balanced with memory considerations. On the one hand, the memory required in the first method to store the paths in feature space scales linearly with the dimension of the feature space. On the other hand, the second only requires dynamic memory, requiring $O(L^2)$ for each computation.

While we have primarily discussed the moment map as the feature map for persistence diagrams, the signature method can be used together with any feature map such as persistence landscapes [Bub15], persistence images [Ada+17], or tropical coordinates [Kal19]. By using the second method outlined above, we can also adapt the use of kernels such as the sliced-Wasserstein kernel [CCO17] or the persistence scale space kernel [Rei+15; Kwi+15].

Furthermore, there is a third approach to computing signature kernels which is discussed in [Cas+21]. In this approach, the untruncated signature kernel is formulated in terms of the solution to a PDE. This approach can be also be extended to time series composed with a feature map $\Phi : \mathcal{X} \to V$ by using explicit computations for its corresponding kernel $\kappa : \mathcal{X} \times \mathcal{X} \to V$. In particular, this method would allow us to compute the *untruncated* signature kernel composed with the *untruncated* moment map. In this case, the approximation error would be purely due to the PDE solution error rather than a truncation error.

4.5 APPLICATION: PARAMETER ESTIMATION IN A SWARM MODEL

In this section, we study the paths of persistence diagrams which arise in models of collective motion as described in the introduction. We perform simulations using a swarm model and use kernel support vector regression to estimate model parameters. In particular, all of our methods only use unidentified position data from the swarms: we have access to position data for a collection of agents at each time step, but we do not have information regarding how individual particles move over time. This restriction disqualifies the use of classical order parameters which make use of velocity information, such as group polarization and (absolute) angular momentum, to classify collective behavior. Thus, we estimate the model parameters purely by using the evolution of the *shape* of the swarm over time.

We study the 3D D'Orsogna model [D'O+06] for collective motion. The evolving shape of a single swarm simulation is seen in Figure 4.4. The swarm is initialized with random positions and velocities, and the disordered nature of collective motion can be seen at $t = 5$. After a period of time, the swarm moves into a more organized structure roughly in the shape of a torus at $t = 150$. Finally, given enough time, the swarm settles into steady-state behavior in the shape of a sphere at $t = 350$. The exact organization of the swarm over time will depend on the (C, ℓ) parameters, which we have chosen to be $(0.6, 0.3)$ in the figure. The stable structures of these swarms are studied and organized into a phase diagram in [NJG12].

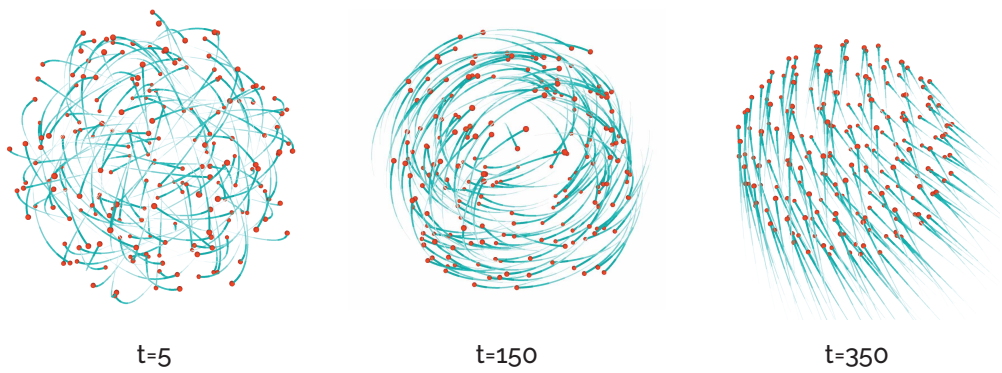


Figure 4.4: A single swarm simulation shown at time points $t = 5$, $t = 150$, and $t = 350$. This is simulated using the 3D D'Orsogna model [D'O+06] with $N_{\text{agent}} = 200$ agents. The model parameters are $m = 1$, $\alpha = 1.0$, $\beta = 0.5$, $C = 0.6$, and $\ell = 0.3$.

Previous studies [Bha+19; Xia+20] have successfully applied persistent homology to classify swarms based on the 2D D'Orsogna and 2D Vicsek models, using a feature map called crocker plots. Both of these studies use a collection of simulations generated using a fixed finite set of model parameters, and use clustering methods such as k-medoids to classify the simulations into clusters based on the fixed set of parameters. In contrast, our experiments use the 3D D'Orsogna model, and our model parameters are chosen uniformly at random within a specified region in parameter space. Rather than using classification algorithms which only work for a finite number of model parameters, we use regression to perform parameter estimation.

Furthermore, we study the generalizability and stability of our methods by considering situations with missing data. We find that our methods remain stable when a fixed number of agents are randomly subsampled at each time step. Moreover, we consider a heterogeneous situation where the training data is incommensurate with the test data: we train a regressor on the full simulation, and use the regressor to estimate parameters for simulations with subsampled data. While crocker plots become unstable for these experiments, we find that signature methods are able to effectively predict parameters in most situations.

4.5.1 *Model Description and Data*

The D'Orsogna model [D'O+06] describes the collective motion of N_{agent} interacting agents with positions $\mathbf{x}_i \in \mathbb{R}^d$, and velocities $\mathbf{v}_i \in \mathbb{R}^d$. Previous work [TZH15; Bha+19] has applied persistent homology techniques to study the two-dimensional ($d = 2$) model [Bha+19].

Here we choose to focus on the three-dimensional ($d = 3$) case. The agents are governed by the following coupled differential equations

$$\begin{aligned}
\frac{d\mathbf{x}_i}{dt} &= \mathbf{v}_i \\
m\frac{d\mathbf{v}_i}{dt} &= \left(\alpha - \beta|\mathbf{v}_i|^2\right)\mathbf{v}_i - \nabla_i U(\mathbf{x}_i), \\
U(\mathbf{x}_i) &= \sum_{j\neq i}^{N}\left(C_r e^{-\mathbf{r}_{i,j}/\ell_r} - C_a e^{-\mathbf{r}_{i,j}/\ell_a}\right),
\end{aligned}$$

where $\mathbf{r}_{i,j} = \|\mathbf{x}_i - \mathbf{x}_j\|$ and ∇_i refers to the gradient with respect to $\mathbf{x}_i$. Each agent has mass m, is self-propelled with propulsion strength α, and experiences a velocity-dependent drag with coefficient β. Furthermore, each particle $\mathbf{x}_i$ experiences a potential $U(\mathbf{x}_i)$ derived from pairwise interactions with all of the other agents. Each pair has a *repulsive* component with strength $C_r > 0$ and characteristic length scale $\ell_r > 0$, and an *attractive* component with strength $C_a > 0$ and characteristic length scale $\ell_a > 0$.

For our simulations, we use $N_{\text{agent}} = 200$ interacting agents. Furthermore, we fix $m = 1$, $\alpha = 1.0$, and $\beta = 0.5$ which corresponds to the parameters chosen in [NJG12]. After nondimensionalizing the equation, we are left with two free parameters: the interaction strength $C = C_r/C_a$ and the characteristic length $\ell = \ell_r/\ell_a$.

We perform $N_{\text{simulation}} = 500$ instances of this simulation, where each instance is run for $t \in [0, 400]$, and is discretized using $T = 400$ uniformly spaced time points. The differential equations are solved using the DifferentialEquations package in Julia, which uses adaptive timestepping. The solution is then sampled at the desired time points using the built-in interpolation function. For each simulation, the initial positions are set by uniformly sampling the unit cube $[0, 1]^3$, and each component of the initial velocities are sampled from a Gaussian distribution with mean 1 and variance 1, as is done in [NJG12], and C and ℓ are chosen uniformly at random from $[0.1, 2]$. As they are easily identified with simpler measures, we reject unbounded phenotypes, which we define to be a simulation in which any particle moves 40 units in any direction before $t = 200$. This restricts the distribution of (C, ℓ), and we show the distribution of the parameters in our simulations in Figure 4.5.

To consider missing data, we take a random subsample of N agents at each time step. We consider three cases of this with a fixed number of subsamples N across all simulations with $N = 150, 100, 50$. We also consider the case where N is fixed across time for individual simulations, but we sample N uniformly from $[50, 200]$ for each simulation. Finally, we consider situations with heterogeneous training/test data. For these experiments we train the regressor with data from the full simulation, and predict parameters for data from the subsampled simulations. These are called the *heterogeneous* experiments, and this is performed using all four sets of subsampled data.

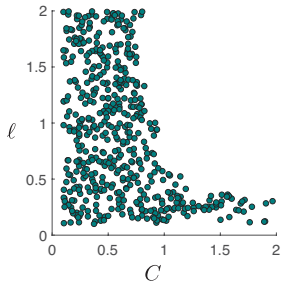


Figure 4.5: Scatter plot of generated parameter pairs (C, ℓ) that resulted in bounded swarm simulations.

1. Full simulation
2. Subsampled (Fixed $N = 150$)
3. Subsampled (Fixed $N = 100$)
4. Subsampled (Fixed $N = 50$)
5. Subsampled (Random N in $[50, 200]$)
6. Heterogeneous Subsampled ($N = 150$)
7. Heterogeneous Subsampled ($N = 100$)
8. Heterogeneous Subsampled ($N = 50$)
9. Heterogeneous Subsampled (Random N in $[50, 200]$)

4.5.2 *Support Vector Regression and Feature Maps*

For each of full and subsampled data sets, we compute persistent homology up to dimension $d = 2$ for all simulations and all time steps in order to obtain three paths of persistence diagrams (at dimensions $d = 0, 1, 2$) for each simulation. Persistent homology is computed using the Eirene software package [HG17] written in Julia.

We use the support vector regression (SVR) algorithm [Dru+97; SS02] to take advantage of signature kernels. Suppose we have a training set $\{(\mathbf{D}_i, y_i)\}_{i=1}^{N_{\text{train}}}$, where

$$\mathbf{D}_i(t) = (D_i^0(t), D_i^1(t), D_i^2(t)) \in C^{1-var}\left(\mathcal{D}_{\text{fin}}^+(\widetilde{\Omega})^3\right)$$

represents paths of persistence diagrams with $d = 0, 1, 2$, and $y_i \in \mathbb{R}$ is the true parameter value. Given a feature map

$$\Phi : C^{1-var}\left(\mathcal{D}_{\text{fin}}^+(\widetilde{\Omega})^3\right) \to \mathcal{H},$$

support vector regression solves for a regression function

$$f_{\mathbf{w},b}(\mathbf{x}) = \langle \mathbf{w}, \Phi(\mathbf{x}) \rangle_{\mathcal{H}} - b,$$

where $\mathbf{w} \in \mathcal{H}$ and $b \in \mathbb{R}$ by minimizing the regularized cost function

$$C(\mathbf{w}, b) = \frac{1}{2}\|\mathbf{w}\|_{\mathcal{H}}^2 + \frac{\lambda}{N_{\text{train}}} \sum_{i=1}^{N_{\text{train}}} |y_i - f_{\mathbf{w},b}(\mathbf{x})|_\epsilon, \tag{4.5.1}$$

where

$$|y|_\epsilon = \max\{0, |y| - \epsilon\},$$

is the *ϵ-insensitive loss function* and $\lambda, \epsilon > 0$ are hyperparameters. We use the SVR algorithm from the Julia implementation of the scikit-learn library. We perform our analyses using three feature maps.

1. **Moments + Signature.** This is the composition of the truncated moment map with the discrete signature kernel described in Section 4.4.2. In order to take into account all three dimensions for homology, we compute moments for all three persistence diagrams, and compute signatures with respect to the concatenated feature vector. We truncate the moments at degree $n = 6$ and the signature at level $m = 8$.

2. **PersPath + Signature.** Here, we use a signature-based feature map for persistence diagrams introduced in [CNO20] using the *Betti embedding*. Given a static persistence diagrams at $d = 0, 1, 2$, the Betti embedding returns a path $\beta = (\beta_0, \beta_1, \beta_2)$ in $\mathbb{R}^3$, also called the *Betti curve*, parameterized by the scale parameter ϵ in persistent homology. At each ϵ, the Betti curve provides the rank of the $d = 0, 1$, and 2 persistent homology. Given static persistence diagrams $\mathbf{D}_i = (D_i^0, D_i^1, D_i^2) \in \left(\mathcal{D}_{\text{fin}}^+(\widetilde{\Omega})\right)^3$ in dimensions $d = 0, 1, 2$, where $D_i^d = \{(x_i^d, y_i^d)\}_{i=1}^{N_d}$ in birth/lifetime coordinates, the Betti curve is defined by

   $$\beta_d(\epsilon) = |\{i \in \{1, \ldots, N_d\} \,:\, x_i^d < \epsilon < x_i^d + y_i^d\}|. \tag{4.5.2}$$

   By taking the truncated signature of the Betti curves along the ϵ parameter, we obtain a path in $T^{(n)}(\mathbb{R}^3)$. Then, we compute the signature a second time along the t parameter. We truncate the first signature computation at level $n = 6$ and the second signature computation at level $m = 8$.

3. **Crocker Plots.** Crocker plots, introduced in [TZH15], treat the discretized Betti curve $\beta = (\beta_0, \beta_1, \beta_2)$ as a feature vector. Following [Bha+19], we discretize the Betti curve on a log scale from $\epsilon = 10^{-4}$ to $\epsilon = 1$ with $N_\epsilon = 200$ points. Furthermore, we uniformly subsample the time coordinate at intervals of $\Delta t = 20$ to obtain $N_t = 20$ points in the time coordinate. Collapsing the discretized Betti curves into a single vector, each path of persistence diagrams coorresponds to an element in

$\mathbb{R}^{3\times 200\times 20} = \mathbb{R}^{12000}$. The kernel is obtained from the standard inner product on $\mathbb{R}^{12000}$.

For the first two signature methods, we also show results using a sliding window embedding, a common preprocessing step for signatures [CK16; Fer21]. The sliding window embedding can be defined as

$$\begin{aligned}\mathrm{SWin}_k : C^{1-var}(\mathcal{H}) &\to C^{1-var}(\mathcal{H}^{k+1}) \\ \gamma(t) &\mapsto \big(\gamma(t), \gamma(t-\tau), \dots, \gamma(t-k\tau)\big)\end{aligned}$$

for some delay $\tau > 0$. We define $\gamma(t) = 0$ for any $t < 0$. We consider the case of $k = 0, 1, 2$. Note that for the PersPath + Signature method, we only consider lags in the second signature computation due to large memory requirements. We note that these memory issues can theoretically be circumvented by iterating the PDE approach to signature computations [Cas+21], but we do not consider this approach here.

4.5.3 *Statistical Methods*

We detail our statistical methodology in this section. We begin by enumerating the $N_{\text{simulation}} = 500$ individual simulations and randomly select a 400 / 100 split for the training and test data respectively. With this choice, we perform a *trial* of our analysis, which consists of three steps:

1. **Hyperparameter selection.** The SVR algorithm uses the hyperparameters (λ, ϵ) and we tune the hyperparameters by grid search with 4-fold cross-validation exclusively using the training data. The grid points of λ are logarithmically spaced between 10^{-3} and 10^{3}, and those of ϵ are logarithmically spaced between 10^{-5} and 10^{1}. We select 13 points for each parameter for a total of 169 parameter pairs.

2. **Training.** We select the parameter pair (λ, ϵ) which has the best performance in terms of mean square error (MSE). We then train the SVR using these hyperparameters on the entire training set.

3. **Prediction.** Using the trained regressor, we predict the (C, ℓ) values of the data in the test set and compute the MSE for each parameter individually.

Note that for the heterogeneous experiments, we use the full simulation data for steps 1 and 2, and use the subsampled data for step 3. We then repeat the trial $N_{\text{trial}} = 1000$ times in order to obtain a distribution of the MSE for a given data set / feature map.

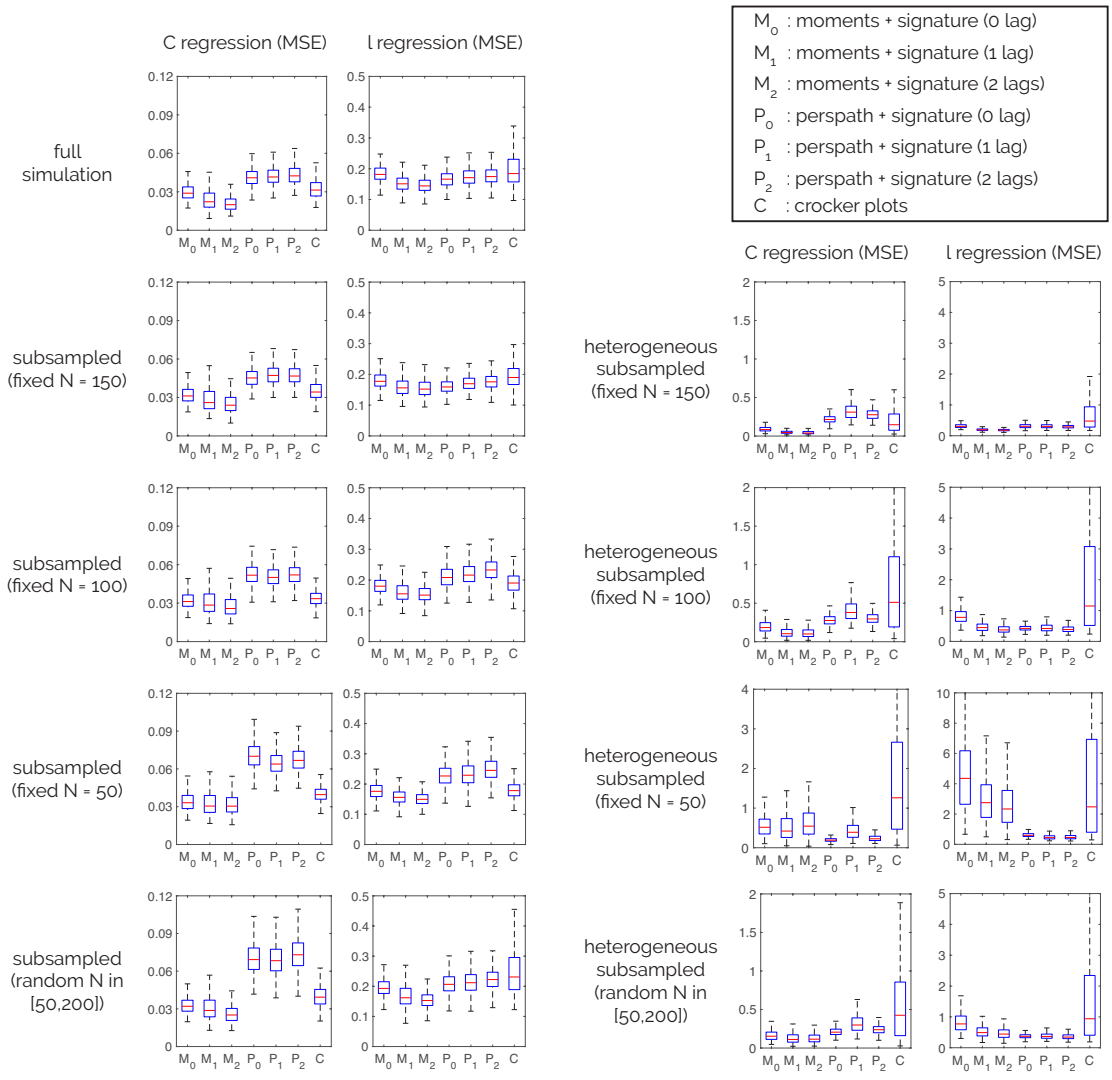


Figure 4.6: Box plots for the 9 variants of the data, and all 7 feature maps. The central red line of each plot denotes the median and the bottom / top edge of the box corresponds to the 25^{th} and 75^{th} percentile respectively.

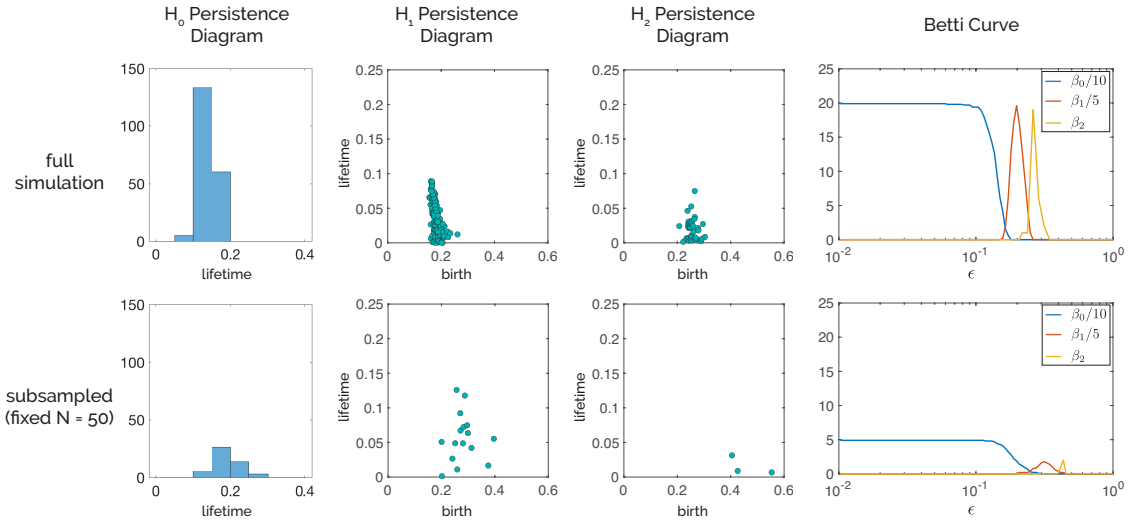


Figure 4.7: Comparison of persistence diagrams and Betti curves between a point cloud from the full simulation (top row) and the subsampled simulation with fixed $N = 50$ (bottom row). The H_0 persistence diagram is shown as a histogram of lifetimes, since all birth times are at $\epsilon = 0$. In the Betti curve, β_0 is scaled by a factor of $1/10$ and β_1 is scaled by a factor of $1/5$ for visualization purposes.

4.5.4 *Results and Discussion*

The results of our experiment are shown as boxplots in Figure 4.6. In the 5 experiments with the same train / test data used (on the left side of Figure 4.6), all of the feature maps are able to predict parameters to some degree. For both the C and ℓ regression, we can see that the Moments + Signatures (2 lags) feature map (M_2) consistently outperforms all other feature maps, though the difference is more prominent for the C regression.

In the 4 experiments with heterogeneous train / test sets (on the right side of Figure 4.6), we observe more unstable behavior. Note that the y axes on these plots are significantly larger; in particular, the axes on the fixed $N = 50$ plot is larger by a factor of 2. We can immediately see that all feature maps perform worse when compared with the experiments using the same train / test data. However, the decrease in performance for the crocker plots (C) is the most striking. The Moments + Signatures (2 lags) (M_2) significantly outperforms the crocker plot in all cases, but when we only use $N = 50$ subsampled data as the test set, its performance also reduces drastically. This is expected due to the disparity between the training and test sets: we are training on data using 200 agents at each time point, and predicting parameters using the locations of only 25% of the data, which dramatically alters the density of the cloud. While the PersPath + Signature method does not provide

the best performance for most experiments, it significantly outperforms all other feature maps for the heterogeneous subsampled (fixed $N = 50$) experiment.

To understand the performance of the feature maps in this $N = 50$ case, we consider in Figure 4.7 a comparison of the persistence diagrams between the full simulation and the subsampled simulation for a single simulation at a single time point. This comparison demonstrates two representative phenomena that occur due to subsampling:

1. the persistence diagrams are much sparser due to fewer points in the point cloud, and
2. there is a shift in the birth coordinate.

The sparsity of the subsampled persistence diagrams leads to a scaling of all feature maps. However, the shift in the birth coordinate results in nontrivial differences for the moment map and crocker plots; for instance, the cross-moments including the birth coordinate will shift. However due to reparametrization invariance, this will not result in a significant change in the signature of the Betti curve. Thus, the persistence paths + signature feature map remains considerably more stable under subsampling noise.

One possible amendment to the moment map that would take into account the subsampling phenomena is to use the number of agents in the swarm, which is a known parameter, to normalize and scale the point cloud and persistence diagrams. These normalizations are motivated by limit theorems for point processes in a unit d-cube in [DP19; DL21]. In particular, let $\mathbb{X}_N = (X_1, \ldots, X_N)$ be a collection of N i.i.d. samples of a bounded, positive distribution on $[0,1]^d$. Let $\mu_N^q = N^{-1}\mathrm{Dgm}_q(N^{1/d}\mathbb{X}_N)$, where $\mathrm{Dgm}_q(N^{1/d}\mathbb{X}_N)$ is the dimension q persistence diagram of the point cloud $\mathbb{X}$ scaled in all coordinates by $n^{1/d}$, and the resulting diagram is then scaled by N^{-1}, viewed as a measure. Then, the sequence of measures $\mu_N^q \xrightarrow{N\to\infty} \mu^q$ converges vaguely to some Radon measure μ^q. This result suggests that by scaling our point clouds by $N^{1/d}$, and the resulting persistence diagrams by N^{-1}, we can effectively normalize our feature maps with respect to subsampling noise. We do not pursue such an analysis here in the interest of concision.

4.6 CONCLUSION

We have presented a theoretical and computational framework for the study of paths of persistence diagrams. By introducing the concept of Lipschitz-free spaces, we were able to integrate several generalizations of the space of persistence diagrams into a single Banach space. This allowed us to define a Banach space of bounded variation paths of persistence diagrams, and thus to define the path signature of these diagrams. The Banach space structure provides a setting to study probabilistic and statistical properties of both static and time-varying persistence diagrams. Furthermore, the path signature is universal

and characteristic, providing a robust theoretical tool kit for characterizing probability measures on paths of persistence diagrams.

In addition, we introduced a new feature map for static persistence diagrams, motivated by the perspective of diagrams as measures, and showed that it is both injective and stable. Coupling this with the discrete path signature, we defined an approximation to the intrinsic continuous signature. This approximation is computable, stable, and valued in a Hilbert space, enabling the application of kernel methods for data analysis. We demonstrated the efficacy of our feature map with a parameter estimation problem for a 3D model of collective motion.

While we specifically considered the moment map, the signature methods we presented is general: the signature can be used in conjunction with any feature map or kernel for persistence diagrams. Moreover, the framework laid out in this paper suggests several possibilities for future work, and we highlight some promising directions here.

- **Persistence vineyards.** Persistence vineyards [CEM06] are paths of persistence diagrams with finer information; in particular it retains information about how individual homology classes evolve over time. Can we further extend the connections between topological data analysis, optimal transport and path signatures to represent and study persistence vineyards? In particular, can we use adapted Wasserstein metrics [Bac+20] and higher rank signatures [BLO20] to characterize vineyards?

- **Rough paths.** The theory of *rough paths* [LCL07] provides a deterministic integration theory for highly irregular paths, and has led to several developments in stochastic differential equations [FV10; FH20]. This theory has been well-developed for paths valued in a Banach space [LQ07]. Therefore, using the Lipschitz-free space as a Banach space for persistence diagrams, we may define and study *rough paths of persistence diagrams*. This opens up the possibility of studying the persistent homology of data driven by stochastic differential equations.

- **Banach space learning.** While kernel methods for Hilbert spaces are a mature set of tools in machine learning, the corresponding generalization to Banach spaces is still fairly recent. As discussed in Proposition 4.2.5, the Lipschitz free space $\mathcal{F}(\widetilde{\Omega})$ is not reflexive, and thus many of the results for Banach space kernel methods [ZXZ09; Sch19a] do not hold. Can we develop kernel methods using the approximate representer theorem [Sch20] for non reflexive Banach spaces? What is the advantage of this compared to other methods?

- **Lipschitz-free spaces.** Although we have discussed several properties of Lipschitz-free spaces which are directly applicable to the study of persistence diagrams, are there other aspects of this connection which we can exploit? One concrete problem is the an explicit description of the Lipschitz-free space of the quotient metric space of

a metric pair (X, d, A). Does Question 4.2.1 have an affirmative answer? Or can we adapt the representations for known Lipschitz-free spaces such as $\mathcal{F}(\mathbb{R}^N)$ [CKK17] to develop a new representation of persistence diagrams?

Part III

GENERALIZATIONS OF THE DOMAIN

5

THE MAPPING SPACE CONSTRUCTION

Chen's construction for path spaces in Chapter 2 provided a cochain model for path spaces and loop spaces from a de Rham-theoretic perspective. The underlying formalism behind this construction is much more general, and has been used to build cochain algebras for mapping spaces in order to study their cohomology. The general idea consists of two steps:

1. Given suitable smooth spaces K and Y, build a cosimplicial model for the smooth mapping space $\mathsf{Diff}(K, Y)$,

2. Given a cosimplcial space $X_\bullet$, build a cochain algebra which is quasi-isomorphic to the singular cochain complex of the totalization of $X_\bullet$.

When we use de Rham cochains for the construction of the cochain algebra for the cosimplicial space, we call the resulting cochain construction a *Chen complex*. One of the earliest papers that considered the cohomology of mapping spaces in this direction was [And72], which announced a generalization of the Eilenberg-Moore spectral sequence, now called the Anderson spectral sequence, though proofs were omitted. Later, Bott and Segal [BS77] provided an explicit description of the cochain construction for mapping spaces, where the source of the mapping space has an open contractible covering (along with finiteness and connectivity conditions). Bendersky and Gitler [BG91] considered the more general case where the source is the realization of a simplicial set, though their proof was not satisfactory for certain spaces. Finally, Patras and Thomas [PT03] consolidated the ideas through Kan adjunction techniques, and proved the quasi-isomorphism for $\mathsf{Top}(K, Y)$ when K is the realization of a simplicial set weakly homotopy equivalent to a finite simplicial set of dimension less than or equal to the connectivity of Y.

While the Patras and Thomas construction is stated for the topological setting and does not explicitly use smooth de Rham cochains, their methods are general, and the results using de Rham cochains and an explicit description of the Chen complex for mapping spaces is provided in [GTZ10]. Motivated by the observation that the path signature coefficients constitute the 0-cochains in Chen's construction for path spaces, we aim to use the 0-cochains of Chen's mapping space construction to define a notion of *mapping space*

signatures. In particular, our goal in this part of the thesis is to define a mapping space signature for

$$\mathsf{Diff}(\square^d, \mathbb{R}^N).$$

A cubical domain is chosen due to its simplicity as well as the existence of compositions in each dimension. While unit d-cubes can certainly be constructed as a simplicial set, the number of nondegenerate top dimensional simplices increases on the order of $d!$, and the composition of d-cubes is more difficult to express in the simplicial setting.

The goal of this chapter is to formulate a cubical variant of Chen's complex for mapping spaces, using cubical sets with connection. We begin with an overview of cubical sets, followed by the topological construction. While we do not show that our cubical variant of the mapping space construction leads to a quasi-isomorphism, we will show that it is indeed a morphism of differential graded algebras. Finally, we extract the 0-cochains from this construction, which we call *monomials*, which will be used to define signatures in the following chapters.

5.1 INTRODUCTION TO CUBICAL SETS WITH CONNECTION

Just as simplicial sets [May92] can be viewed as a generalization of simplicial complexes, cubical sets can be thought of as a generalization to cubical complexes in algebraic topology. Simplicial and cubical sets can be defined as a sets of simplices or cubes, along with information about how these objects are glued, in the form of *face maps* and *degeneracy maps*. We obtain a topological space from a simplicial or cubical set through *geometric realization*. In fact, we can define a homotopy theory (closed model category) for simplicial sets [GJ12] and cubical sets [Jar02] that is equivalent to the standard homotopy theory for topological spaces. Thus, simplicial and cubical sets are used as a combinatorial approach to the homotopy theory of topological spaces.

However, a major defect in the homotopy theory for cubical sets is that the geometric realization does not preserve products. In other words, the realization of a product of two cubical sets does not have the same homotopy type as the product of their realizations. *Cubical sets with connections* were introduced in [BH81] to address this issue by adding additional degeneracies to cubical sets called *connections*. Thus, cubical sets with connections provide a more appropriate model of topological spaces.

In this section, we provide an introduction to cubical sets with connection which does not require prior knowledge of simplicial sets. This section is entirely expository and primarily references [BHS11; Ant02].

5.1.1 *The Box Category and Connections*

Definition 5.1.1. The *box category* $\square$ is a subcategory of Top where the objects are the standard n-cubes $\square^n := [0,1]^n$ for $n \geq 0$, and the morphisms $\square(\square^n, \square^m)$ are obtained by compositions of *coface maps* $\delta_i^\epsilon : \square^n \to \square^{n+1}$ defined by

$$\delta_i^\epsilon(x_1, \ldots, x_n) := (x_1, \ldots, x_{i-1}, \epsilon, x_i, \ldots, x_n), \qquad i \in [n],\ \epsilon \in \{0,1\}, \tag{5.1.1}$$

as well as *codegeneracy maps* $\sigma_i : \square^n \to \square^{n-1}$ defined by

$$\sigma_i(x_1, \ldots, x_n) := (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n), \qquad i \in [n]. \tag{5.1.2}$$

The coface and codegeneracy maps satisfy the following *cocubical identites*:

$$\delta_j^{\epsilon_2}\delta_i^{\epsilon_1} = \delta_i^{\epsilon_1}\delta_j^{\epsilon_2}, \quad i < j, \tag{5.1.3}$$

$$\sigma_j\sigma_i = \sigma_i\sigma_{j+1}, \quad i \leq j, \tag{5.1.4}$$

$$\sigma_j\delta_j^\epsilon = \begin{cases} \delta_i^\epsilon\sigma_{j-1} & i < j, \\ \delta_{i-1}^\epsilon\sigma_j & i > j \\ \mathrm{id} & i = j \end{cases} \tag{5.1.5}$$

Note that while we use $\square$ to denote both the box category and the unit interval, the meaning will be clear from context. There are two ways to view the box category:

1. as an *abstract category*, ignoring the explicit definitions of the objects and morphisms as those internal to Top, and
2. as a *cocubical space*, which is a functor $F : \square \to \mathsf{Top}$, defined as the identity functor, which preserves the explicit structure of the objects and morphisms.

We begin by describing its structure as an abstract category, and return to the perspective of a cocubical space when we discuss geometric realization. As an abstract category, we ignore the topological structure of the individual objects, and to emphasize this view, we describe the set of objects as the natural numbers $\mathbb{N}$, and denote the objects using underlined integers such as $\underline{n} := \square^n$. Furthermore, we also ignore the topological meaning of the morphisms as maps between topological spaces; instead we *define* the coface and codegeneracy maps as those which satisfy the cocubical identites.

We can now define cubical sets as a functor. Note that the topologial structure of the box category is irrelevant in this definition, so we only need to consider it as an abstract category.

Definition 5.1.2. The category $\mathsf{Set}^{\square^{\mathrm{op}}}$ of *cubical sets* is defined to be the functor category $\mathsf{Cat}(\square^{\mathrm{op}}, \mathsf{Set})$. Thus, a cubical set is a functor

$$K : \square^{\mathrm{op}} \to \mathsf{Set}$$

and a map of cubical sets is a natural transformation of functors. We will assume that all cubical sets are *finite*, defined in Remark 5.1.1.

More concretely, a cubical set $K : \square^{\mathrm{op}} \to \mathsf{Set}$ is a collection of objects $K_n \coloneqq K(\underline{n})$ where $n \in \mathbb{N}$, and a collection of maps between the K_n generated by the *face maps* $d_i^\epsilon \coloneqq K(\delta_i^\epsilon) : K_{n+1} \to K_n$ for $i \in [n]$ and $\epsilon \in \{0,1\}$, and the *degeneracy maps* $s_i \coloneqq K(\sigma_i) : K_{n-1} \to K_n$ for $i \in [n]$. The face and degeneracy maps satisfy the dual of the cocubical identities, which are called the *cubical identities*:

$$d_i^{\epsilon_1} d_j^{\epsilon_2} = d_{j-1}^{\epsilon_2} d_i^{\epsilon_1}, \quad i < j, \tag{5.1.6}$$

$$s_i s_j = s_{j+1} s_i, \quad i \leq j, \tag{5.1.7}$$

$$d_j^\epsilon s_j = \begin{cases} s_{j-1} d_i^\epsilon & i < j \\ s_j d_{i-1}^\epsilon & i > j \\ \mathrm{id} & i = j. \end{cases} \tag{5.1.8}$$

Suppose $K_\bullet \in \mathsf{Set}^{\square^{\mathrm{op}}}$ and $a \in K_n$. If there exists some $b \in K_{n-1}$ and some $i \in [n]$ such that $a = s_i b$, then we say that a is a *degenerate cube*; otherwise, we say that a is a *nondegenerate cube*. Furthermore, every degenerate cube a can be written in a canonical form as

$$a = s_{i_k} s_{i_{k-1}} \dots s_{i_1} b \tag{5.1.9}$$

where $1 \leq i_1 < i_2 < \dots < i_n \leq n$, and $b \in K_{n-k}$ is a nondegenerate cube. In this case, we say that a is a *degeneracy* of b.

Remark 5.1.1. All cubical sets $K_\bullet \in \mathsf{Set}^{\square^{\mathrm{op}}}$ considered in this thesis are *finite*: each $K_\bullet$ has a finite number of non-degenerate cubes.

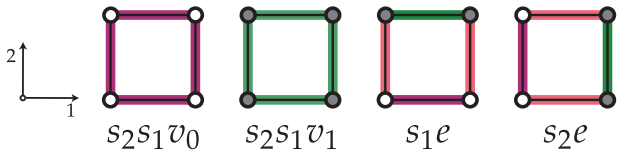
Intuitively, we can think of a cubical set $K_\bullet \in \mathsf{Set}^{\square^{\mathrm{op}}}$ as a combinatorial model for a cubical complex, where the elements $a \in K_n$ are combinatorial n-cubes where each degree k subcube is labelled by an element of K_k, specifying the gluing data.

Example 5.1.1. We describe a cubical set $I_\bullet \in \mathsf{Set}^{\square^{\mathrm{op}}}$ that models the unit interval $[0,1]$. As a cubical complex, the unit interval $[0,1]$ can be describing using two vertices, v_0 and v_1, modelling the endpoints, and a single edge, e, modelling the interval. Indeed, the 0-cubes of the cubical set are $I_0 = \{v_0, v_1\}$, which we represent visually using white and gray points.

The 1-cubes in $I_\bullet$ contain the expected edge e, but it also contains the degeneracies of the 0-cubes. Thus, we have $I_1 = \{s_1v_0, s_1v_1, e\}$. We visually represent each of these 1-cubes using colored edges.

The arrow on the left is used to specify a numbered coordinate direction, which is used to read off the face maps. In general, an n-cube has two faces in each of the n-coordinate directions, and each face is specified using the face maps d_i^ϵ. The index $i \in [n]$ corresponds to the coordinate direction, and $\epsilon \in \{0, 1\}$ specifies the "back" or "front" face. The arrow points from 0 to 1. Thus from the figure, we can read off $d_1^0 e = v_0$ as the white point, and $d_1^1 e = v_1$ as the black point.

While we do not expect any non-degenerate 2-cubes, we must include all possible degeneracies of v_0, v_1, and e. Therefore, the 2-cubes are $I_2 = \{s_2s_1v_0, s_2s_1v_1, s_1e, s_2e\}$, represented visually in the following figure.



The axes on the left now specify the two possible coordinate directions, and we can read off the face maps for each 2-cube as before. As we can see in this figure, the degeneracy s_ia of some $n-1$ cube a is an n-cube where the cube a is copied along the i direction.

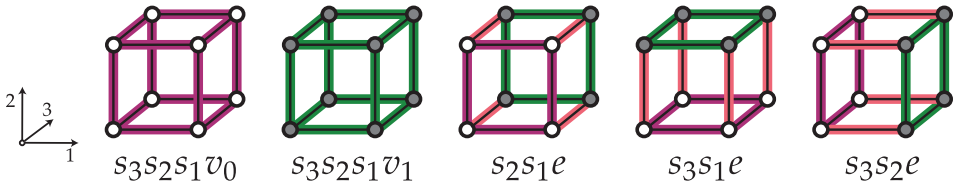
The higher dimensions will also contain only degenerate cubes, and they can be represented in the same manner. The 3-cubes can be listed as

$$I_3 = \{s_3s_2s_1v_0,\ s_3s_2s_1v_1,\ s_2s_1e,\ s_3s_1e,\ s_3s_2e\},$$

where we write all degeneracies in their canonical form specified by Equation (5.1.9).

As stated in the introduction, ordinary cubical sets lack certain properties when we consider their realization as topological spaces. Thus, we consider a modification introduced by Brown and Higgins [BH81].

Definition 5.1.3. The *box category with connections* $\square_c$ is a subcategory of Top with the same objects as the objects in $\square$. The morphisms $\hom_{\square_c}(\square^n, \square^m)$ are obtained by compositions of



the coface and codegeneracy maps from $\square$ as well as the *coconnection maps* $\gamma_i : \square^n \to \square^{n-1}$ defined by

$$\gamma_i(x_1, \ldots, x_n) \coloneqq (x_1, \ldots, x_{i-1}, \max(x_i, x_{i+1}), x_{i+2}, \ldots, x_n), \qquad i \in [n-1], \tag{5.1.10}$$

where $n \geq 2$. As before, the coface and codegeneracy maps satisfy the cocubical identites. Along with the coconnection maps, the three classes of maps satisfy the additional *coconnection identites*:

$$\gamma_j \delta_i^\epsilon = \begin{cases} \delta_i^\epsilon \gamma_{j-1} & i < j, \\ \delta_{i-1}^\epsilon \gamma_j & i > j+1 \end{cases} \tag{5.1.11}$$

$$\gamma_j \delta_j^0 = \delta_{j+1}^0 \gamma_j = id \tag{5.1.12}$$

$$\gamma_j \delta_j^1 = \delta_{j+1}^1 \gamma_j = \delta_j^1 \sigma_j \tag{5.1.13}$$

$$\sigma_j \gamma_i = \begin{cases} \gamma_i \sigma_{j-1} & i < j, \\ \gamma_{i-1} \sigma_j & i > j \end{cases} \tag{5.1.14}$$

$$\sigma_j \gamma_j = \sigma_j \sigma_j = \sigma_j \sigma_j + 1 \tag{5.1.15}$$

$$\gamma_j \gamma_i = \gamma_i \gamma_{j+1} \tag{5.1.16}$$

As with the ordinary box category, we may interpret the box category with connections as an abstract category or as a cocubical space with connections. We defer the discussion of the cocubical space, and focus on the abstract category perspective. As before, the object set of $\square_c$ can be specified by $\mathbb{N}$, and the morphisms of $\square_c$ are defined by the cocubical and coconnection identities. Thus, the only difference is the existence of additional morphisms.

Definition 5.1.4. The category $\mathsf{Set}^{\square_c^{\mathrm{op}}}$ of *cubical sets with connections* is defined to be the functor category $\mathsf{Cat}(\square_c^{\mathrm{op}}, \mathsf{Set})$. Thus, a cubical set with connections is a functor

$$K : \square_c^{\mathrm{op}} \to \mathsf{Set}$$

and a map of cubical sets with connections is a natural transformation of functors.

Concretely, a cubical set with connection $K : \square_c^{\mathrm{op}} \to \mathrm{Set}$ has the same objects K_n, face maps $d_i^\epsilon : K_{n+1} \to K_n$, and degeneracy maps $s_i : K_{n-1} \to K_n$ as a cubical set. In addition, we also have the *connection maps* $g_i := K(\gamma_i) : K_{n-1} \to K_n$, which only exist for $n \geq 2$. In addition to the cubical identites, these objects satisfy the *connection identities*:

$$d_i^\epsilon g_j = \begin{cases} g_{j-1} d_i^\epsilon & i < j, \\ g_j d_{i-1}^\epsilon & i > j+1 \end{cases} \tag{5.1.17}$$

$$d_j^0 g_j = g_j d_{j+1}^0 = \mathrm{id} \tag{5.1.18}$$

$$d_j^1 g_j = g_j d_{j+1}^1 = s_j d_j^1 \tag{5.1.19}$$

$$g_i s_j = \begin{cases} s_{j-1} g_i & i < j, \\ s_j & i g_{i-1} > j \end{cases} \tag{5.1.20}$$

$$g_j s_j = s_j s_j = s_j + 1 s_j \tag{5.1.21}$$

$$g_i g_j = g_{j+1} g_i \tag{5.1.22}$$

Suppose $K_\bullet \in \mathrm{Set}^{\square_c^{\mathrm{op}}}$ and $a \in K_n$. If there exists some $b \in K_{n-1}$ and some $i \in [n]$ such that $a = s_i b$ or some $j \in [n-1]$ such that $a = g_j b$, then we say that a is a *degenerate cube*; otherwise, we say that a is a *nondegenerate cube*. In this way, we view connection maps as "extra degeneracies." Furthermore, every degenerate cube a can be written in a canonical form as

$$a = s_{i_k} s_{i_{k-1}} \dots s_{i_1} g_{j_l} g_{j_{l-1}} \dots g_{j_1} b \tag{5.1.23}$$

where

$$1 \leq i_1 < i_2 < \dots < i_n \leq n$$
$$1 \leq j_1 < j_2 < \dots < j_l \leq n-k-1$$

and $b \in K_{n-k-l}$ is a nondegenerate cube. In this case, we say that a is a *degeneracy* of b.

Remark 5.1.2. If $\overline{I} = (i_1, \dots, i_n)$ and $\overline{J} = (j_1, \dots, j_l)$, we will simplify the notation and write

$$s_{\overline{I}} g_{\overline{J}} b := s_{i_k} s_{i_{k-1}} \dots s_{i_1} g_{j_l} g_{j_{l-1}} \dots g_{j_1} b. \tag{5.1.24}$$

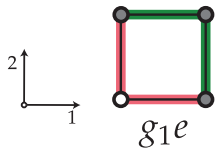
Therefore, we can identify degenerate cubes in terms of their unique non-degenerate cube b, as well as the list of degeneracy and connection maps indexed by $\mathcal{I}$ and $\mathcal{J}$. Thus, we can express $a \in K_n$ as $a = (\overline{I}, \overline{J}, b)$.

Intuitively, we can view cubical sets with connection in the same manner as ordinary cubical sets, but we have additional degeneracies.

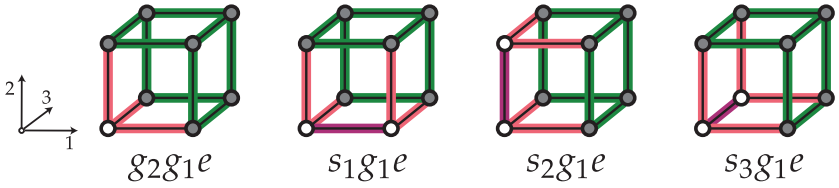
Example 5.1.2. We expand on Example 5.1.1 to specify a model of the unit interval $[0,1]$ using a cubical set with connections $I^c_\bullet$. This new cubical set contains all of the cubes from $I_\bullet$, but we must also include the additional connection degeneracies. Note that the connection maps $g_i : I^c_{n-1} \to I^c_n$ only exist for $n \geq 2$, so the 0-cubes and 1-cubes remain the same:

$$I^c_0 = I_0, \quad I^c_1 = I_1.$$

We have one additional 2-cube, $g_1 e$, which is represented below. Note that any connection degeneracies can be rewritten as ordinary degeneracies according to the connection identities.



We have four additional degenerate 3-cubes, which we specify using the canonical form in Equation (5.1.23).



5.1.2 *Composition*

One advantage of using cubical sets rather than simplicial sets is that cubical sets can be equipped with a *composition structure*.

Definition 5.1.5. A *cubical set with connections and compositions* is a cubical set $K_\bullet : \square_c^{\mathrm{op}} \to$ Set such each for each $n \geq 1$, the set of n-cubes has n partial compositions $+_i$. For $a, b \in K_n$, the composition $a +_i b$ is defined if and only if $d^1_i a = d^0_i b$. The relationships between these partial compositions and the cubical maps are given by the following *composition identites*

$$d_i^\epsilon(a +_j b) = \begin{cases} d_i^\epsilon a +_{j-1} d_i^\epsilon b & : i < j \\ d_i^0 a & : i = j, \epsilon = 0 \\ d_i^1 b & : i = j, \epsilon = 1 \\ d_i^\epsilon a +_j d_i^\epsilon b & : i > j \end{cases} \tag{5.1.25}$$

$$s_i(a +_j b) = \begin{cases} s_i a +_{j+1} s_i b & : i \leq j \\ s_i a +_j s_i b & : i > j \end{cases} \tag{5.1.26}$$

$$g_i(a +_j b) = \begin{cases} g_i a +_{j+1} g_i b & : i < j \\ (g_i a +_{i+1} s_i b) +_i (s_{i+1} b +_{i+1} g_i b) & : i = j \\ g_i a +_j g_i b & : i > j. \end{cases} \tag{5.1.27}$$

Finally, we also have the following *interchange law* when both sides are defined:

$$(a +_i b) +_j (c +_i d) = (a +_j c) +_i (b +_j d). \tag{5.1.28}$$

5.1.3 *Geometric Realization and Evaluation Maps*

In this section, we will consider the standard geometric realization of cubical sets (with connection) as a topological space.

We can define cubical objects and cubical objects with connection in an arbitrary category C as functors.

Definition 5.1.6. Let C be a category. A *cubical object in* C (resp. *cubical object with connection in* C) is defined as a functor

$$K_\bullet : \square^{\mathrm{op}} \to \mathsf{C}, \quad (K_\bullet : \square_c^{\mathrm{op}} \to \mathsf{C}).$$

The category of such objects is denoted by $\mathsf{C}^{\square^{\mathrm{op}}}$ (resp. $\mathsf{C}^{\square_c^{\mathrm{op}}}$).

A *cocubical object in* C (resp. *cubical object with connection in* C) is defined to be a functor

$$K^\bullet : \square \to \mathsf{C}, \quad (K^\bullet : \square_c \to \mathsf{C}).$$

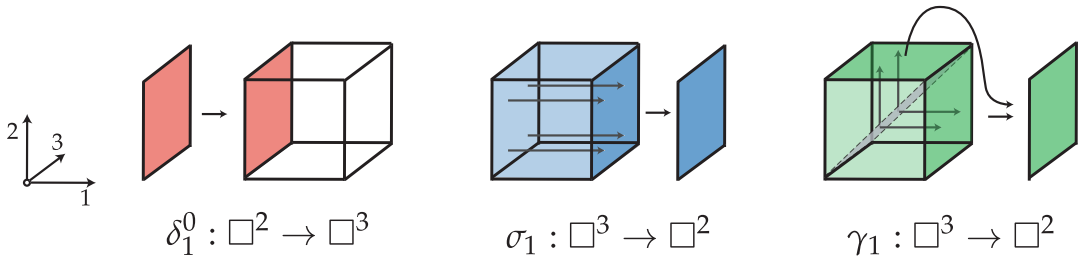
The category of such objects is denoted by $\mathsf{C}^{\square}$ (resp. $\mathsf{C}^{\square_c}$).

(Co)cubical objects in Top are called *(co)cubical spaces*. By using the definition of $\square$ and $\square_c$ as a subcategory of Top, we can interpret both of these objects as cocubical spaces

via the identity functor. Recall the topological definition of the coface, codegeneracy and coconnection maps:

$$\begin{aligned}
\delta_i^\epsilon(x_1,\dots,x_n) &\coloneqq (x_1,\dots,x_{i-1},\epsilon,x_i,\dots,x_n), \qquad i\in[n],\ \epsilon\in\{0,1\}\\
\sigma_i(x_1,\dots,x_n) &\coloneqq (x_1,\dots,x_{i-1},x_{i+1},\dots,x_n), \qquad i\in[n]\\
\gamma_i(x_1,\dots,x_n) &\coloneqq (x_1,\dots,x_{i-1},\max(x_i,x_{i+1}),x_{i+2},\dots,x_n), \qquad i\in[n-1].
\end{aligned}$$

The following figure demonstrates these maps between $\square^2$ and $\square^3$.



$\delta_1^0 : \square^2 \to \square^3$ $\qquad$ $\sigma_1 : \square^3 \to \square^2$ $\qquad$ $\gamma_1 : \square^3 \to \square^2$

Definition 5.1.7. The *geometric realization* $|\cdot|$ is defined as follows. For $K_\bullet \in \mathsf{Set}^{\square^{\mathrm{op}}}$ (resp. $K_\bullet \in \mathsf{Set}^{\square_c^{\mathrm{op}}}$), the geometric realization of $K_\bullet$ is

$$|K_\bullet| \coloneqq \coprod_{n=0}^{\infty} K_n \times \square^n \Big/ \sim, \tag{5.1.29}$$

where K_n is equipped with the discrete topology, and $\sim$ is an equivalence relation on $\coprod_{n=0}^{\infty} K_n \times \square^n$ defined by

$$(a(x),t) \sim (x,\alpha(t)), \quad x\in K_m, \quad t\in\square^n \tag{5.1.30}$$

where $\alpha : \square^n \to \square^m$ is any morphism in the box category $\square$ (resp. the box category with connection $\square_c$), and $a = K(\alpha) : K_m \to K_n$.

In all of these cases, the geometric realization of $K_\bullet$ is a CW-complex with an n-cell for each n-cube in K_n. Then, for each n-cube in K_n, there exists an inclusion called the *evaluation map* into the geometric realization, which will be crucial to the mapping space construction.

Definition 5.1.8. For each n-cube $a \in K_n$, we define its *evaluation map* $\eta[a] : \square^n \to |K_\bullet|$ to be

$$\eta[a] : \square^n \hookrightarrow \coprod_{n=0}^{\infty} K_n \times \square^n \twoheadrightarrow |K_\bullet|, \tag{5.1.31}$$

where the first map is an inclusion into $\{a\} \times \square^n$, and the second map is the projection onto the quotient.

Note that the CW-complex may be described using only the non-degenerate cubes since all degenerate cubes are quotiented out by the equivalence relation. Indeed, for a cubical set $K_\bullet \in \mathsf{Set}^{\square_c^{\mathrm{op}}}$, recall from Equation (5.1.23) that any degenerate cube $a \in K_n$ can be canonically written as

$$a = s_{i_k} s_{i_{k-1}} \dots s_{i_1} g_{j_l} g_{j_{l-1}} \dots g_{j_1} b$$

where

$$1 \leq i_1 < i_2 < \dots < i_n \leq n$$
$$1 \leq j_1 < j_2 < \dots < j_l \leq n-k-1,$$

and $b \in K_{n-k-l}$ is a nondegenerate cube. We can explicitly define the evaluation map for this degenerate cube. For $x \in \square^n$, we have

$$\begin{aligned} \eta[a](x) &= [(a, x)] \\ &= [(s_{i_k} s_{i_{k-1}} \dots s_{i_1} g_{j_l} g_{j_{l-1}} \dots g_{j_1} b, x)] \\ &= [(b, \gamma_{j_1} \dots \gamma_{j_l} \sigma_{i_1} \dots \sigma_{i_k} x)], \end{aligned}$$

where the square brackets denote the equivalence class in $|K_\bullet|$. This shows that each degenerate n-cube a is mapped onto its unique underlying non-degenerate cube b. Furthermore, this shows the following factorization of the evaluation map.

Lemma 5.1.1. *Let $K_\bullet \in \mathsf{Set}^{\square_c^{\mathrm{op}}}$ be a cubical set with connection and $a \in K_n$ be an n-cube, written in its canonical form as*

$$a = s_{i_k} s_{i_{k-1}} \dots s_{i_1} g_{j_l} g_{j_{l-1}} \dots g_{j_1} b.$$

Let $\overline{I} = (i_1, \dots, i_k)$ and $\overline{J} = (j_1, \dots, j_l)$. The evaluation map $\eta[a] : \square^n \to |K|$ can be factored as the composition

$$\eta[a] : \square^n \xrightarrow{\eta_n^{\overline{I},\overline{J}}} \square^{n-k-l} \xrightarrow{\eta[b]} |K|, \tag{5.1.32}$$

where $\eta_n^{\overline{I},\overline{J}} : \square^n \to \square^{n-k-l}$ is defined by the following composition of codegeneracy and coconnection maps

$$\eta_n^{\overline{I},\overline{J}}(\mathbf{t}) = \gamma_{j_1} \dots \gamma_{j_l} \sigma_{i_1} \dots \sigma_{i_k}(\mathbf{t}), \tag{5.1.33}$$

for all $\mathbf{t} \in \square^n$.

In fact, we can explicitly define $\eta_n^{\overline{I},\overline{J}}$.

Definition 5.1.9. A *k-ordered subset* of $[n]$ is an ordered subset $I \subset [n]$ equipped with a partition into k nonempty subsets, denoted by $I^{(r)}$ for $r = 1, \dots, k$, such that if $r < s$, and $x \in I^{(r)}, y \in I^{(s)}$, then $x < y$.

Lemma 5.1.2. *Let $n \in \mathbb{N}$, $\overline{I} = (i_1, \dots, i_k) \subset [n]$ be an ordered subset and $\overline{J} = (j_1, \dots, j_l) \subset [n-k-1]$ be another ordered subset. There exists a $(n-k-l)$-ordered subset $I \subset [n]$ such that the underlying subset I is the complement of $\overline{I}$ in $[n]$ and*

$$\eta_n^I(t_1, \dots, t_n) := \eta_n^{\overline{I},\overline{J}}(t_1, \dots, t_n) = (s_1, \dots, s_{n-k-l}), \tag{5.1.34}$$

where

$$s_r = \max_{j \in I^{(r)}} \{t_j\}. \tag{5.1.35}$$

Proof. The proof of this lemma follows by expanding the definitions of the codegeneracy and coconnection maps in Equation (5.1.2) and Equation (5.1.10). We begin by considering the case where $\overline{I} = \emptyset$ so that the complement is $I = [n]$. In this case, we have $\overline{J} = (j_1, \dots, j_l) \subset [p-1]$ and we need to consider the iterated coconnection map

$$\eta_n^{\emptyset,\overline{J}} = \gamma_{j_1} \dots \gamma_{j_l}.$$

A consecutive subsequence in $\overline{J}$ corresponds to consecutive coconnection maps, which can be written as

$$\begin{aligned} \gamma_{j-m}\gamma_{j-m+1} \dots \gamma_j(t_1, \dots, t_n) &= \gamma_{j-m} \dots \gamma_{j-1}(t_1, \dots, \max(t_j, t_{j+1}), \dots, t_n) \\ &= \gamma_{j-m} \dots \gamma_{j-2}(t_1, \dots, \max(t_{j-1}, t_j, t_{j+1}), \dots, t_n) \\ &= (t_1, \dots, \max(t_{j-m}, \dots, t_{j+1}), \dots, t_n). \end{aligned}$$

Thus, a maximal consecutive sequence $(j-m, \dots, j) \subset \overline{J}$ corresponds to one part of the partition $I^{(r)} = (j-m, \dots, j, j+1)$. For each maximal consecutive sequence in $\overline{J}$ (including sequences of length 1), we add the corresponding subset of I to the partition. Once every element of $\overline{J}$ has been accounted for, we add the remaining elements of I into the partition as singletons. Note that this procedure will provide a partition of I made up of exactly $p-l$ subsets, since each element of J decreases the number of singletons in the partition by 1.

Next, we consider the case where $\overline{I} = (i_1, \dots, i_p)$, so we must consider the coconnections as well as the codegeneracies. However, note that codegeneracies simply omit coordinates, so that

$$\sigma_{i_1} \dots \sigma_{i_p}(t_1, \dots, t_n) = (t_j)_{j \in I},$$

where I is the complement of $\overline{I}$ in $[n]$. Now, we can simply repeat the previous construction of the partition with a smaller set I. □

Example 5.1.3. As a concrete example, consider $n = 8$, $\overline{I} = (5,7)$ and $\overline{J} = (1,4,5)$. We begin by taking the complement $I = (1,2,3,4,6,8)$. Then, we consider the two maximal consecutive subsequences in $\overline{J}$: (1) and $(4,5)$. For the subsequence (1), we add the first and second element of I into a subset $I^{(1)} = (1,2)$ in the partition. For the subsequence $(4,5)$, we add the fourth, fifth and sixth element of I into a subset $I^{(3)} = (4,6,8)$ in the partition. The only element in I unaccounted for is the third element, so we add it to the partition as a singleton $I^{(2)} = (3)$. Thus, the evaluation map can be explicitly written as

$$\eta_8^{\overline{I},\overline{J}}(t_1, \dots, t_8) = (\max(t_1, t_2), t_3, \max(t_4, t_6, t_8)).$$

Proposition 5.1.1. *Let $K_\bullet \in \mathsf{Set}^{\square_c^{\mathrm{op}}}$. Every n-cube $a \in K_n$ can be represented by a pair (I, b), where b is the unique non-degenerate k-cube corresponding to a, and I is a k-ordered subset. Furthermore, every such pair (I, b) corresponds to an n-cube $a \in K_n$. This will be called the* ordered subset notation *for n-cubes.*

Proof. By writing a in canonical form in terms of the degeneracy and connection maps, we can uniquely represent a using two subsets $\overline{I}, \overline{J}$ as well as the non-degenerate cube b. From Lemma 5.1.2, there exists a unique k-ordered subset I for every pair of subsets $\overline{I}$ and $\overline{J}$. Thus, we can use (I, b) as a representation for a. □

Given an n-cube $a = (I, b)$, where $b \in K_q$ is non-degenerate, the factorization of $\eta[a]$ given in Lemma 5.1.1 can be rewritten as

$$\eta[(I,b)] : \square^n \xrightarrow{\eta_n^I} \square^q \xrightarrow{\eta[b]} \to |K|, \tag{5.1.36}$$

where η_n^I is defined in Equation (5.1.34).

Remark 5.1.3. We note that we can describe geometric realization as a colimit. Indeed, for $K_\bullet \in \mathsf{Set}^{\square_c^{\mathrm{op}}}$, the geometric realization of $K_\bullet$ is

$$|K_\bullet| = \mathsf{coeq}\left[\coprod_{\alpha : \square^n \to \square^m} K_m \times \square^n \rightrightarrows \coprod_n K_n \times \square^n \right]. \tag{5.1.37}$$

The first coproduct is taken over all maps in the box category $\square_c$. For a given map $\alpha : \square^n \to \square^m$ in $\square_c$, we have two maps from $K_m \times \square^n$ into $\coprod_n K_n \times \square^n$:

1. $K(\alpha) \times \mathrm{id} : K_m \times \square^n \to K_n \times \square^n$
2. $\mathrm{id} \times \alpha : K_m \times \square^n \to K_m \times \square^m$.

The condition enforced by this coequalizer is the same as the equivalence relation given in the original definition of geometric realization. This formulation emphasizes the duality between geometric realization and the totalization of cocubical spaces defined in the next section.

5.1.4 *Smooth Geometric Realization*

Because we will be working with differential forms, we require a slight variation on the geometric realization in order to produce a differentiable space, introduced in Definition 2.2.1. Note that the geometric realization is defined as the quotient of a coproduct. The coproduct of differentiable spaces (as in Example 2.2.4) and the quotient of a differentiable space (as in Example 2.2.3) are both differentiable spaces.

We wish to formulate a smooth version of geometric realization of cubical sets with connection as a colimit as in Remark 5.1.3. However, the obstruction to such a definition is the fact that the coconnection maps defined using maximum functions are not smooth with respect to the ordinary smooth structure (the subspace smooth structure for $\square^n \subset \mathbb{R}^N$ as in Example 2.2.2), and thus are not morphisms in Diff. We will remedy this be defining a new smooth structure on the geometric n-cubes $\square^n$ such that the maximum functions are smooth.

To begin, we note that each n-cube $\square^n$ can be viewed as a simplicial complex, where the simplices are defined using the ordering of the coordinates. These simplices can be described using functions

$$f : [n] \to [n] \cup \{\epsilon_0, \epsilon_1\}. \tag{5.1.38}$$

Such a function corresponds to the following subsimplex of $\square^n$,

$$\Delta_f := \left\{ (t_1, \ldots, t_n) \in \square^n \; : \; \begin{array}{ll} t_i = 0 & : f(i) = \epsilon_0 \\ t_i = 1 & : f(i) = \epsilon_1 \\ t_i = t_j & : f(i) = f(j) \\ t_i \leq t_j & : f(i) < f(j). \end{array} \right\} \tag{5.1.39}$$

In most cases, we will only need to work with the n-subsimplices of $\square^n$, and these subsimplices can be defined using permutations. Given $\sigma \in \Sigma_n$, we define

$$\Delta^n_\sigma := \{(t_1, \ldots, t_n) \in \square^n \,:\, 0 \leq t_{\sigma(1)} \leq \ldots \leq t_{\sigma(n)} \leq 1\}. \tag{5.1.40}$$

In particular, the union of the n-subsimplices defined over all permutations $\sigma \in \Sigma_n$ is the entire n-cube

$$\bigcup_{\sigma \in \Sigma_n} \Delta^n_\sigma = \square^n,$$

and the intersections occur at the lower dimensional subsimplices. Note that maximum functions and in particular coconnection maps restricted to a n-subsimplex Δ^n_σ is smooth. This is due to the ordering of the coordinates being fixed in Δ^n_σ, and thus maximum functions restricted to Δ^n_σ are simply projection maps.

This leads to the motivating property of the desired smooth structure on $\square^n$. In the following lemma, we define a smooth structure on $\square^n$ such that a map $f : \square^n \to X$ is smooth if and only if the restrictions $f|_{\Delta^n_\sigma} : \Delta^n_\sigma \to X$ are smooth for all $\sigma \in \Sigma_n$.

Definition 5.1.10. The *unit n-cube with the simplicial smooth structure*, denoted by $\square^n_\Delta \in \mathsf{Diff}$, will be defined as follows. For each $\sigma \in \Sigma_n$, let $\Delta^n_\sigma \in \mathsf{Diff}$ be equipped with the subspace smooth structure as a subspace $\Delta^n_\sigma \subset \mathbb{R}^n$. The disjoint union

$$U = \coprod_{\sigma \in \Sigma_n} \Delta^n_\sigma \in \mathsf{Diff}$$

is a differentiable space using the coproduct smooth structure from Example 2.2.4. We define an equivalence relation on the disjoint union U as follows. Let $\mathbf{s} \in \Delta^n_\sigma$ and $\mathbf{t} \in \Delta^n_\tau$ be elements from distinct simplices. We set $\mathbf{s} \sim \mathbf{t}$ if and only if $\mathbf{s} = \mathbf{t}$ in the parametrizations. Finally, we define

$$\square^n_\Delta := \coprod_{\sigma \in \Sigma_n} \Delta^n_\sigma \Big/ \sim \quad \in \mathsf{Diff}, \tag{5.1.41}$$

equipped with the quotient smooth structure from Example 2.2.3.

In particular, a map $\phi : C \to \square^n_\Delta$ is a plot if and only if there exists there exists an open cover $\{C_i\}$ of C, along with plots $\phi_i : C_i \to \Delta^n_{\sigma_i}$ into the subsimplices such that the following diagram commutes

$$\begin{array}{ccc} C_i & \xrightarrow{\phi_i} & \Delta^n_{\sigma_i} \\ {\scriptstyle \iota_i}\downarrow & & \downarrow{\scriptstyle \iota} \\ C & \xrightarrow[\phi]{} & \square^n_\Delta, \end{array}$$

where $\iota_i : C_i \hookrightarrow C$ are the inclusion maps for the cover and $\iota : \Delta^n_{\sigma_i} \hookrightarrow \square^n_\Delta$ is the inclusion of the subsimplex into the cube. This smooth structure on the unit cube has the desired property of smooth maps.

Lemma 5.1.3. *Let X be a differentiable space. A map $f : \square^n_\Delta \to X$ is smooth if and only if the restriction $f_\sigma : \Delta^n_\sigma \to X$ is smooth for all $\sigma \in \Sigma_n$, where $\Delta^n_\sigma \subset \mathbb{R}^n$ has the subspace smooth structure.*

Proof. Suppose $f : \square^n_\Delta \to X$ is smooth. Suppose $\phi : C \to \Delta^n_\sigma$ is a plot and let $\iota_\sigma : \Delta^n_\sigma \hookrightarrow \square^n$ be the inclusion. Note that ι_σ is smooth and thus the restriction of f to Δ^n_σ, defined as $f \circ \iota_\sigma : \Delta^n_\sigma \to X$, is a composition of smooth functions and therefore smooth.

Next, suppose that $f : \square^n_\Delta \to X$ is a map such that the restrictions f_σ are smooth for all σ. In order to show that f is smooth, we must show that $f \circ \phi : C \to X$ is a plot for every plot $\phi : C \to \square^n_\Delta$. Indeed, by definition of a plot for $\square^n_\Delta$, there exists an open cover $\{C_i\}$ of C, along with plots $\phi_i : C_i \to \Delta^n_{\sigma_i}$. Since the restrictions f_σ are smooth, each composition $f_{\sigma_i} \circ \phi_i : C_i \to X$ is a plot for X. Note that the plots $f_{\sigma_i} \circ \phi_i$ cover the map $f \circ \phi$. By the covering condition for differentiable spaces, $f \circ \phi$ is also a plot. □

An immediate corollary of this lemma is the following.

Corollary 5.1.1. *The coface maps δ^ϵ_i, the codegeneracy maps σ_i and the coconnection maps γ_i are all smooth when the unit n-cubes are equipped with the simplicial smooth structure.*

This allows us to define the smooth geometric realization in the same manner as the topological geometric realization.

Definition 5.1.11. The *smooth geometric realization functor* $|\cdot|_\infty : \mathsf{Set}^{\square_c^{\mathrm{op}}} \to \mathsf{Diff}$ is defined as follows. For $K_\bullet \in \mathsf{Set}^{\square_c^{\mathrm{op}}}$, the smooth geometric realization of $K_\bullet$ is

$$|K_\bullet|_\infty \coloneqq \coprod_{n=0}^{\infty} K_n \times \square^n_\Delta \Big/ \sim, \tag{5.1.42}$$

where the equivalence relation $\sim$ is defined in Equation (5.1.30), which is taken over all morphisms in the box category $\square$. The product, coproduct and quotient are taken as differentiable spaces.

As a topological space, the smooth geometric realization can be interpreted in the same manner as the topological geometric realization: as a CW-complex with an n-cell for each n-cube in $K_\bullet$. However, the smooth geometric realization gains the structure of a differentiable space. In this setting, the evaluation maps from Definition 5.1.8 are defined by the composition of an inclusion and quotient map, both of which are smooth. This shows that the evaluation maps themselves are smooth.

Lemma 5.1.4. *Let $K_\bullet \in \mathsf{Set}^{\square^{\mathrm{op}}+c}$ and $a \in K_n$. The evaluation map $\eta[a] : \square^n_\Delta \to |K_\bullet|_\infty$ is smooth.*

Remark 5.1.4. The smooth geometric realization for a cubical set with connection, $K_\bullet \in \mathsf{Set}^{\square_c^{\mathrm{op}}}$, can be expressed as a colimit in the same manner as the ordinary geometric realization. The box category with connection $\square_c$ can be viewed as a category internal to Diff since the coface, codegeneracy, and coconnection maps are smooth, and thus $\square_c$ is a cocubical object in Diff. Furthermore, we treat $K_\bullet$ as a cubical object in Diff by equipping each set K_n with the discrete smooth structure. Then, $|K_\bullet|_\infty$ can be defined as the coequalizer in Equation (5.1.37), viewed as a colimit in Diff. Because Diff is cocomplete from Proposition 2.2.2, the colimit exists.

Remark 5.1.5. The unit n-cube with the simplicial smooth structure defined in Definition 5.1.10 can also be defined as the smooth geometric realization of the unit n-cube as a simplicial set.

Smooth maps from $|K|_\infty$ can be characterized in the same manner as smooth maps from $\square^n_\Delta$, where the proof follows similarly.

Proposition 5.1.2. *Let $K_\bullet \in \mathsf{Set}^{\square_c^{\mathrm{op}}}$ be a cubical set with connection and X be a differentiable space. A map $f : |K|_\infty \to X$ is smooth if and only if for each nondegenerate cube $b \in K_n$, and each $\sigma \in \Sigma_n$, the restriction to the subsimplex Δ^n_σ of the geometric n-cube corresponding to b, defined as*

$$f_{b,\sigma} : \Delta^n_\sigma \overset{\iota_\sigma}{\hookrightarrow} \square^n_\Delta \overset{\eta[b]}{\longhookrightarrow} |K|_\infty \overset{f}{\to} X, \tag{5.1.43}$$

is smooth.

Proof. First, suppose that $f : |K|_\infty \to X$ is smooth. Note that $\iota_\sigma : \Delta^n_\sigma \hookrightarrow \square^n_\Delta$ is a plot. Because the evaluation map $\eta[b]$ is smooth, $f_{b,\sigma}$ is a composition of smooth maps, and is therefore also smooth.

Next, suppose that $f_{b,\sigma} : \Delta^n_\sigma \to X$ is smooth for all degenerate cubes $b \in K_n$ and $\sigma \in \Sigma_n$. Let $\phi : C \to |K|_\infty$ be a plot. By definition of the quotient smooth structure on

$|K|_\infty$ with respect to face, degeneracy and connection maps, there exists an open cover $\{C_i\}$ of C, along with plots $\phi_i : C_i \to \square^{n_i}_\Delta$, which are plots for the geometric n_i-cubes $\square^{n_i}_\Delta$ corresponding to non-degenerate cubes $b_i \in K_{n_i}$. These plots satisfy the following commutative diagram, where $\iota_i : C_i \to C$ are the inclusions.

$$\begin{array}{ccc} C_i & \xrightarrow{\phi_i} & \square^{n_i}_\Delta \\ {\scriptstyle \iota_i}\downarrow & & \downarrow{\scriptstyle \eta[b_i]} \\ C & \xrightarrow[\phi]{} & |K|_\infty, \end{array}$$

By the simplicial smooth structure of $\square^{n_i}_\Delta$, there exists an open cover $\{C_{i,j}\}$ of each C_i, along with plots $\phi_{i,j} : C_{i,j} \to \Delta^{n_i}_{\sigma_{i,j}}$, where $\sigma \in \Sigma_{n_i}$. These satisfy the following commutative diagram

$$\begin{array}{ccc} C_{i,j} & \xrightarrow{\phi_{i,j}} & \Delta^{n_i}_{\sigma_{i,j}} \\ {\scriptstyle \iota_i}\downarrow & & \downarrow{\scriptstyle \iota_{\sigma_{i,j}}} \\ C_i & \xrightarrow[\phi_i]{} & \square^{n_i}_\Delta, \end{array}$$

By composing this diagram with the first diagram above, we find that $\{\phi_{i,j} : C_{i,j} \to \Delta^{n_i}_{\sigma_{i,j}}\}$ forms a cover of $\phi : C \to |K|_\infty$. Then, since each $f_{b,\sigma}$ is smooth, the composition $f_{b_i,\sigma_{i,j}} \circ \phi_{i,j} : C_{i,j} \to X$ is a plot, which implies that $f \circ \phi : C \to X$ is a plot. Thus, $f : |K|_\infty \to X$ is smooth. □

Remark 5.1.6. Here we consider an alternate way to frame this result. In the above proof, we used the fact that the inclusion of a subsimplex $\iota_\sigma : \Delta^n_\sigma \to \square^n_\Delta$ is a plot and the evaluation map $\eta[b] : \square^n_\Delta \to |K|_\infty$ are smooth, for any nondegenerate n-cube $b \in K_n$ and any $\sigma \in \Sigma_n$. This implies that the composition

$$\phi_{b,\sigma} : \Delta^n_\sigma \to |K|_\infty \tag{5.1.44}$$

is a plot. Suppose $f : |K|_\infty \to X$ is a plot. Rather than checking that $f \circ \phi : C \to X$ is a plot for every plot $\phi : C \to |K|_\infty$, the above proposition implies that we only need to check that $f \circ \phi_{b,\sigma} : \Delta^n_\sigma \to X$ are plots for all nondegenerate $b \in K_n$ and $\sigma \in \Sigma_n$. We will use the same perspective to define differential forms on $|K|_\infty$ in Proposition 5.2.3.

5.2 THE CUBICAL CHEN MAPPING SPACE CONSTRUCTION

In this section, we describe the cubical Chen construction to obtain a cochain algebra for mapping spaces. This section will require some familiarity with simplicial homotopy theory, as we adapt simplicial methods to the cubical context. The full construction will not be used in this thesis, and we provide a discussion of the 0-cochains in this model in the following section.

As described in the introduction to this chapter, the construction of a cochain algebra for mapping spaces is done in three steps:

1. Given a cubical set $K_\bullet \in \mathsf{Set}^{\square_c^{\mathrm{op}}}$ and a differentiable space $Y \in \mathsf{Diff}$, build a cocubical space model $\mathsf{Diff}^\bullet(K, Y)$ of the mapping space such that $\|\mathsf{Diff}^\bullet(K, Y)\|_\infty \cong \mathsf{Diff}(|K|, Y)$, where $\|\cdot\|_\infty$ denotes the smooth totalization of a smooth cocubical space;

2. Given a smooth cocubical space $Z^\bullet \in \mathsf{Diff}^{\square_c}$, and a cochain algebra functor for smooth spaces $C^\bullet : \mathsf{Diff}^{\mathrm{op}} \to \mathsf{dga}$ construct a cochain algebra $T^\bullet(Z)$ for the cocubical space;

3. Using the de Rham cochain algebra functor $\Omega^\bullet : \mathsf{Diff}^{\mathrm{op}} \to \mathsf{dga}$ and the cocubical mapping space $\mathsf{Diff}^\bullet(K, X)$, show that we have a map of differential graded algebras

$$\mathcal{IT} : T^\bullet(\mathsf{Diff}(K, X)) \to \Omega^\bullet(\mathsf{Diff}(|K|_\infty, X)).$$

As we will be building a cochain algebra for mapping spaces using differential forms, our constructions will be performed in the category of differentiable spaces Diff. Using the simplicial smooth structure on the unit n-cube $\square^n_\Delta$, the constructions for Top from [PT03] can be directly generalized to category of differentiable spaces Diff.

Notation: From this section onwards, cubical sets will refer to cubical sets with connection unless otherwise specified. Furthermore, we will always assume that the unit n-cube $\square^n$ is equipped with the simplicial smooth structure from Definition 5.1.10. Therefore, we will supress the Δ from the notation, and denote this differentiable space as $\square^n := \square^n_\Delta$.

5.2.1 *Cocubical Mapping Spaces*

In this subsection, we fix $K_\bullet \in \mathsf{Set}^{\square_c^{\mathrm{op}}}$ to be a cubical set and $Y \in \mathsf{Diff}$ to be a smooth manifold (though the construction continues to hold for any differentiable space $Y \in \mathsf{Diff}$), and we define a smooth cocubical mapping space $\mathsf{Diff}^\bullet(K, Y)$ and show there is a smooth isomorphism

$$\|\mathsf{Diff}^\bullet(K, Y)\|_\infty \cong \mathsf{Diff}(|K|_\infty, Y).$$

We begin with the definition of the smooth totalization of a smooth cocubical space.

Definition 5.2.1. The *smooth totalization functor* $\|\cdot\| : \mathsf{Diff}^{\square_c} \to \mathsf{Diff}$ is defined as follows. For a smooth cocubical space $Z^\bullet \in \mathsf{Diff}^{\square_c}$, the totalization is a subspace of the sequences of maps

$$\|Z^\bullet\|_\infty \subset \prod_n \mathsf{Diff}(\square^n, Z^n) \tag{5.2.1}$$

where $\square^n \subset \mathbb{R}^n$ is equipped with the simplicial smooth structure and $(f_n : \square^n \to Z^n)_{n=0}^\infty \in \|Z^\bullet\|_\infty$ if

$$\begin{array}{ccc} \square^n & \xrightarrow{f_n} & Z^n \\ {\scriptstyle\alpha}\downarrow & & \downarrow{\scriptstyle Z(\alpha)} \\ \square^m & \xrightarrow{f_m} & Z^m \end{array} \tag{5.2.2}$$

commutes for all morphisms $\alpha : \square^n \to \square^m$ in the box category with connection $\square_c$.

Remark 5.2.1. As with smooth geometric realization, we can reframe totalization as a limit as

$$\|Z^\bullet\|_\infty := \mathsf{eq}\left[\prod_n \mathsf{Diff}(\square^n, Z^n) \rightrightarrows \prod_{\alpha:\square^n\to\square^m} \mathsf{Diff}(\square^n, Z^m)\right]. \tag{5.2.3}$$

The second product in the equalizer runs through all maps $\alpha : \square^n \to \square^m$ in $\square_c$. The two maps

$$\phi, \psi : \prod_n \mathsf{Diff}(\square^n, Z^n) \to \prod_{\alpha:\square^n\to\square^m} \mathsf{Diff}(\square^n, Z^m)$$

in the equalizer are defined as:

$$\begin{aligned} \phi\left((f_n : \square^n \to Z^n)_{n=0}^\infty\right) &:= (Z(\alpha) \circ f_n : \square^n \to Z^m)_{\alpha\in\square_c} \\ \psi\left((f_n : \square^n \to Z^n)_{n=0}^\infty\right) &:= (f_m \circ \alpha : \square^n \to Z^m)_{\alpha\in\square_c} \end{aligned}$$

Both of these maps are smooth (and thus morphisms in Diff) since $\square^n$ is equipped with the simplicial smooth structure. Thus, this is a limit in the category of differentiable spaces Diff, and since Diff is complete from Proposition 2.2.2, the totalization exists.

While the definition of totalization is less transparent than that of geometric realization, this construction is well-suited to the modelling of mapping spaces.

Definition 5.2.2. Let $K_\bullet \in \mathsf{Set}^{\square_c^{\mathrm{op}}}$ and $Y \in \mathsf{Diff}$. The *smooth cocubical mapping space* $\mathsf{Diff}^\bullet(K, Y) : \square \to \mathsf{Diff}$ is defined by $\mathsf{Diff}^n(K, Y) = \mathsf{Diff}(K_n, Y)$, where K_n is given the discrete smooth structure. All of the cocubical maps are induced by those in $K_\bullet$.

In order to understand how this construction corresponds to the mapping space, we explicitly write out its totalization. First, we consider a sequence of maps

$$(f_n : \square^n \to \mathsf{Diff}(K_n, Y))_{n=0}^\infty \in \prod_n \mathsf{Diff}(\square^n, \mathsf{Diff}(K_n, Y)).$$

By using the product-exponential adjunction in Diff from Proposition 2.2.1, we can rewrite this sequence as

$$(\tilde{f}_n : K_n \times \square^n \to Y)_{n=0}^\infty \in \prod_n \mathsf{Diff}(K_n \times \square^n, Y).$$

Thus, $\tilde{f}_n$ can be viewed as a map from each geometric n-cube in the realization of $K_\bullet$ into Y. Furthermore, we may combine this into a map

$$\tilde{f} : \coprod_n K_n \times \square^n \to Y.$$

Rewriting the compatibility condition for totalization in Equation (5.2.2) in terms of $\tilde{f}$, we get

$$\tilde{f}(K(\alpha)(x), t) = \tilde{f}(x, \alpha(t)), \quad x \in K_m, \quad t \in \square^n,$$

for all morphisms $\alpha : \square^n \to \square^m$ in $\square$. This implies that the map $\tilde{f}$ respects the equivalence relation specified by the geometric realization of $K_\bullet$. Thus, the map $\tilde{f}$ is well defined under the quotient, and we obtain

$$\tilde{f} : |K|_\infty \to Y.$$

This shows that there is a bijection $\|\mathsf{Diff}^\bullet(K, Y)\|_\infty \cong \mathsf{Diff}(|K|_\infty, Y)$. In fact, this is smooth isomorphism, which we show using category theoretic methods.

Proposition 5.2.1. *Let $K_\bullet \in \mathsf{Set}^{\square_c^{\mathrm{op}}}$ and $Y \in \mathsf{Diff}$. Then, there exists smooth isomorphism (an isomorphism in* Diff*)*

$$\|\mathsf{Diff}^\bullet(K, Y)\|_\infty \cong \mathsf{Diff}(|K|_\infty, Y). \tag{5.2.4}$$

Proof. The following category-theoretic proof mimics the discussion above. We use the fact that we are working with a cartesian closed category of differentiable spaces Diff, given by Proposition 2.2.1. In particular, $\mathsf{Diff}(-,-)$ specifies the internal hom, which preserves

limits in both variables. We treat the sets K_n as differentiable spaces equipped with the discrete smooth structure. Beginning with the equalizer definition of totalization and using the product-hom adjunction, we get

$$\begin{aligned}
\|\mathsf{Diff}^\bullet(K,Y)\|_\infty &= \mathsf{eq}\left[\prod_n \mathsf{Diff}(\square^n, \mathsf{Diff}(K_n, Y)) \rightrightarrows \prod_{\alpha:\square^n\to\square^m} \mathsf{Diff}(\square^n, \mathsf{Top}(K_m, Y))\right] \\
&\cong \mathsf{eq}\left[\prod_n \mathsf{Diff}(K_n \times \square^n, Y) \rightrightarrows \prod_{\alpha:\square^n\to\square^m} \mathsf{Diff}(K_m \times \square^n, Y)\right] \\
&\cong \mathsf{eq}\left[\mathsf{Diff}\left(\coprod_n UK_n \times \square^n, Y\right) \rightrightarrows \mathsf{Diff}\left(\coprod_{\alpha:\square^n\to\square^m} K_m \times \square^n, Y\right)\right] \\
&\cong \mathsf{Diff}\left(\mathsf{coeq}\left[\coprod_{\alpha:\square^n\to\square^m} K_m \times \square^n \rightrightarrows \coprod_n K_n \times \square^n\right], Y\right) \\
&= \mathsf{Diff}(|K|_\infty, Y).
\end{aligned}$$

□

5.2.2 *Cochain Algebras for Cocubical Mapping Spaces*

Now that we are equipped with a cocubical model for mapping spaces, $\mathsf{Diff}^\bullet(K,Y)$, we introduce a construction that associates a cochain algebra to $\mathsf{Diff}^\bullet(K,Y)$. This section is a reformulation of the simplcial construction discussed in [BS77; PT03] adapted to cubical sets and differentiable spaces.

We introduce the general construction using arbitrary cochains and smooth cocubical spaces. Let $Z^\bullet \in \mathsf{Diff}^{\square_c}$ be a smooth cocubical space, and let $C^\bullet : \mathsf{Diff}^{\mathrm{op}} \to \mathsf{Ch}$ be a cochain functor into Ch, the category of nonnegatively graded cochain complexes. By applying the cochain functor $C^\bullet$ to the cocubical space $Z^\bullet$, we obtain a cubical cochain complex $C^\bullet(Z^\bullet) \in \mathsf{Ch}^{\square_c^{\mathrm{op}}}$. Specifically, at each degree, we have

$$\underline{p} \mapsto C^\bullet(Z^p),$$

and the face, degeneracy and connection maps are induced by the coface and codegeneracy maps in $Z^\bullet$ since $C^\bullet$ is a contravariant functor,

$$\begin{aligned}
\partial_i^\epsilon &= C^\bullet(\delta_i^\epsilon) : C^\bullet(Z^p) \to C^\bullet(Z^{p-1}), \\
s_i &= C^\bullet(\sigma_i) : C^\bullet(Z^p) \to C^\bullet(Z^{p+1}), \\
g_i &= C^\bullet(\gamma_i) : C^\bullet(Z^p) \to C^\bullet(Z^{p+1}).
\end{aligned}$$

We may also view this as a double complex

$$\begin{array}{ccccccc} \vdots & & \vdots & & \vdots & & \\ \uparrow d & & \uparrow d & & \uparrow d & & \\ C^1(Z^0) & \xleftarrow{\partial} & C^1(Z^1) & \xleftarrow{\partial} & C^1(Z^2) & \longleftarrow & \cdots \\ \uparrow d & & \uparrow d & & \uparrow d & & \\ C^0(Z^0) & \xleftarrow{\partial} & C^0(Z^1) & \xleftarrow{\partial} & C^0(Z^2) & \longleftarrow & \cdots \end{array} \tag{5.2.5}$$

where the vertical differential, d, is the coboundary map from the cochain functor and the horizontal differential is the alternating sum of face maps

$$\partial = \sum_{i=1}^{m} (-1)^i \left(\partial_i^0 - \partial_i^1 \right) : C^k(Z^m) \to C^k(Z^{m-1}). \tag{5.2.6}$$

By taking the total complex, we obtain a cochain complex associated to $Z^\bullet$.

Definition 5.2.3. The *normalized bicomplex* of the cubical cochain complex $C^\bullet(Z^\bullet)$ is the bicomplex shown in Equation (5.2.5), where the cochains $C^k(Z^m)$ are replaced with the *normalized cochains*,

$$\overline{C}^k(Z^p) := C^k(Z^p)/C^k_{\text{degen}}(Z^p), \tag{5.2.7}$$

where $C^k_{\text{degen}}(Z^p)$ is generated by the image of the degeneracy and connection maps of $C^\bullet(Z^\bullet)$.

The *realization* of $C^\bullet(Z^\bullet)$, denoted $T^\bullet(Z)$, is the total complex of the normalized bicomplex defined by

$$T^r(Z) := \bigoplus_{q-p=r} \overline{C}^q(Z^p). \tag{5.2.8}$$

The differential of $T^\bullet(Z)$, denoted by $D : T^\bullet(Z) \to T^{\bullet+1}(Z)$ is defined for $\alpha \in \overline{C}^q(Z^p)$ by

$$D(\alpha) := (-1)^p (d - \partial)(\alpha). \tag{5.2.9}$$

Lemma 5.2.1. *The realization $T^\bullet(Z)$ of a smooth cocubical space $Z^\bullet \in \mathsf{Diff}^\square$ is a cochain complex.*

Proof. In order to show that $D^2 = 0$, it suffices to prove that $d\partial - \partial d = 0$. We note that the face maps $\partial_i^\epsilon : C^\bullet(Z^p) \to C^\bullet(Z^p)$ are cochain maps, and thus they commute with the cochain differential d. Because ∂ is a linear combination of face maps, it also commutes with d. □

If the underlying cochain functor is equipped with a multiplicative structure so that $C^\bullet : \mathsf{Diff}^{\mathrm{op}} \to \mathsf{dga}$ sends a differentiable space to a differential graded algebra, the realization will also be a differential graded algebra.

Proposition 5.2.2. *Suppose $Z^\bullet \in \mathsf{Diff}^{\square_c}$ is a smooth cocubical space and $C^\bullet : \mathsf{Diff}^{\mathrm{op}} \to \mathsf{dga}$ is a cochain algebra functor. Then, the realization of $C^\bullet(Z^\bullet)$ defined in Equation (5.2.8), $T^\bullet(Z)$, is a cochain algebra.*

Proof. First, we will define a product on the unnormalized bicomplex $C^\bullet(Z^\bullet)$. For each pair $p_1, p_2 \in \mathbb{N}$, we have the map

$$\begin{aligned}\theta^{p_1,p_2} = (\theta_1^{p_1,p_2}, \theta_2^{p_1,p_2}) : \square^{p_1+p_2} &\to \square^{p_1} \times \square^{p_2} \\ (x_1, \dots, x_{p_1+p_2}) &\mapsto \big((x_1, \dots, x_{p_1}), (x_{p_1+1}, \dots, x_{p_1+p_2})\big).\end{aligned}$$

Within the cocubical space $Z^\bullet$, this provides maps

$$Z(\theta^{p_1,p_2}) = (Z(\theta_1^{p_1,p_2}), Z(\theta_2^{p_1,p_2})) : Z^{p_1+p_2} \to Z^{p_1} \times Z^{p_2}$$

Then, given $\sigma \in C^{k_1}(Z^{p_1})$ and $\tau \in C^{k_2}(Z^{p_2})$, the product $\sigma \cup \tau \in C^{k_1+k_2}(Z^{p_1+p_2})$ is given by

$$\sigma \cup \tau := Z(\theta_1^{p_1,p_2})^* \sigma \cup Z(\theta_2^{p_1,p_2})^* \tau, \tag{5.2.10}$$

where the $\cup$ product on the right hand side is the product from the cochain algebra $C^\bullet$.

Next, we must check that the product respects degeneracies and descends to a product in the normalized bicomplex $\overline{C}^\bullet(Z^\bullet)$. Suppose that $\sigma = Z(a)^*\sigma'$, where $\sigma' \in C^{k_1}(Z^{p_1-1})$ and $a : \square^{p_1} \to \square^{p_1-1}$ is either a codegeneracy or coconnection map in the box category $\square_c$. Then, we have

$$(Z(a)^*\sigma') \cup \tau = Z(\theta_1^{p_1,p_2} \circ a)^* \sigma' \cup Z(\theta_2^{p_1,p_2})^* \tau.$$

There exists another codegeneracy or connection map $a' : \square^{p_1+p_2} \to \square^{p_1+p_2-1}$ such that $\theta^{p_1-1,p_2} \circ a' = (a, 1) \circ \theta^{p_1,p_2}$. In particular, note that $\theta_2^{p_1-1,p_2} \circ a' = \theta_2^{p_1,p_2} : \square^{p_1+p_2} \to \square^{p_2}$ since the new codegeneracy only affects the first p_1 coordinates. Then, the above becomes

$$\begin{aligned}(Z(a)^*\sigma') \cup \tau &= Z(a')^*(\theta_1^{p_1-1,p_2})^*\sigma' \cup Z(a')^* Z(\theta_2^{p_1-1,p_2})^*\tau \\ &= Z(a')^* \left(Z(\theta_1^{p_1-1,p_2})^*\sigma' \cup Z(\theta_2^{p_1-1,p_2})^*\tau \right),\end{aligned}$$

where the final equality is given by the functoriality of $C^\bullet$. A similar computation can be done for the case where τ is degenerate. This shows that the cochain product is well defined on the normalized bicomplex, and thus $T^\bullet(Z)$ is a cochain algebra. □

5.2.3 *Chen Construction for Cocubical Mapping Spaces*

In this section, we will restrict ourselves to the primary objects of study: differential forms on mapping spaces $\mathsf{Diff}(|K|_\infty, X)$, where $K_\bullet \in \mathsf{Set}^{\square_c^{\mathrm{op}}}$ and X is a smooth manifold. The main goal of this section is to establish a morphism of differential graded algebras

$$\mathcal{IT} : T^\bullet(\mathsf{Diff}^\bullet(K, X)) \to \Omega^\bullet(\mathsf{Diff}(|K|_\infty, X)), \tag{5.2.11}$$

where the realization $T^\bullet$ is constructed using the de Rham complex $\Omega^\bullet : \mathsf{Diff}^{\mathrm{op}} \to \mathsf{dga}$ as the cochain algebra functor. Before we begin, we briefly recall the structure of differential forms on the mapping space $\mathsf{Diff}(|K|_\infty, X)$.

Recall from Definition 2.2.2 that a differential k-form on a differentiable space Y is a collection of forms $\omega_\phi \in \Omega^k(C)$ on each plot $\phi : C \to Y$ such that if $f : C' \to C$ is smooth, then $\omega_{\phi \circ f} = f^* \omega_\phi$. For the smooth geometric realization of a cubical set $|K|_\infty$, differential forms can be fully defined by specifying forms for the generating plots

$$\phi_{b,\sigma} : \Delta^n_\sigma \xrightarrow{\iota_\sigma} \square^n_\Delta \xrightarrow{\eta[b]} |K|_\infty,$$

for all non-degenerate $b \in K_n$ and $\sigma \in \Sigma_n$, which were previously discussed in Remark 5.1.6.

Proposition 5.2.3. *Let $K_\bullet \in \mathsf{Set}^{\square_c^{\mathrm{op}}}$. A differential k-form on $|K|_\infty$ is a collection of differential forms $\omega_{b,\sigma} \in \Omega^k(\Delta^n_\sigma)$ for each non-degenerate cube $b \in K_n$ and $\sigma \in \Sigma_n$ which agree on intersections. In other words, if $b_1, b_2 \in K_n$ are nondegenerate cubes which share a boundary $c \in K_{n-1}$, then for any $\sigma_1, \sigma_2 \in \Sigma_n$ and $\sigma \in \Sigma_{n-1}$ such that*

$$\mathrm{im}(\phi_{c,\sigma}) \subset \mathrm{im}(\phi_{b_1,\sigma_1}) \cap \mathrm{im}(\phi_{b_2,\sigma_2}),$$

then

$$\omega_{c,\sigma} = f_1^* \omega_{b_1,\sigma_1} = f_2^* \omega(b_2, \sigma_2),$$

where $f_1 : \Delta^{n-1}_\sigma \to \Delta^n_{\sigma_1}$ and $f_2 : \Delta^{n-1}_\sigma \to \Delta^n_{\sigma_2}$ are the inclusions.

Proof. We must show that given a collection of forms $\omega_{b,\sigma} \in \Omega^k(\Delta^n_\sigma)$ as above, we obtain differential forms $\omega_\phi \in \Omega^k(C)$ for each plot $\phi : C \to |K|_\infty$. By the same argument as the

proof of Proposition 5.1.2, there exists an open cover $\{C_{i,j}\}$ of C with plots $\phi_{i,j} : C_{i,j} \to \Delta^{n_i}_{\sigma_{i,j}}$ such that

$$\begin{array}{ccc} C_{i,j} & \xrightarrow{\phi_{i,j}} & \Delta^{n_i}_{\sigma_{i,j}} \\ {\scriptstyle \iota_{i,j}}\downarrow & & \downarrow{\scriptstyle \phi_{b_i,\sigma_{i,j}}} \\ C & \xrightarrow[\phi]{} & |K|_\infty \end{array}$$

commutes. Here, $b_i \in K_{n_i}$ are nondegenerate cubes, and $\sigma_{i,j} \in \Sigma_{n_i}$. Now, we can define forms $\omega_{i,j} \in \Omega^k(C_{i,j})$ via the pullback

$$\omega_{i,j} = \phi^*_{i,j}\omega_{b_i,\sigma_{i,j}}.$$

Because the forms $\omega_{b,\sigma} \in \Omega^k(\Delta^n_\sigma)$ agree on intersections, and the sets $C_{i,j}$ form an open cover of C, there exists a unique form $\omega_\phi \in \Omega^k(C)$ such that $\iota^*_{i,j}\omega_\phi = \omega_{i,j}$. Thus, ω_ϕ is the desired differential form for the plot ϕ. □

We will now define the map

$$\mathcal{IT} : T^\bullet(\mathsf{Diff}(K, X)) \to \Omega^\bullet(\mathsf{Diff}(|K|_\infty, X))$$

which associates a differential form on the mapping space $\mathsf{Diff}(|K|_\infty, X)$ to each element of $T^\bullet(\mathsf{Diff}(K, X))$, where

$$T^r(\mathsf{Diff}(|K|_\infty, X)) = \bigoplus_{q-p=r} \overline{\Omega}^q(\mathsf{Diff}(K_p, X)).$$

Because K_p is equipped with the discrete smooth structure, we have

$$\mathsf{Diff}(K_p, X) \cong X^{K_p},$$

where X^{K_p} is a finite Cartesian product of X, since we have assumed $K_\bullet$ to be a finite cubical set. We begin with the definition for the unnormalized cochain complex, and we show that it remains well defined in the normalized complex in Lemma 5.2.2. Fix a q-form $\omega \in \Omega^q(X^{K_p})$, where $q - p = r$, and we define the associated differential form.

By the totalization of cocubical spaces, we have a natural evaluation map

$$\begin{aligned} \mathrm{ev}_p : \square^p \times \|\mathsf{Diff}^\bullet(K, X)\|_\infty &\to X^{K_p} \\ \left(\mathbf{t}, (f_n : \square^n \to \mathsf{Diff}(K_n, X))_{n=0}^\infty\right) &\mapsto f_p(\mathbf{t}) \end{aligned} \tag{5.2.12}$$

for each p. The pullback along this evaluation map provides a map

$$\mathrm{ev}_p^* : \Omega^q(X^{K_p}) \to \Omega^q(\square^p \times \mathsf{Diff}(|K|_\infty, X)).$$

In order to get a differential form on $\mathsf{Diff}(|K|_\infty, X)$, we integrate over $\square^p$ denoted by

$$\int_{\square^p} : \Omega^q(\square^p \times \mathsf{Diff}(|K|_\infty, X)) \to \Omega^{q-p}(\mathsf{Diff}(|K|_\infty, X)). \tag{5.2.13}$$

Note that the resulting differential form on $\mathsf{Diff}(|K|_\infty, X)$ is of degree $q - p = r$, which coincides with the degree of the element $\omega \in \overline{\Omega}^q(\mathsf{Diff}(K_p, X)) \subset T^r(\mathsf{Diff}(K, X))$. The following result shows that this construction is well defined for the normalized cochains.

Lemma 5.2.2. *Suppose $\omega \in \Omega^q(X^{K_p})$ is a degenerate element such that there exists some $\nu \in \Omega^q(X^{K_{p-1}})$ such that $\omega = u^*(\nu)$, where $u : X^{K_p} \to X^{K_{p-1}}$ is induced by either a degeneracy or connection map in $K_\bullet$ corresponding to the codegeneracy/coconnection map $\mu : \square^p \to \square^{p-1}$. Then,*

$$\int_{\square^p} \mathrm{ev}_p^* \omega = 0.$$

Proof. From the definitions of the codegeneracy/coconnection map μ and u, as well as the definition of the evaluation map ev_p, the following diagram commutes

$$\begin{array}{ccc} \square^p \times \mathsf{Diff}(|K|_\infty, X) & \xrightarrow{\mathrm{ev}_p} & X^{K_p} \\ \downarrow{\scriptstyle (\mu, \mathrm{id})} & & \downarrow{\scriptstyle u} \\ \square^{p-1} \times \mathsf{Diff}(|K|_\infty, X) & \xrightarrow[\mathrm{ev}_{p-1}]{} & X^{K_{p-1}}. \end{array}$$

Then, the integrand can be expressed as

$$\begin{aligned} \mathrm{ev}_p^* \omega &= \mathrm{ev}_p^* u^* \nu \\ &= (\mu, \mathrm{id})^* \mathrm{ev}_{p-1}^* \nu. \end{aligned}$$

Because the pullbacks factor through $\Omega^\bullet(\square^{p-1} \times \mathsf{Diff}(|K|_\infty, X)$, there does not exist a p-volume form component on $\square^p$. Therefore, the resulting integral is zero. □

The map from the cochain algebra of the cocubical mapping space to the de Rham complex of the mapping space can be summarized as follows.

Definition 5.2.4. Let $K_\bullet \in \mathsf{Set}^{\square_c^{\mathrm{op}}}$ and $X \in \mathsf{Diff}$. The *integration map*

$$\mathcal{IT} : T^\bullet(\mathsf{Diff}(K, X)) \to \Omega^\bullet(\mathsf{Diff}(|K|_\infty, X))$$

is defined by

$$\mathcal{IT}(\omega) := \int_{\square^p} \mathrm{ev}_p^* \omega,$$

where $\omega \in \overline{\Omega}^q(X^{K_p}) \subset T^{q-p}(\mathsf{Diff}(K, X))$. The image of the integration map is called the *Chen complex of* $\mathsf{Diff}(|K|_\infty, X)$, and is denoted by $\mathrm{Chen}^\bullet(\mathsf{Diff}(|K|_\infty, X))$.

While the integration map is defined for arbitrary elements in $\overline{\Omega}^q(X^{K_p})$, the construction can be made slightly more concrete for differential forms which can be factored into a product of forms on each individual factor of X in X^{K_p}. This coincides with our definition of the cochain model for path spaces in Section 2.3.

In this case, we choose a total order on the set of p-cubes K_p, and denote the cubes by $\{a_i\}_{i=1}^{N_p}$, where N_p is the number of p-cubes. Let $\pi_m : X^{K_p} \to X$ be the projection onto the m^{th} factor. This induces an inclusion

$$\begin{aligned} \rho_p : \Omega^\bullet(X)^{\otimes K_p} &\to \Omega^\bullet(X^{K_p}) \\ (\omega_1, \ldots, \omega_{N_p}) &\mapsto \pi_1^* \omega_1 \wedge \ldots \wedge \pi_{N_p}^* \omega_{N_p}. \end{aligned} \tag{5.2.14}$$

Next, the evaluation map in Equation (5.2.12) can be defined in terms of the cubical evaluation maps arising from the geometric realization of $K_\bullet$ in Definition 5.1.8. Using the isomorphism in Proposition 5.2.1, we can equivalently define the evaluation map ev_p as

$$\begin{aligned} \mathrm{ev}_p : \square^p \times \mathsf{Diff}(|K|_\infty, X) &\to X^{K_p} \\ (\mathbf{t}, f) &\mapsto (f \circ \eta[a](\mathbf{t}))_{a \in K_p}. \end{aligned} \tag{5.2.15}$$

Because differential forms are defined plotwise, we fix a plot $\tilde{\phi} : C \to \mathsf{Diff}(|K|_\infty, X)$ in order to make the construction more explicit. Recall from the definition of the mapping space smooth structure in Example 2.2.6 that $\tilde{\phi}$ is a plot if the adjoint map $\phi : C \times |K|_\infty \to X$ is a plot. The evaluation map with respect to the plot $\tilde{\phi}$ is

$$\begin{aligned} \mathrm{ev}_{p,\tilde{\phi}} : C \times \square^p &\to X^{K_p} \\ (\mathbf{x}, \mathbf{t}) &\mapsto \big(\phi(\mathbf{x}, \eta[a](\mathbf{t}))\big)_{a \in K_p} \end{aligned} \tag{5.2.16}$$

The pullback of the composition of the inclusion from the tensor product in Equation (5.2.14), the evaluation map Equation (5.2.15) and the plot $\tilde{\phi}$ can be written as

$$\mathrm{ev}_{p,\tilde{\phi}} \circ \rho_p(\omega_1 \otimes \ldots \otimes \omega_{N_p}) = \bigwedge_{j=1}^{N_p} (\phi \circ \eta[a_j])^* \omega_j. \tag{5.2.17}$$

Thus, for these factored differential forms, we can define the integration map from Definition 5.2.4 as

$$\begin{aligned} \mathcal{IT}_p : \Omega^\bullet(X)^{\otimes K_p} &\to \Omega^\bullet(\mathsf{Diff}(|K|_\infty, X)) \\ \omega_1 \otimes \ldots \otimes \omega_{N_p} &\mapsto \left(\int_{\square^p} \bigwedge_{j=1}^{N_p} (\phi \circ (\mathrm{id}, \eta[a_j]))^* \omega_j \right)_{\tilde{\phi}}, \end{aligned} \tag{5.2.18}$$

where on the right, we use the subscript $\tilde{\phi}$ to denote the differential form with respect to the plot $\tilde{\phi}$.

Returning to the setting of arbitrary elements in $T^\bullet(\mathsf{Diff}(K, X))$, the previous discussion provides a definition of the integration map as a degree 0 linear map. In the remainder of this section, we will show that $\mathcal{IT}$ is a map of differential graded algebras. First, we show that it respects the differential.

Proposition 5.2.4. *The integration map $\mathcal{IT} : T^\bullet(\mathsf{Diff}(K, X)) \to \Omega^\bullet(\mathsf{Diff}(|K|_\infty, X))$ is a cochain map.*

Proof. The proof of this relies on Stokes' theorem for fiberwise integration (also used in both [BS77] and [GTZ10]), which states that for a differential form $\nu \in \Omega^\bullet(\square^p \times C)$, then

$$d \int_{\square^p} \nu = (-1)^p \left(\int_{\square^p} d\nu - \int_{\partial \square^p} d\nu \right). \tag{5.2.19}$$

Because the exterior derivative of a differential form on a differentiable space is defined plotwise, we fix a plot $\tilde{\phi} : C \to \mathsf{Diff}(|K|_\infty, X)$. Given $\omega \in \overline{\Omega}^q(X^{K_p})$, the fibrewise Stokes' theorem shows that

$$\begin{aligned} d\left(\mathcal{IT}(\omega)\right)_{\tilde{\phi}} &= \left(\mathcal{IT}(\omega)\right)_{\tilde{\phi}} \\ &= \int_{\square^p} \mathrm{ev}^*_{p,\tilde{\phi}} \omega \\ &= (-1)^p \left(\int_{\square^p} d(\mathrm{ev}^*_{p,\tilde{\phi}} \omega) - \int_{\partial \square^p} \mathrm{ev}^*_{p,\tilde{\phi}} \omega \right) \\ &= (-1)^p \left(\int_{\square^p} \mathrm{ev}^*_{p,\tilde{\phi}} d\omega - \int_{\square^{p-1}} \mathrm{ev}^*_{p-1,\tilde{\phi}} \left(\sum_{i=1}^{p} (-1)^i (\partial^0_i - \partial^1_i) \omega \right) \right) \\ &= \left(\mathcal{IT}(D\omega)\right)_{\tilde{\phi}}. \end{aligned}$$

□

Finally, we will show that the integration map is a morphism of differential graded algebras.

Proposition 5.2.5. *The integration map $\mathcal{IT} : T^\bullet(\mathsf{Diff}(K,X)) \to \Omega^\bullet(\mathsf{Diff}(|K|_\infty, X))$ is morphism of differential graded algebras.*

Proof. The only part left to show is that the integration map preserves the product structure of the of $T^\bullet(\mathsf{Diff}(K,X))$. Recall from the proof of Proposition 5.2.2, specifically in Equation (5.2.10), that the product structure on $T^\bullet(\mathsf{Diff}(K,X))$ is given as follows. Suppose $\omega_1 \in \overline{\Omega}^{q_1}(X^{K_{p_1}})$ and $\omega_2 \in \overline{\Omega}^{q_2}(X^{K_{p_2}})$.

Let $\theta_1^{p_1,p_2} : \square^{p_1+p+2} \to \square^{p_1}$ and $\theta_2^{p_1,p_2} : \square^{p_1+p_2} \to \square^{p_2}$ be the projection maps onto the first p_1, and last p_2 coordinates respectively. Both of these maps are iterated codegeneracy maps, and are therefore morphisms of $\square_c$. These induce morphisms

$$\tilde{\theta}_1^{p_1+p_2} : \Omega^\bullet(X^{K_{p_1}}) \to \Omega^\bullet(X^{K_{p_1+p_2}}),$$
$$\tilde{\theta}_2^{p_1+p_2} : \Omega^\bullet(X^{K_{p_2}}) \to \Omega^\bullet(X^{K_{p_1+p_2}}),$$

and the product of ω_1 and ω_2 in $T^\bullet(\mathsf{Diff}(K,X))$ is given by

$$\omega_1 \wedge \omega_2 := \tilde{\theta}_1^{p_1+p_2}\omega_1 \wedge \tilde{\theta}_2^{p_1+p_2}\omega_2.$$

Then, the integration map applied to this product is

$$\mathcal{IT}(\omega_1 \wedge \omega_2) = \int_{\square^{p_1+p_2}} \mathrm{ev}^*_{p_1+p_2}\tilde{\theta}_1^{p_1+p_2}\omega_1 \wedge \mathrm{ev}^*_{p_1+p_2}\tilde{\theta}_2^{p_1+p_2}\omega_2.$$

The differential form

$$\mathrm{ev}^*_{p_1+p_2}\tilde{\theta}_1^{p_1+p_2}\omega_1 \in \Omega^{q_1}(\square^{p_1+p_2} \times \mathsf{Diff}(|K|_\infty, X))$$

depends only on the $(t_1, \ldots, t_{p_1})$ coordinates of $\square^{p_1+p_2}$ and

$$\mathrm{ev}^*_{p_1+p_2}\tilde{\theta}_1^{p_1+p_2}\omega_1 \in \Omega^{q_1}(\square^{p_1+p_2} \times \mathsf{Diff}(|K|_\infty, X))$$

depends only on the $(t_{p_1+1}, \ldots, t_{p_1+p_2})$ coordinates of $\square^{p_1+p_2}$. This implies that the integral above can be factored as

$$\begin{aligned}\mathcal{IT}(\omega_1 \wedge \omega_2) &= \left(\int_{\square^{p_1}} \mathrm{ev}^*_{p_1}\omega_1\right) \wedge \left(\int_{\square^{p_2}} \mathrm{ev}^*_{p_2}\omega_1\right) \\ &= \mathcal{IT}(\omega_1) \wedge \mathcal{IT}(\omega_2),\end{aligned}$$

which is the wedge product of differential forms in $\Omega^\bullet(\mathsf{Diff}(|K|_\infty, X))$. □

5.3 CUBICAL MONOMIALS FOR MAPPING SPACES

In this section, we discuss the 0-cochains in the cubical mapping space construction detailed in the previous section. From this point onwards, we will also call individual 0-cochains *signature monomials*, or simply *monomials*. This terminology comes from the analogy between the moment generating function and the expected path signature. Just as moments (expectations of monomials) are used to characterize measures on finite dimensional vector spaces, the expected path signature coefficients are used to characterize measures on path spaces. We reserve the term *signature* for a collection of monomials arranged into a single map, such as the full path signature $S : P\mathbb{R}^N \to T((\mathbb{R}^N))$.

We begin by providing the general definition of the cubical monomials from the mapping space model. Then, we refine this definition into the form which we will use in the next two chapters to define the path signature and the mapping space signature. We will conclude by considering the product structure on cubical monomials, which corresponds to the shuffle product of the ordinary path signature. This section can be read independently from the previous section on the topological construction and contains the foundational material which will be used for the rest of this thesis.

5.3.1 *Definition of Cubical Monomials*

Throughout this section, we fix a cubical set $K_\bullet \in \mathsf{Set}^{\square_c^{\mathrm{op}}}$ and a smooth manifold X. Recall that $\mathsf{Diff}(|K|_\infty, X)$ is the mapping space of smooth maps from the smooth geometric realization of $K_\bullet$ into X, which are characterized by Proposition 5.1.2.

We will introduce a class of monomials from the Chen construction obtained using factored differential forms; this coincides with the definition in Equation (5.2.18). The 0-cochains (monomials) of the Chen construction $\mathsf{Chen}^0(\mathsf{Diff}(|K|_\infty, X)$ have a grading, analogous to the grading by levels in the ordinary path signature. Fix a level $n \in \mathbb{N}$. Recall from Section 2.3 that a level n monomial in Chen's construction for path spaces is defined using a collection of n 1-forms on the underlying manifold X. For the mapping space construction, the input for a 0-cochain is a collection of m differential forms $\omega_i \in \Omega^{q_i}(X)$ paired with a collection of m n-cubes $a_i \in K_n$ such that the degrees of the differential forms satisfy

$$\sum_{j=1}^{m} q_i = n.$$

From Proposition 5.1.1, each n-cube $a_i \in K_n$ can be represented by a pair (I_j, b_j), where b_j is a non-degenerate q_j-cube and I_j is a q_j-ordered subset, defined in Definition 5.1.9.

Definition 5.3.1. Let $n \in \mathbb{N}$, and consider the ordered sets

$$\mathcal{I} = \Big((I_1, b_1), \dots, (I_m, b_m)\Big),$$
$$P = (\omega_1, \dots, \omega_m).$$

The ordered set $\mathcal{I}$ specifies the choice of n-cubes, and P specifies the associated differential forms. Each (I_j, b_j) specifies a n-cube, where b_j is a non-degenerate q_j-cube, and I_j is a q_j-ordered subset of $[n]$. Furthermore, $\omega_j \in \Omega^{q_j}(M)$ such that $\sum_{j=1}^m q_j = n$. Let $\gamma \in \mathsf{Diff}(|K|_\infty, X)$. The *cubical monomial of γ with respect to $\mathcal{I}$ and P* is

$$M^{\mathcal{I},P}(\gamma) = \int_{\square^n} \bigwedge_{j=1}^m \left(\eta[b_j] \circ \eta_n^{I_j}\right)^* (\gamma^* \omega_j) \tag{5.3.1}$$

$$= \sum_{\sigma \in \Sigma_n} \int_{\Delta^n_\sigma} \bigwedge_{j=1}^m \left(\eta[b_j] \circ \eta_n^{I_j}\right)\Big|^*_{\Delta^n_\sigma} (\gamma^* \omega_j). \tag{5.3.2}$$

The definition of the cubical monomial in Equation (5.3.1) uses the factorization of the evaluation map with respect to the n-cube $a_j = (I_j, b_j)$ given in Equation (5.1.36), while the second expression in Equation (5.3.2) uses the structure of differential forms on $\mathsf{Diff}(|K|_\infty, X)$ from Proposition 5.2.3. This Proposition states that a differential form on $\square^n$ equipped with the simplicial smooth structure from Definition 5.1.10 is determined by its restriction to the subsimplices Δ^n_σ, for all $\sigma \in \Sigma_n$.

While this definition is still quite involved for general mapping spaces, we provide a host of explicit examples for $\mathsf{Diff}([0,1], \mathbb{R}^N)$ in Section 6.1.1, which recover the standard path signature monomials, and for $\mathsf{Diff}(\square^d, \mathbb{R}^N)$ in Section 7.1.3, which lead to the monomials used to define the mapping space signature. When we consider these explicit examples in the following chapters, we work with a single non-degenerate cube b. In this case, the indexing set $\mathcal{I}$ can be described as a collection of k-ordered subsets $\mathcal{I} = (I_1, \dots, I_m)$. Furthermore, since we will be considering the case where $K_\bullet$ is a cubical model for a geometric d-cube, the evaluation map for the non-degenerate cube b will be the identity map, further simplifying the formula.

5.3.2 *Product Structure for Cubical Monomials*

Our final task in this chapter is to consider the product structure on the collection of cubical monomials, similar to the shuffle product for the standard path signature monomials. Suppose we have two cubical monomials, one at level n_1 indexed by

$$\begin{aligned} \mathcal{I} &= \Big((I_1, b_1), \ldots, (I_{m_1}, b_{m_1})\Big), \\ P &= (\omega_1, \ldots, \omega_{m_1}), \end{aligned} \tag{5.3.3}$$

and another at level n_2 indexed by

$$\begin{aligned} \mathcal{J} &= \Big((J_1, c_1), \ldots, (J_{m_2}, c_{m_2})\Big), \\ Q &= (\nu_1, \ldots, \nu_{m_2}). \end{aligned} \tag{5.3.4}$$

By using the definition of the cubical monomials in Equation (5.3.1), the product of these monomials for a map $\gamma \in M^K$ is

$$\begin{aligned} &M^{\mathcal{I},P}(\gamma) M^{\mathcal{J},Q}(\gamma) \\ &\quad = \left(\int_{\square^{n_1}} \bigwedge_{k=1}^{m} \left(\eta[b_k] \circ \eta_{n_1}^{I_k} \right)^* (\gamma^* \omega_k) \right) \left(\int_{\square^{n_2}} \bigwedge_{k=1}^{m_2} \left(\eta[c_k] \circ \eta_{n_2}^{J_k} \right)^* (\gamma^* \nu_k) \right). \end{aligned}$$

Let $\rho_1 : \square^{n_1+n_2} \to \square^{n_1}$ and $\rho_2 : \square^{n_1+n_2} \to \square^{n_2}$ be the two projection maps. By taking the pullback of the two integrands along these projection maps, we can write the product as a single integral,

$$\begin{aligned} &M^{\mathcal{I},P}(\gamma) M^{\mathcal{J},Q}(\gamma) \\ &\quad = \int_{\square^{n_1+n_2}} \left(\bigwedge_{k=1}^{m} \left(\eta[b_k] \circ \eta_{n_1}^{I_k} \circ \rho_1 \right)^* (\gamma^* \omega_k) \right) \wedge \left(\bigwedge_{k=1}^{m_2} \left(\eta[c_k] \circ \eta_{n_2}^{J_k} \circ \rho_2 \right)^* (\gamma^* \nu_k) \right). \end{aligned} \tag{5.3.5}$$

In order to interpret this integral as a cubical monomial, we must rewrite $\eta_{n_1}^{I_k} \circ \rho_1$ and $\eta_{n_2}^{J_k} \circ \rho_2$ as evaluation maps for $n_1 + n_2$ cubes.

Lemma 5.3.1. *Let $\rho_1 : \square^{n_1+n_2} \to \square^{n_1}$ and $\rho_2 : \square^{n_1+n_2} \to \square^{n_2}$ be the projection maps onto $\square^{n_1}$ and $\square^{n_2}$. Let I be a k-ordered subset of $[n_1]$ and J be a k-ordered subset of $[n_2]$. Let $\widetilde{I}, \widetilde{J}$ be k-ordered subsets of $[n_1 + n_2]$ defined by*

$$\begin{aligned} \widetilde{I} &= I \\ \widetilde{J} &= J + n_1 = \{j + n_1 \,:\, j \in J\}. \end{aligned}$$

In particular, $\widetilde{I}$ is equivalent to I, but reinterpreted as a k-ordered subset of $[n_1+n_2]$. The ordered partitions of $\widetilde{I}$ and $\widetilde{J}$ are induced by those of I and J. Then,

$$\eta^{\widetilde{I}}_{n_1+n_2} = \eta^{I}_{n_1} \circ \rho_1 \tag{5.3.6}$$
$$\eta^{\widetilde{J}}_{n_1+n_2} = \eta^{J}_{n_2} \circ \rho_2. \tag{5.3.7}$$

Proof. Equation (5.3.6) is straightforward to show by definition:

$$\eta^{I}_{n_1} \circ \rho_1(t_1, \dots, t_{n_1+n_2}) = (s_1, \dots, s_k), \qquad s_r = \max_{j \in I^{(r)}}\{t_j\},$$

which is equal to $\eta^{\widetilde{I}}_{n_1+n_2}$ by definition. Expanding the right hand side of Equation (5.3.7), we get

$$\begin{aligned}\eta^{J}_{n_2} \circ \rho_2(t_1, \dots, t_{n_1+n_2}) &= \eta^{J}_{n_2}(t_{n_1+1}, \dots, t_{n_1+n_2}) \\ &= (s_1, \dots, s_k), \qquad s_r = \max_{j \in J^{(r)}}\{t_{n_1+j}\},\end{aligned}$$

which is equal to $\eta^{\widetilde{J}}_{n_1+n_2}$. □

Theorem 5.3.1. *Let $M^{\mathcal{I},P}$ be a level n_1 cubical monomial and $M^{\mathcal{J},Q}$ be a level n_2 cubical monomial, where the index sets are defined in Equation (5.3.3) and Equation (5.3.4) respectively. Let*

$$\begin{aligned}\mathcal{I} * \mathcal{J} &= \Big((I_1, b_1), \dots, (I_{m_1}, b_{m_1}), (J_1 + n_1, c_1), \dots, (J_{m_2} + n_1, c_{m_2})\Big) \\ P * Q &= (\omega_1, \dots, \omega_{m_1}, \nu_1, \dots, \nu_{m_2}),\end{aligned}$$

where each I_k and $J_k + n_1$ are viewed as k-ordered subsets of $[n_1+n_2]$. Then for any $\gamma \in M^K$,

$$M^{\mathcal{I},P}(\gamma) M^{\mathcal{J},Q}(\gamma) = M^{\mathcal{I}*\mathcal{J},P*Q}(\gamma). \tag{5.3.8}$$

Proof. This product formula is obtained by applying Lemma 5.3.1 to the expansion of the product in Equation (5.3.5). □

In Section 6.3, we will show how this product structure on $\mathsf{Diff}([0,1], \mathbb{R}^N)$ reduces to the ordinary shuffle product for the path signature monomials. Furthermore, in Section 7.3, we will derive a generalized shuffle product formula for cubical monomials on the mapping space $\mathsf{Diff}(\square^d, \mathbb{R}^N)$.

6

THE CUBICAL FORMULATION OF THE PATH SIGNATURE

In the previous chapter, we developed a cubical variant of the Chen model for mapping spaces and discussed the resulting 0-cochains for a general mapping space $\mathsf{Diff}(|K|_\infty, X)$, where $|K|_\infty$ is the realization of a finite cubical set $K_\bullet$, and X is a smooth manifold. In this chapter, we consider the path space of $X = \mathbb{R}^N$, so we take $K = [0,1]$. Our goal in this section is to demonstrate how to extract the standard path signature from the 0-cochains of the cubical model. Furthermore, we will show that the shuffle product arises naturally from the algebraic structure of the model and Chen's identity can be derived using the composition structure of cubical sets. While we do not derive any new properties of the path signature in this section, we provide an overview of how to work with the 0-cochains in a familiar setting. Many aspects of the mapping space signature, which we introduce in the next chapter, will be studied in a similar manner.

The general goal of both this chapter and the next is to determine a collection of the 0-cochains of the cubical mapping space model which are rich enough to separate points in the path space or mapping space; has some algebraic structure such as the shuffle product; and is equipped with a composition law such as Chen's identity. From this point onward, we will call individual 0-cochains *signature monomials*, or simply *monomials*. This terminology comes from the analogy between the moment generating function and the expected path signature. Just as moments (expectations of monomials) are used to characterize measures on finite dimensional vector spaces, the expected path signature coefficients are used to characterize measures on path spaces.

We begin this chapter by providing explicit descriptions of the full set of cubical monomials as well as several examples. Next, we consider the monomials indexed by flags, which correspond to the coefficients of the path signature. We then derive the shuffle product and Chen's identity using the algebra structure on cochains and the composition structure on cubical sets respectively.

Notation: As we will be comparing the monomials obtained from the cubical model with those obtained from the simplicial model, we will use S^P_Δ to denote the ordinary signature with respect to the multi-index P. We will also call these *simplicial signature monomials* or simply *simplicial monomials*. We will continue to use "cubical sets" to refer to

cubical sets with connection and implicitly assume that the geometric n-cube is equipped with the simplicial smooth structure from Definition 5.1.10.

6.1 CUBICAL MONOMIALS

Throughout this section, we fix $K_\bullet \in$ to be the cubical model for the unit interval $[0,1]$ as described in Example 5.1.1 and Example 5.1.2, and we fix $X = \mathbb{R}^N$. Recall from Definition 5.3.1 that a cubical monomial of level n is defined by choosing a collection of m differential forms $\omega_j \in \Omega^{q_j}(\mathbb{R}^N)$ paired with m n-cubes in K_n such the total degree is $\sum_{j=1}^m q_j = n$. Thus, there are two pieces of information which are required to specify a level n monomial:

1. A collection of n-cubes denoted by

$$\mathcal{I} = \Big((I_1, b_1), \dots, (I_m, b_m)\Big), and$$

2. A collection of differential forms

$$P = (\omega_1, \dots, \omega_m)$$

such that $\omega_j \in \Omega^{q_j}(\mathbb{R}^N)$ associated to each of the chosen n-cubes.

We begin by discussing the first step, which is slightly more involved. Recall that our cubical model $K_\bullet$ for the unit interval only has three non-degenerate cubes: two 0-cubes v_0, v_1 representing the endpoints, and a single 1-cube e representing the interval. Each of the nondegenerate 0-cubes has exactly one degeneracy at level p, written as

$$s_p s_{p-1} \dots s_1 v_j.$$

The evaluation maps $\eta[s_p s_{p-1} \dots s_1 v_j] : \square^p \to [0,1]$ associated to these degenerate cubes are the constant functions at the end points of $[0,1]$. Thus the pullback along these evaluation maps are trivial, and so we omit these cubes from our consideration.

This leaves the degree p degeneracies of the single 1-cube e, which can be indexed using ordered subsets $I \subset [p]$ as discussed at the end of the previous chapter. Note that we do not consider any partitions since the corresponding non-degenerate cube is of degree 1, thus resulting in a partition made up of a single subset.

We must then associate differential forms $\omega_I \in \Omega^{q_I}(\mathbb{R}^N)$ to collection of chosen subsets $I \subset [p]$. However, because the domain of our mapping space $[0,1]$ is only one dimensional, the only forms $\omega \in \Omega^\bullet(\mathbb{R}^N)$ with a nontrivial pullback are the 1-forms. This forces each of the chosen forms ω_I to be 1-forms, and thus for a level p monomial, we must choose exactly

p degeneracies of the edge to associate 1-forms to. Furthermore, we restrict ourselves to using the standard 1-forms dx_i of $\mathbb{R}^N$.

Definition 6.1.1. The *level n cubical monomials* of the path space $\mathbf{Map}([0,1],\mathbb{R}^N)$ are indexed by two objects:

$$\mathcal{I} = (I_1, \dots, I_n)$$
$$P = (p_1, \dots, p_n)$$

where $I_j \subset [n]$ such that $I_j \neq \emptyset$ and the I_j are all distinct, and P is a multi-index where $p_j \in [N]$. Let $\gamma = (\gamma_1, \dots, \gamma_N) : [0,1] \to \mathbb{R}^N$. The *cubical signature monomial of γ with respect to $\mathcal{I}$ and P* is defined to be

$$S_{\square}^{\mathcal{I},P}(\gamma) := \int_{\mathbf{t}\in\square^n} \bigwedge_{k=1}^{n} \gamma'_{p_k}(s_k)ds_k, \quad s_k = \max_{i\in I_k}\{t_i\}. \tag{6.1.1}$$

In this definition, the $\mathcal{I}$ specifies the choice of level n degeneracies of the edge e which in turn specifies the coordinates over which the s variables take their maximum, and p_j specifies that the 1-form dx_{p_j} should be associated to I_j. Because we are using standard 1-forms on $\mathbb{R}^N$, we can explicitly compute one of the pullbacks from the general formula for monomials in Equation (5.3.1) as $\gamma^* dx_j = \gamma_j(t)dt$. Furthermore, we denote the pullback along the evaluation map explicitly in terms of the s variables

$$(\eta_n^{I_k})^*(\gamma'_{p_k}(t)dt) = \gamma'_{p_k}(s_k)ds_k. \tag{6.1.2}$$

As discussed in the previous section, differential forms on $\square^n$ equipped with the simplicial smooth structure are determined by their restrictions onto each subsimplex $\Delta^n_\sigma \subset \square^n$, and the signature is determined by summing over all subsimplices as in Equation (5.3.1).

6.1.1 *Examples*

We now provide several explicit examples of these cubical monomials. We focus providing some intuition for various choices of $\mathcal{I}$. We begin with a couple of examples where the choice of a simple $\mathcal{I}$ leads to trivial information. We then explore some other choices that will lead us back to the ordinary path signature.

Example 6.1.1. We begin with the case where $\mathcal{I}$ consists of singletons. Let's consider a level $n = 2$ monomial with $I_1 = \{1\}$ and $I_2 = \{2\}$. Let $\gamma = (\gamma_1, \dots, \gamma_N) : \square^1 \to \mathbb{R}^N$, and take $P = (p_1, p_2)$. In this case, $s_i = t_i$, and Equation (6.1.1) reduces to

$$\begin{aligned} S_{\square}^{\mathcal{I},P}(\gamma) &= \int_{\square^2} \gamma'_{p_1}(t_1)\gamma'_{p_2}(t_2)dt_1dt_2 \\ &= \big(\gamma_{p_1}(1) - \gamma_{p_1}(0)\big)\big(\gamma_{p_2}(1) - \gamma_{p_2}(0)\big). \end{aligned}$$

Example 6.1.2. We can generalize the previous example for an arbitrary level n. Let $\mathcal{I} = (I_1, \dots, I_n)$ consist of singletons $I_j = \{j\}$, and let $P = (p_1, \dots, p_n)$ be a multiset of $[N]$. As in the previous example, we have $s_i = t_i$, and the cubical monomial is

$$\begin{aligned} S_{\square}^{\mathcal{I},P} &= \int_{\square^n} \gamma'_{p_1}(t_1) \dots \gamma'_{p_n}(t_n)dt_1 \dots dt_n \\ &= \prod_{k=1}^{n} \left(\gamma_{p_k}(1) - \gamma_{p_k}(0)\right). \end{aligned}$$

These examples show that when $\mathcal{I}$ consist entirely of singletons, the cubical signature only provides information about the displacement of the path in various dimensions. Thus, in order to gain any other information, we must consider other subsets in $\mathcal{I}$.

Example 6.1.3. Next, we consider an example at level $n = 3$, with $I_1 = \{1\}, I_2 = \{2, 3\}, I_3 = \{1, 2, 3\}$. In this case, the pullbacks $(\eta_3^{I_j})^*(\gamma'_{p_j}(t)dt)$ are computed piecewise to be

$$\begin{aligned} \gamma'_{p_1}(s_1)ds_1 &= \gamma'_{p_1}(t_1)dt_1 \\ \gamma'_{p_2}(s_2)ds_2 &= \begin{cases} \gamma'_{p_2}(t_2)dt_2 & : t_3 < t_2 \\ \gamma'_{p_2}(t_3)dt_3 & : t_2 < t_3, \end{cases} \\ \gamma'_{p_3}(s_3)ds_3 &= \begin{cases} \gamma'_{p_3}(t_1)dt_1 & : t_2, t_3 < t_1 \\ \gamma'_{p_3}(t_2)dt_2 & : t_1, t_3 < t_2 \\ \gamma'_{p_3}(t_3)dt_3 & : t_1, t_2 < t_3. \end{cases} \end{aligned}$$

We first note that in the region where t_1 is the largest, the wedge product of these three will be 0 since both the first and third pullbacks contribute a dt_1 term. Then in the simplices where t_2 (resp. t_3) is the largest, both the second and third pullback contribute a dt_2 (resp. dt_3) term, and thus must be 0. Therefore, since the wedge product is 0 for all subsimplices, it is zero over the entire integration domain, so

$$S_{\square}^{\mathcal{I},P}(\gamma) = 0$$

for all $\gamma \in \mathsf{Diff}([0,1], \mathbb{R}^N)$. This demonstrates that certain nontrivial choices of $\mathcal{I}$ can result in a trivial monomial.

This example leads to a key observation. If we wish to construct a nontrivial level n monomial with $\mathcal{I} = (I_1, \ldots, I_n)$, there must exist at least one subsimplex Δ^n_σ, given by the ordering $\sigma \in \Sigma_n$ of coordinates, such that the collection of maximum coordinates s_k each projects onto a distinct coordinate in $\mathbf{t} = (t_1, \ldots, t_n)$. As seen in the above example, if s variables project onto the same coordinate t_i, the integrand will have a $dt_i \wedge dt_i$ term, leading to 0.

Example 6.1.4. Now, we consider the case where $\mathcal{I}$ is a flag. Specifically, we consider $n = 2$ and take $I_1 = \{1\}$ and $I_2 = \{1, 2\}$. By considering the two subsimplices of the unit square individually as in the previous example, the pullbacks can be defined piecewise as

$$\gamma'_{p_1}(s_1)ds_1 = \gamma'_{p_1}(t_1)dt_1$$
$$\gamma'_{p_2}(s_2)ds_2 = \begin{cases} \gamma'_{p_2}(t_1)dt_1 & : t_2 < t_1 \\ \gamma'_{p_2}(t_2)dt_2 & : t_1 < t_2. \end{cases}$$

Thus, we have

$$\bigwedge_{k=1}^{2} \gamma'_{p_k}(s_k)ds_k = \begin{cases} 0 & : t_2 < t_1 \\ \gamma'_{p_1}(t_1)\gamma'_{p_2}(t_2)dt_1 \wedge dt_2 & : t_1 < t_2 \end{cases}$$

where we use the fact that $dt_1 \wedge dt_1 = 0$. Then in this case, since the integrand is 0 when $0 \leq t_2 < t_1 \leq 1$, the integration domain reduces to the standard 2-simplex, and the cubical signature monomial is

$$S^{\mathcal{I},P}_{\square}(\gamma) = \int_{\vec{t} \in \Delta^2} \gamma'_{p_1}(t_1)\gamma'_{p_2}(t_2)dt_1 dt_2$$

Note that this is equivalent to $S^P_\Delta(\gamma)$, one of the simplicial signature monomials.

Example 6.1.5. Here, we consider the case of a flag at level $n = 3$ so that $I_1 = \{1\}, I_2 = \{1,2\}, I_3 = \{1,2,3\}$. In this case, the three 1-forms can be written piecewise as

$$\begin{aligned}
\gamma'_{p_1}(s_1)ds_1 &= \gamma'_{p_1}(t_1)dt_1 \\
\gamma'_{p_2}(s_2)ds_2 &= \begin{cases} \gamma'_{p_2}(t_1)dt_1 & : t_2 < t_1 \\ \gamma'_{p_2}(t_2)dt_2 & : t_1 < t_2, \end{cases} \\
\gamma'_{p_3}(s_3)ds_3 &= \begin{cases} \gamma'_{p_3}(t_1)dt_1 & : t_2, t_3 < t_1 \\ \gamma'_{p_3}(t_2)dt_2 & : t_1, t_3 < t_2 \\ \gamma'_{p_3}(t_3)dt_3 & : t_1, t_2 < t_3. \end{cases}
\end{aligned}$$

We will now need to take the wedge product of these three 1-forms. The only region where the wedge product is nonzero is when $t_1 < t_2 < t_3$.

$$\bigwedge_{k=1}^{3} \gamma'_{p_k}(s_k)ds_k = \begin{cases} \gamma'_{p_1}(t_1)\gamma'_{p_2}(t_2)\gamma'_{p_3}(t_3)dt_1 \wedge dt_2 \wedge dt_3 & : t_1 < t_2 < t_3 \\ 0 & : \text{otherwise.} \end{cases}$$

Finally, we see that the resulting monomial coincides with the ordinary path signature

$$\begin{aligned}
S_{\square}^{\mathcal{I},P}(\gamma) &= \int_{\vec{t} \in \Delta^3} \gamma'_{p_1}(t_1)\gamma'_{p_2}(t_2)\gamma'_{p_3}(t_3)dt_1dt_2dt_3 \\
&= S_{\Delta}^{P}(\gamma).
\end{aligned}$$

Example 6.1.6. We will conclude with one more example at level $n = 3$. We take $I_1 = \{1\}, I_2 = \{1,2\}, I_3 = \{3\}$. In this case, the forms are

$$\begin{aligned}
\gamma'_{p_1}(s_1)ds_1 &= \gamma'_{p_1}(t_1)dt_1 \\
\gamma'_{p_2}(s_2)ds_2 &= \begin{cases} \gamma'_{p_2}(t_1)dt_1 & : t_2 < t_1 \\ \gamma'_{p_2}(t_2)dt_2 & : t_1 < t_2, \end{cases} \\
\gamma'_{p_3}(s_3)ds_3 &= \gamma'_{p_3}(t_3)dt_3
\end{aligned}$$

In this case, the wedge product will be zero in all subsimplices in which $t_1 < t_2$, since both the first and second form will contribute a dt_1 term. This leaves us with the three subsimplices $\Delta^3_{\sigma_j}$ where $\sigma_j \in \Sigma_3$ is defined as

$$\begin{aligned}
\sigma_1 &= (123) \\
\sigma_2 &= (132) \\
\sigma_3 &= (312).
\end{aligned}$$

Let $R = \Delta^3_{\sigma_1} \cup \Delta^3_{\sigma_2} \cup \Delta^3_{\sigma_3}$. The integrand can be written as

$$\bigwedge_{k=1}^{3} \gamma'_{p_k}(s_k)ds_k = \begin{cases} \gamma'_{p_1}(t_1)\gamma'_{p_2}(t_2)\gamma'_{p_3}(t_3)dt_1 \wedge dt_2 \wedge dt_3 & : \vec{t} \in R \\ 0 & : \text{otherwise.} \end{cases}$$

Because R consists of three simplices, we can write the cubical monomial as a sum of three simplicial monomials.

$$S^{\mathcal{I},P}_{\square}(\gamma) = \sum_{j=1}^{3} \int_{\Delta^3_{\sigma_j}} \gamma'_{p_1}(t_1)\gamma'_{p_2}(t_2)\gamma'_{p_3}(t_3)d\vec{t}$$

Furthermore, we can perform a change of variables on the integrals such that each integration domain is the standard 3-simplex, and we can thus interpret this as the sum of three ordinary path signature terms

$$\begin{aligned} S^{\mathcal{I},P}_{\square}(\gamma) &= \int_{\vec{t}\in\Delta^3} \gamma'_{p_1}(t_1)\gamma'_{p_2}(t_2)\gamma'_{p_3}(t_3)d\vec{t} + \int_{\vec{t}\in\Delta^3} \gamma'_{p_1}(t_1)\gamma'_{p_2}(t_3)\gamma'_{p_3}(t_2)d\vec{t} \\ &\quad + \int_{\vec{t}\in\Delta^3} \gamma'_{p_1}(t_3)\gamma'_{p_2}(t_1)\gamma'_{p_3}(t_2)d\vec{t} \\ &= S^{(p_1,p_2,p_3)}_{\Delta}(\gamma) + S^{(p_1,p_3,p_2)}_{\Delta}(\gamma) + S^{(p_3,p_1,p_2)}_{\Delta}(\gamma). \end{aligned}$$

This example demonstrates that in general, we should be able to write the cubical monomials as a linear combination of ordinary path signature monomials. Moreover, note that the resulting sum of three signature monomials corresponds exactly to the shuffle product for $S^{(p_1,p_2)}_{\Delta}(\gamma)S^{p_3}_{\Delta}(\gamma)$. We will make both of these statements precise in the following sections.

6.2 RELATIONSHIP BETWEEN CUBICAL AND SIMPLICIAL MONOMIALS

As we saw in Example 6.1.4 and Example 6.1.5, certain choices of $\mathcal{I}$ result in cubical monomials which are equivalent to the simplicial monomials. We will now show that this is true in general: the simplicial path signature can be recovered from the cubical monomials.

Definition 6.2.1. Let $n \in \mathbb{N}$. An *n-flag* is an ordered set of subsets of $[n]$ defined by

$$\mathcal{F}^n := ([1], [2], \ldots, [n-1], [n]). \tag{6.2.1}$$

Proposition 6.2.1. *Let $\gamma : [0,1] \to \mathbb{R}^N$, and $P = (p_1, \ldots, p_n)$ be a multi-index with values in $[N]$. Then*

$$S_{\square}^{\mathcal{F}^n,P}(\gamma) = S_{\Delta}^{P}(\gamma).$$

Proof. First, we consider the pullback of forms along the evaluation maps restricted to the standard n-simplex

$$\Delta^n = \{0 \leq t_1 < t_2 < \ldots < t_n \leq 1\}.$$

Restricted to Δ^n, the s variables are $s_k = t_k$ by definition. Thus, within Δ^n, the pullbacks are evaluated as

$$\gamma'_{p_k}(s_k)ds_k = \gamma'_{p_k}(t_k)dt_k.$$

Therefore, within Δ^n, the wedge product is

$$\bigwedge_{k=1}^{n} \gamma'_{p_k}(s_k)ds_k = \left(\gamma'_{p_1}(t_1)\ldots\gamma'_{p_n}(t_n)\right) dt_1 \wedge \ldots \wedge dt_n,$$

and the integral over Δ^n is exactly the simplicial signature $S_{\Delta}^{P}(\gamma)$.

Next, we must show that the wedge product is 0 outside of the standard n-simplex. Consider the n-simplex Δ_σ^n, where $\sigma \in \Sigma_n$ is not the identity. Then, there exist distinct $j, k \in [N]$ where $s_j = \max_{i\in[j]}\{t_i\} = \max_{i\in[k]}\{t_i\} = s_j$. Therefore, the wedge product in this subsimplex will contain a $dt_j \wedge dt_j$ term, and will be 0. □

Moreover Example 6.1.6 suggested that a stronger statement holds: that all cubical monomials can be represented as a linear combination of simplicial monomials. By following the computations in Example 6.1.6, this fact is not too surprising. Indeed, let $\mathcal{I} = (I_1, \ldots, I_n)$. Restricted to a simplex Δ_σ^n, each of the s variables s_k are mapped to a fixed t variable $t_{k'}$. The specific coordinates are given by the *maximal coordinates*.

Definition 6.2.2. Let $\mathcal{I} \subset 2^{[n]}$ such that $\emptyset \notin \mathcal{I}$, and let $\sigma \in \Sigma_n$. The *maximal coordinates with respect to $\mathcal{I}$ and σ* is

$$\mu(\mathcal{I}, \sigma) := \left(\text{argmax}_{i\in I_1}\{t_i\}, \ldots, \text{argmax}_{i\in I_n}\{t_i\}\right) \in [n]^n, \tag{6.2.2}$$

where the argmax is taken within Δ_σ^n.

Thus, restricted to each Δ_σ^n, the s variables can be explicitly written as

$$s_k = t_{\mu(\mathcal{I},\sigma)(k)}.$$

A priori, $\mu(\mathcal{I}, \sigma)$ is only an element of $[n]^n$, since there may be duplicate of a single coordinate. However, as we saw in the examples, the integrand restricted to Δ^n_σ is trivial when $\mu(\mathcal{I}, \sigma)$ includes a duplicate of any coordinate due to a $dt \wedge dt$ term.

Definition 6.2.3. Let $\mathcal{I} \subset 2^{[n]}$ such that $\emptyset \notin \mathcal{I}$. Given $\sigma \in \Sigma_n$, the subsimplex Δ^n_σ is a *nontrivial domain* if $\mu(\mathcal{I}, \sigma)$ has no duplicate elements, and thus $\mu(\mathcal{I}, \sigma) \in \Sigma_n$.

In order to obtain an explicit formula for an arbitrary cubical monomial in terms of the simplicial monomials, we only need to consider the nontrivial domains of $\square^n$. We will characterize the nontrivial domains using the perspective of posets and lattices. Given a set A, we let 2^A denote the power set of A. We define the index set $\mathcal{I}$ for a level n cubical monomial as a subset $\mathcal{I} \subset 2^{[n]}$ such that $|\mathcal{I}| = n$ and $\emptyset \notin \mathcal{I}$. Any subset $\mathcal{I} = \{I_1, \ldots, I_n\} \subset 2^{[n]}$ can be viewed as a poset by setting $I_j \leq I_k$ if $I_j \subset I_k$. Such a poset is called a *lattice* if any two elements I_j and I_k have a least upper bound, called the *join*, and a greatest upper bound, called the *meet*. For example, the full power set $2^{[n]}$ is a lattice. Indeed, given any two subsets $A, B \in 2^{[n]}$, the join is the union $A \cup B$ and the meet is the intersection $A \cap B$.

Definition 6.2.4. Let $\mathcal{I} \subset 2^{[n]}$. The *union completion of* $\mathcal{I}$ is defined to be

$$U(\mathcal{I}) := \left\{ \bigcup_{k=1}^{n} I_k \,:\, I_k \in \mathcal{I} \right\} \cup \{\emptyset\}. \tag{6.2.3}$$

While we do not need the structure of the union completion to proceed, we will include it for completeness.

Lemma 6.2.1. *Let $\mathcal{I} \subset 2^{[n]}$. The union completion of $\mathcal{I}$, $U(\mathcal{I})$, is a lattice.*

Proof. Given any $I, J \in U(\mathcal{I})$, the join is $I \cup J$, which is contained in $U(\mathcal{I})$ by definition. Within $U(\mathcal{I})$, the meet of I and J is defined to be

$$\bigcup \{A \in \mathcal{I} \,:\, A \subset I, A \subset J\}.$$

This element is contained in $U(\mathcal{I})$ by definition. Note that the meet between two elements is not necessarily the intersection within $U(\mathcal{I})$. □

Given $\mathcal{I} \subset 2^{[n]}$, we can use the union completion $U(\mathcal{I})$ to enumerate the subsimplices with nonzero integrands.

Definition 6.2.5. Let $\mathcal{I} \subset 2^{[n]}$ and $U(\mathcal{I})$ be its union completion. A *maximal chain* of $U(\mathcal{I})$ is a sequence of n nonempty subsets of $U(\mathcal{I})$ ordered by inclusions:

$$I_1 \subset \ldots \subset I_n = [n].$$

The cardinality of each subset must be $|I_k| = k$, and thus each maximal chain corresponds to a permutation $\sigma \in \Sigma_n$ determined by

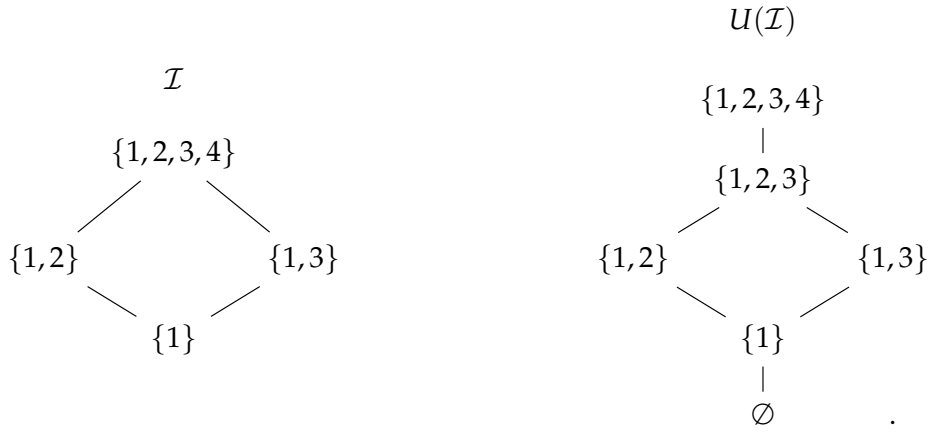
$$\sigma = (I_1, I_2 - I_1, \ldots, I_n - I_{n-1}) .$$

The collection of maximal chains of $U(\mathcal{I})$ is denoted $M(\mathcal{I})$, and we view this as a subset of the symmetric group Σ_n.

Example 6.2.1. Let $\mathcal{I} \subset 2^{[4]}$ be defined by

$$\mathcal{I} = \{\{1\}, \{1,2\}, \{1,3\}, \{1,2,3,4\}\} .$$

We draw the Hasse diagrams of $\mathcal{I}$ as well as its union completion $U(\mathcal{I})$ in order to visualize its poset structure:



There are two maximal chains in $U(\mathcal{I})$,

$$\{1\} \subset \{1,2\} \subset \{1,2,3\} \subset \{1,2,3,4\},$$
$$\{1\} \subset \{1,3\} \subset \{1,2,3\} \subset \{1,2,3,4\},$$

with the corresponding permutations

$$\sigma_1 = (1234)$$
$$\sigma_2 = (1324).$$

Thus, we have $M(\mathcal{I}) = \{\sigma_1, \sigma_2\} \subset \Sigma_4$.

We claim that $M(\mathcal{I})$ indexes all of the nontrivial domains Δ^n_σ. We will prove this with a sequence of lemmas.

Lemma 6.2.2. *Let $\mathcal{I} = (I_1, \dots, I_n) \subset 2^{[n]}$ such that $\emptyset \notin \mathcal{I}$. Suppose $\sigma \in \Sigma_n$, and let $\sigma_k \subset [n]$ denote the first k elements of σ as a set. The maximal coordinates $\mu(\mathcal{I}, \sigma)$ defined in Equation (6.2.2) has distinct elements if and only if σ satisfies the* nonzero condition*: for all $k \in [n]$, there exists exactly one $j \in [n]$ such that*

$$\sigma_k \cap I_j = I_j, \qquad \sigma_{k-1} \cap I_j \neq I_j. \tag{6.2.4}$$

Proof. First, let suppose $\sigma \in \Sigma_n$ satisfies the nonzero condition. For each $j \in [n]$, Equation 6.2.4 implies that

$$\operatorname{argmax}_{i \in I_j}(t_i) = t_{\sigma(k)}$$

on Δ^n_σ. This implies that the largest coordinate corresponding to each I_j will be distinct, and thus $\mu(\mathcal{I}, \sigma) \in \Sigma_n$.

Next, suppose that $\sigma \in \Sigma_n$ does not satisfy the nonzero condition. This implies that for some $k \in [n]$, there exist two distinct $j_1, j_2 \in [n]$ such that Equation 6.2.4 holds. This means that the maximal coordinate for I_{j_1} and I_{j_2} are both k, so $\mu(\mathcal{I}, \sigma)$ contains a repetition, and is not a permutation. □

We can now show that the permutations corresponding to maximal chains satisfy the nonzero condition.

Lemma 6.2.3. *Let $\mathcal{I} = (I_1, \dots, I_n) \subset 2^{[n]}$ such that $\emptyset \notin \mathcal{I}$. Let $\sigma \in \Sigma_n$. Then σ satisfies the nonzero condition if and only if it corresponds to a maximal chain so that $\sigma \in M(\mathcal{I})$.*

Proof. Suppose σ satisfies the nonzero condition. We will continue to let $\sigma_k \subset [n]$ denote the first k elements of σ as a set. The maximal chain that corresponds to σ is

$$\emptyset = \sigma_0 \subset \sigma_1 \subset \ldots \subset \sigma_n,$$

so we must show that $\sigma_k \in U(\mathcal{I})$ for all $k = 0, \dots, n$. First, we have $\emptyset \in U(\mathcal{I})$ by definition, so $\sigma_0 = \emptyset \in U(\mathcal{I})$. Next, we proceed by induction and suppose that $\sigma_k \in U(\mathcal{I})$, and we aim to show that $\sigma_{k+1} \in U(\mathcal{I})$. By the nonzero condition, we know there exists exactly one $I_j \in \mathcal{I}$ such that Equation 6.2.4 holds for σ_{k+1}. We claim that

$$\sigma_{k+1} = \sigma_k \cup I_j.$$

Indeed, by the nonzero condition, we know that $\sigma(k+1) \in I_j$ and $I_j \subset \sigma_{k+1}$, so these two sets are equal. Thus, $\sigma_{k+1} \in U(\mathcal{I})$.

Next, suppose that σ corresponds to a maximal chain in $U(\mathcal{I})$, meaning $\sigma_k \in U(\mathcal{I})$ for all k. Note that $\sigma_1 \in \mathcal{I}$ by the definition of $U(\mathcal{I})$ (we only add unions, so if a singleton

exists, it must be in $\mathcal{I}$), so the nonzero condition holds for $k = 1$. Let's consider the nonzero condition at k. Note that we can write

$$\sigma_k = I_{j_1} \cup \ldots \cup I_{j_m}$$

where $I_{j_1} \ldots I_{j_m} \in \mathcal{I}$. Thus, one of these sets $I' \in \{I_{j_1}, \ldots, I_{j_m}\}$, contains the element $\sigma(k)$. Then, I' satisfies Equation 6.2.4. This shows that for each k, there exists at least one set $I' \in \mathcal{I}$ that satisfies Equation 6.2.4. However, each set in $\mathcal{I}$ satisfies the equation for exactly one k. Thus, for each k, there is exactly one I_j that satisfies the equation. Therefore, σ satisfies the nonzero condition. □

Putting these two lemmas together, we find that the permutations given by the maximal chains correspond to each of the nonzero domains Δ^n_σ.

Corollary 6.2.1. *Let $\mathcal{I} = (I_1, \ldots, I_n) \subset 2^{[n]}$ such that $\emptyset \notin \mathcal{I}$ and $\sigma \in \Sigma_n$. Furthermore, let P be a multiset of length n valued in $[N]$. There exists a path $\gamma : [0,1] \to \mathbb{R}^N$ such that the integrand of $S^{\mathcal{I},P}_\square(\gamma)$ restricted to Δ^n_σ is nonzero,*

$$\bigwedge_{k=1}^{n} (\gamma'_{p_k}(s_k) ds_k) \neq 0,$$

if and only if $\sigma \in M(\mathcal{I})$.

Now that we have a characterization for the simplices with nonzero integrand, we can provide an explicit formula for the cubical monomials as a linear combination of the simplicial monomials.

Theorem 6.2.1. *Let $\mathcal{I} = (I_1, \ldots, I_n) \subset 2^{[n]}$ such that $\emptyset \notin \mathcal{I}$ and $P = (p_1, \ldots, p_n)$ be a multi-index with values in $[N]$. Then, for all $\gamma : [0,1] \to \mathbb{R}^N$,*

$$S^{\mathcal{I},P}_\square(\gamma) = \sum_{\sigma \in M(\mathcal{I})} S^{(\mu(\mathcal{I},\sigma)^{-1} \circ \sigma)P}_\Delta(\gamma),$$

where $\mu(\mathcal{I}, \sigma)$ is given in Definition 6.2.2.

Proof. As we have shown in Corollary 6.2.1, $M(\mathcal{I})$ indexes the nontrivial domains of $\square^n$. Let's fix $\sigma \in M(\mathcal{I})$ and to simplify notation, we set $\mu := \mu(\mathcal{I}, \sigma)$. To prove the result, we must rewrite the integral over Δ^n_σ as an integral over Δ^n (the standard n-simplex). The integral corresponding to σ is

$$\int_{\vec{t} \in \Delta^n_\sigma} \gamma'_{p_1}(t_{\mu(1)}) \ldots \gamma'_{p_n}(t_{\mu(n)}) dt_1 \ldots dt_n.$$

We perform a change of variables by permuting the coordinates $\vec{t} \mapsto \sigma^{-1}\vec{t}$ so that the domain of integration is mapped $\Delta^n_\sigma \mapsto \Delta^n$. Then, by moving the permutations from the coordinates t_i to the multi-index P, we get

$$\begin{aligned}
&\int_{\vec{t}\in\Delta^n_\sigma} \gamma'_{p_1}(t_{\mu(1)}) \dots \gamma'_{p_n}(t_{\mu(n)}) dt_1 \dots dt_n \\
&\qquad = \int_{\vec{t}\in\Delta^n} \gamma'_{p_1}(t_{\sigma^{-1}\circ\mu(1)}) \dots \gamma'_{p_n}(t_{\sigma^{-1}\circ\mu(n)}) dt_1 \dots dt_n \\
&\qquad = \int_{\vec{t}\in\Delta^n} \gamma'_{p_{\mu^{-1}\circ\sigma(1)}}(t_1) \dots \gamma'_{p_{\mu^{-1}\circ\sigma(n)}}(t_n) dt_1 \dots dt_n \\
&\qquad = S_\Delta^{(\mu^{-1}\circ\sigma)P}(\gamma)
\end{aligned}$$

Adding up all the subsimplices with nonzero integrands indexed by $M(\mathcal{I})$, we get the desired result. □

6.3 SHUFFLE PRODUCT

From the product structure for cubical monomials given in Theorem 5.3.1, we immediately get a product structure for the cubical monomials for the path space.

Theorem 6.3.1. *Let $\mathcal{I} = (I_1, \dots, I_{n_1}) \subset 2^{[n_1]}$ and $\mathcal{J} = (J_1, \dots, J_{n_2}) \subset 2^{[n_2]}$. Let $P = (p_1, \dots, p_{n_1})$ and $Q = (q_1, \dots, q_{n_2}$ be multi-indices in $[N]$ of length n_1 and n_2 respectively. Define*

$$\begin{aligned}
\mathcal{I} * \mathcal{J} &= \Big(I_1, \dots, I_{n_1}, J_1 + n_1, \dots, J_{n_2} + n_1) \\
P * Q &= (p_1, \dots, p_{n_1}, q_1, \dots, q_{n_2}).
\end{aligned}$$

Then for any $\gamma : [0,1] \to \mathbb{R}^N$,

$$S_\square^{\mathcal{I},P}(\gamma) S_\square^{\mathcal{J},Q}(\gamma) = S_\square^{\mathcal{I}*\mathcal{J},P*Q}(\gamma).$$

Proof. This is a special case of Theorem 5.3.1. Note that for the cubical path signature monomials, each of the elements of $\mathcal{I}$ and $\mathcal{J}$ correspond to degeneracies of the 1-cell e. □

When we restrict the subsets $\mathcal{I}$ and $\mathcal{J}$ to be flags, so that the monomials correspond to the simplicial signature monomials, we recover the shuffle product. The set of (n_1, n_2) shuffles, denoted $Sh(n_1, n_2)$ is defined in Definition 1.2.7.

Corollary 6.3.1. *Let $P = (p_1, \dots, p_{n_1})$ and $Q = (q_1, \dots, q_{n_2})$ be multi-indices in $\mathbb{R}^N$. Let $\mathcal{F}_{n_1} \subset 2^{[n_1]}$ and $\mathcal{F}_{n_2} \subset 2^{[n_2]}$ be n_1- and n_2- flags respectively. Then*

$$S_{\square}^{\mathcal{F}_{n_1},P}(\gamma) S_{\square}^{\mathcal{F}_{n_1},P}(\gamma) = \sum_{\sigma \in Sh(n_1,n_2)} S_{\Delta}^{\sigma(P*Q)}(\gamma).$$

Proof. By the product structure for cubical path signature monomials from Theorem 6.3.1 and the expression of cubical monomials as simplicial monomials from Theorem 6.2.1, we immediately get

$$\begin{aligned} S_{\square}^{\mathcal{F}_{n_1},P}(\gamma) S_{\square}^{\mathcal{F}_{n_1},P}(\gamma) &= S_{\square}^{\mathcal{F}_{n_1} * \mathcal{F}_{n_2}, P*Q}(\gamma) \\ &= \sum_{\sigma \in \mathcal{M}(\mathcal{F}_{n_1} * \mathcal{F}_{n_2})} S_{\Delta}^{\mu\left(\mathcal{F}_{n_1} * \mathcal{F}_{n_2}, \sigma\right)^{-1} \circ \sigma(P*Q)}(\gamma). \end{aligned}$$

First, we show that $\mathcal{M}(\mathcal{F}_{n_1} * \mathcal{F}_{n_2}) = Sh(n_1, n_2)$. First, suppose that $\sigma \in Sh(n_1, n_2)$, and recall that $\sigma_k = \{\sigma(1), \dots, \sigma(k)\}$ denotes the set containing the first k elements of σ. The maximal chain corresponding to σ is

$$\emptyset = \sigma_0 \subset \sigma_1 \subset \ldots \subset \sigma_{n_1+n_2}.$$

By the definition of a (n_1, n_2)-shuffle, $\sigma_k = [m_1] \cup ([m_2] + n_1)$ for some $m_1 \in [n_1]$ and $m_2 \in [n_2]$, thus σ_k is in the union completion $U(\mathcal{F}_{n_1} * \mathcal{F}_{n_2})$, so σ corresponds to a maximal chain in $U(\mathcal{F}_{n_1} * \mathcal{F}_{n_2})$. Conversely, every element of $U(\mathcal{F}_{n_1} * \mathcal{F}_{n_2})$ can be written in the form

$$[m_1] \cup ([m_2] + n_1),$$

and a maximal chain of such subsets corresponds to a permutation $\sigma \in Sh(n_1, n_2)$. Thus $\mathcal{M}(\mathcal{F}_{n_1} * \mathcal{F}_{n_2}) = Sh(n_1, n_2)$.

To complete the proof, we must show that $\mu(\mathcal{F}_{n_1} * \mathcal{F}_{n_2}, \sigma) = \mathrm{id}$, the identity permutation, for all $\sigma \in Sh(n_1, n_2)$. Because

$$\mathcal{F}_{n_1} * \mathcal{F}_{n_2} = \Big([1], \dots, [n_1], [1] + n_1, \dots, [n_2] + n_1\Big)$$

and the elements of $[n_1]$ and $[n_2] + n_1$ appear in ascending order in σ since it is a shuffle, we get

$$\begin{aligned} \operatorname{argmax}_{j \in [m_1]} \{t_j\} &= j, \quad m_1 \in [n_1] \\ \operatorname{argmax}_{j \in [m_2]+n_1} \{t_j\} &= m_2 + n_1, \quad m_2 \in [n_2] \end{aligned}$$

and thus

$$\mu(\mathcal{F}_{n_1} * \mathcal{F}_{n_2}, \sigma) = (1, 2, \ldots, n_1 + n_2) = \mathrm{id}.$$

□

6.4 COMPOSITION AND CHEN'S IDENTITY

In the previous section, we derived the shuffle product for the path signature using the product structure on cubical monomials. In this section, we will derive Chen's identity, another fundamental property of the path signature, by using the composition structure for cubical sets. Given two paths $\gamma^{01}, \gamma^{02} : [0,1] \to \mathbb{R}^N$ such that $\gamma^{01}(1) = \gamma^{12}(0)$, the concatenation of γ^{01} and γ^{02} is a path $\gamma^{02} = \gamma^{01} * \gamma^{12} : [0,2] \to \mathbb{R}^N$ defined by

$$\gamma^{02}(t) = \begin{cases} \gamma^{01}(t) & : t \in [0,1] \\ \gamma^{12}(t) & : t \in (1,2]. \end{cases}$$

Our aim is to derive a formula for the cubical monomials of γ^{02} using the cubical monomials of γ^{01} and γ^{02}.

6.4.1 *A Modified Cubical Model for the Interval*

In order to derive this using the composition structure of cubical sets, we require a modified cubical set $L_\bullet$ for the interval $[0,2]$. The 0-cubes of this cubical set are

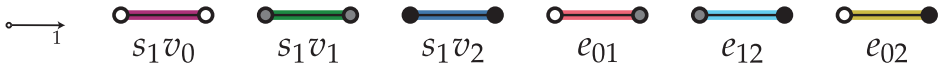
$$L_0 = \{v_0, v_1, v_2\},$$

representing vertices at the points 0, 1, and 2 on the interval. Following the visual representation of cubical sets introduced in Example 5.1.1 and Example 5.1.2, we can represent this visually as shaded points.

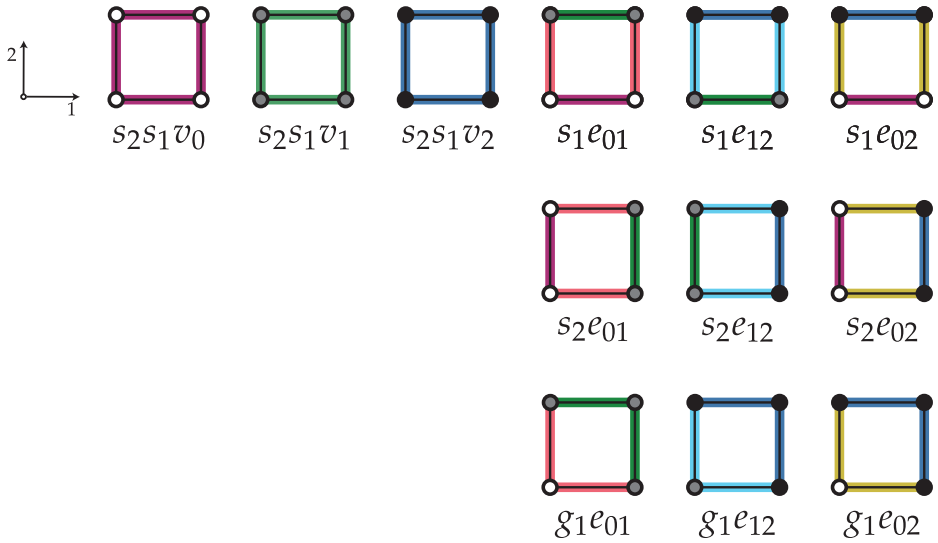
Next, our cubical model has three non-degenerate 1-cells: e_{01} for the interval $[0,1]$, e_{12} for the interval $[1,2]$, and e_{02} for the interval $[0,2]$, so

$$L_1 = \{s_1 v_0, s_1 v_1, s_1 v_2, e_{01}, e_{12}, e_{02}\}.$$

Keeping the same colors for the 1-cubes s_1v_0, s_1v_1 and e_{01} as Example 5.1.1, the 1-cubes can be represented as follows.



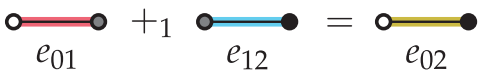
There are no other non-degenerate cubes. In general, we will have a single degeneracy of each vertex in L_n, and a collection of degeneracies for each non-degenerate 1-cube. For example, the collection of 2-cubes is shown below.



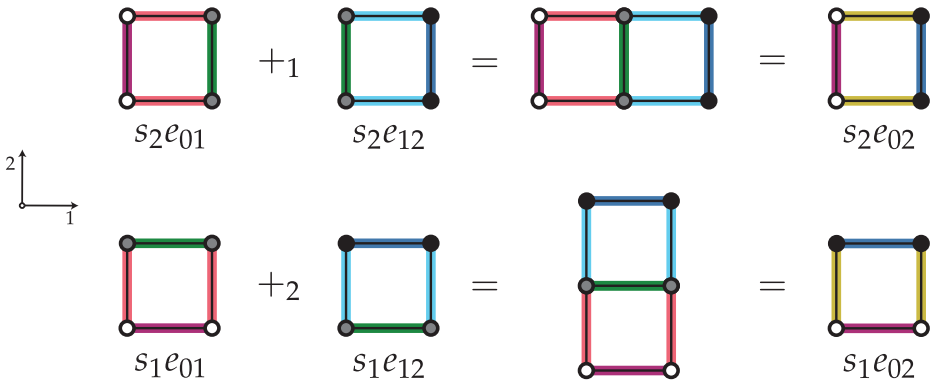
This cubical model for the interval $[0,2]$ allows us to consider the cubical monomials of γ^{01}, γ^{12} and the concatenated γ^{02} within a single model. Furthermore, this cubical model has a nontrivial composition structure introduced in Definition 5.1.5. In particular, e_{01} and e_{12} are composable with respect to $+_1$ since $d^1_1(e_{01}) = d^1_1(e_{01}) = v_1$, and we have

$$e_{01} +_1 e_{12} = e_{02} \tag{6.4.1}$$

which can be visually represented as follows.



In general, we have n partial compositions for the n-cubes which are defined for $a, b \in L_n$ if and only if $d^1_i a = d^0_i b$. Each of the compositions $+_i$ corresponds to composition along various dimensions. For example, consider the following two compositions of 2-cubes.



We define certain degeneracies to be the units for the $+_i$ composition.

Definition 6.4.1. Let $a \in L_n$. The *left i-unit for a* is $s_i d^0_i a$, which has the property that

$$s_i d^0_i a +_i a = a,$$

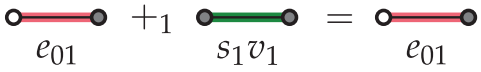
and the *right i-unit for a* is $s_i d^1_i a$, which has the property that

$$a +_i s_i d^1_i a = a.$$

For example, if $a = e_{01}$, the right 1-unit of a is $s_1 d^1_1 a = s_1 v_1$, and here we get

$$e_{01} +_1 s_1 v_1 = e_{01},$$

represented visually as follows.



Next, we discuss a modified geometric realization of this model. Recall from Section 5.1.3, the ordinary geometric realization of $L_\bullet$ can be described as a CW-complex with an n-cell for each non-degenerate cube, as all degenerate cubes are identified with non-degenerate cube in the equivalence relation from Definition 5.1.7. We can describe the geometric realization by defining the evaluation maps $\eta[a]$ for each non-degenerate cube $a \in L_\bullet$. However, we must take compositions into account in $L_\bullet$.

Definition 6.4.2. Let $L_\bullet$ be the cubical set with compositions as defined above. Suppose $a, b \in L_n$ are i-composable, a is not a left i-unit for b, and b is not a right i-unit for a. Then, the evaluation map for the composition $\eta[a +_i b] : \square^n \to [0,2]$ is defined to be

$$\eta[a +_i b](t_1, \dots, t_n) = \begin{cases} \eta[a](t_1, \dots, 2t_i, \dots, t_n) & : t_i \in [0, 1/2] \\ \eta[b](t_1, \dots, 2t_i - 1, \dots, t_n) & : t_i \in (1/2, 1]. \end{cases} \tag{6.4.2}$$

This definition implies that for our modified geometric realization can be defined by providing the evaluation maps for all indecomposable and nondegenerate cubes. In particular, the evaluation maps for the three non-degenerate 1-cubes in $L_\bullet$ are defined by

$$\begin{array}{lll} \eta[e_{01}] : \square^1 \to [0,2] & \eta[e_{12}] : \square^1 \to [0,2] & \eta[e_{02}] : \square^1 \to [0,2] \\ \qquad t \mapsto t, & \qquad t \mapsto 1 + t, & \qquad t \mapsto 2t, \end{array}$$

where the evaluation map $\eta[e_{02}]$ is defined using Equation (6.4.2). The evaluation maps for degenerate cubes can also be derived from this definition. Indeed, given a degeneracy $a \in L_n$ of e_τ, where $\tau \in \{01, 12, 02\}$, we have

$$\eta[a] : \square^n \xrightarrow{\eta^I_n} \square^1 \xrightarrow{\eta[e_\tau]} [0,2],$$

where $I \subset 2^{[n]}$ is given by Lemma 5.1.2.

While we cannot use Equation (6.4.2) for composition with units in general, there is a special case for degenerate cubes when this definition still holds.

Lemma 6.4.1. *Let $a \in L_{n-1}$ and $p \in [n]$. Then,*

$$\eta[s_p a] = \eta[s_p a +_p s_p a]. \tag{6.4.3}$$

Proof. The evaluation map $\eta[s_p a]$ is factored as

$$\eta[s_p a] : \square^n \xrightarrow{\eta^{\{p\},\varnothing}_n} \square^{n-1} \xrightarrow{\eta[a]} [0,2],$$

where $\eta^{\{p\},\varnothing}_n(t_1, \dots, t_n) = (t_1, \dots, t_{p-1}, t_{p+1}, \dots, t_n)$. Therefore, $\eta[s_p a]$ does not depend on coordinate t_p, so applying Equation (6.4.2), we get

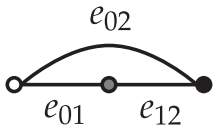
$$\eta[s_p a +_p s_p a] = \eta[s_p a].$$

□

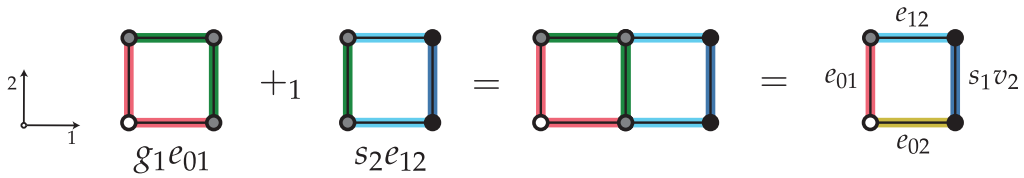
Remark 6.4.1. As we have previously mentioned, this definition of geometric realization for cubical sets with composition does not coincide with ordinary geometric realization.

One notion of geometric realization is considered in [BHS11] for ω-groupoids, which are cubical sets with composition and inverses. A loose analogy is that our modified geometric realization is "strict" like a colimit, while the ordinary geometric realization for ω-groupoids is "weak" like a homotopy colimit.

Indeed, suppose we took the ordinary geometric realization of $L_\bullet$. The 1-skeleton of the geometric realization would consist of 3 distinct edges, and is homeomorphic to S^1.



However, we can obtain nontrivial 2-cubes by composing degenerate 1-cubes. For example, consider the composition $g_1 e_{01} +_1 s_2 e_{12}$, shown visually below.



The composition is not a degeneracy of any of the edges, so we have labelled its faces on the right. In particular, since $g_1 e_{01} +_1 s_2 e_{12}$ is a non-degenerate 2-cube, there will be a 2-cell in the ordinary geometric realization which has e_{01}, e_{12} and e_{02} on its boundary. In fact, if we consider the completion of $L_\bullet$ into an ω-groupoid, we obtain other non-degenerate 2-cubes which will yield additional 2-cells in the realization with e_{01}, e_{12}, and e_{02} on its boundary. Thus, the topology of the 2-skeleton will also be nontrivial, and higher nondegenerate cubes are required to fill in the voids.

Thus, while the ordinary geometric realization of the ω-groupoid corresponding to $L_\bullet$ is homotopy equivalent to $[0,2]$, our modified geometric realization is homeomorphic to $[0,2]$.

Remark 6.4.2. While our modified cubical model for $[0,2]$ is sufficient to consider the concatenation of two paths, we would need a different model to consider concatenations of three (or more) paths. However, even a model for $[0,3]$ with 1 cubes e_{01}, e_{12}, e_{23} and their compositions raises difficulties regarding associativity. Our definition for the modified geometric realization explicitly parametrizes the evaluation maps, and it would not be possible to define consistent evaluation maps for

$$(e_{01} +_1 e_{12}) +_1 e_{23}$$

and

$$e_{01} +_1 (e_{12} +_1 e_{23}).$$

While we show in the following sections that we can use the composition structure of $L_\bullet$ and the modified geometric realization to obtain Chen's identity, these two remarks suggest that this approach may not be the most general. A more general approach may be to consider an ω-groupoid which models the entire real line, such as the fundamental ∞-groupoid of $\mathbb{R}$ interpreted as cubical set (or ω-groupoid). However, such an approach would yield an infinite dimensional geometric realization which is only homotopy equivalent to the real line. Thus, in order to obtain a theory along these lines, one must first interpret cubical monomials of this much larger model, and relate them to the ordinary signature monomials.

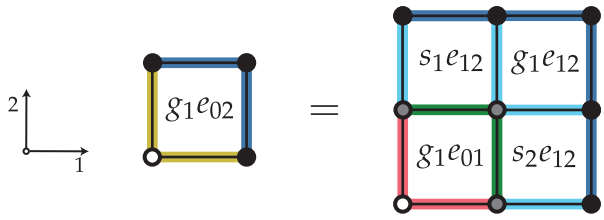
6.4.2 *Decomposition Functions*

In this section, we consider the decomposition of degeneracies of e_{12} into sums of degeneracies of e_{01} and e_{12}.

Example 6.4.1. Consider the degeneracy g_1e_{02}. From the composition identities for connections in Equation (5.1.27), this degeneracy can be decomposed as

$$\begin{aligned} g_1e_{02} &= g_1(e_{01} +_1 e_{12}) \\ &= (g_1e_{01} +_2 s_1e_{12}) +_1 (s_2e_{12} +_2 g_1e_{12}). \end{aligned}$$

This is a composition of four 2-cubes arranged in a square as shown below. Note that by the interchange law in Equation (5.1.28), the order that the $+_1$ and $+_2$ compositions are taken is irrelevant.



A third way of representing this decomposition is by using a function $D[g_1 e_{02}] : \mathbb{Z}_2^2 \to L_2$. Here, $\mathbb{Z}_2^2$ is used to index the four squares in the decomposed square, and the function $D[g_1 e_{02}]$ is defined to be

$$\begin{aligned} D[g_1 e_{02}](0,0) &= g_1 e_{01} \\ D[g_1 e_{02}](0,1) &= s_1 e_{12} \\ D[g_1 e_{02}](1,0) &= s_2 e_{12} \\ D[g_1 e_{02}](1,1) &= g_1 e_{12}. \end{aligned}$$

This functional representation of the decomposition is the most convenient to use in practice, and we introduce this in general.

Definition 6.4.3. Let $L_n^{01,12} \subset L_n$ denote the n-cubes which are degeneracies of e_{01} and e_{12}, and let $L_n^{02} \subset L_n$ denote the n-cubes which are degeneracies of e_{02}. For any $a \in L_n^{02}$, the *decomposition function of* a is the map

$$D[a] : \mathbb{Z}_2^n \to L_n^{01,12}$$

which is defined by the decomposition of a according to the composition identities in Definition 5.1.5. The decomposition function satisfies

$$\underset{i_1 \in \mathbb{Z}_2}{+_1} \underset{i_2 \in \mathbb{Z}_2}{+_2} \cdots \underset{i_n \in \mathbb{Z}_2}{+_n} D[a](i_1, i_2, \ldots, i_n) = a. \tag{6.4.4}$$

The decomposition function provides a piecewise definition of the evaluation map for any degenerate $a \in L_n^{02}$ in terms of the evaluation maps for elements in $L_n^{01,12}$, which is the key to proving Chen's identity. The unit n-cube can be decomposed into 2^n cubes of side length $1/2$. These subcubes can be indexed by $\mathbb{Z}_2^n$ in precisely the same way as $\mathbb{Z}_2^n$ was used to index elements an n-dimensional composition in Definition 6.4.3. Specifically, given $\epsilon = (\epsilon_1, \ldots, \epsilon_n) \in \mathbb{Z}_2^n$, we define

$$\square_\epsilon^n := \prod_{i=1}^n \left[\frac{\epsilon_i}{2}, \frac{\epsilon_i + 1}{2} \right]. \tag{6.4.5}$$

Because

$$\square^n = \bigcup_{\epsilon \in \mathbb{Z}_2^n} \square_\epsilon^n,$$

we can define a function on $\square^n$ by defining functions on each $\square_\epsilon^n$ individually.

Proposition 6.4.1. *Let $a \in L_n^{02}$, and $D[a]$ be its decomposition function. We assume that none the compositions in $D[a]$ are compositions with left/right units, except for the case in Lemma 6.4.1. Let $\epsilon = (\epsilon_1, \dots, \epsilon_n) \in \mathbb{Z}_2^n$, then the evaluation map $\eta[a]$ restricted to $\square_\epsilon^n$ is*

$$\eta[a](t_1, \dots, t_n) = \eta[D[a](\epsilon)](2t_1 - \epsilon_1, \dots, 2t_n - \epsilon_n). \tag{6.4.6}$$

Proof. From the defining property of the decomposition function in Equation (6.4.4), we can apply the definition of the evaluation map for composite cubes in Equation (6.4.2) for each coordinate. If necessary, we also apply Lemma 6.4.1 to deal with the special units in the decomposition function. □

In order to obtain a formula for the composition of cubical monomials, we require an explicit expression for the decomposition function of an arbitrary degeneracy of e_{02}. This will be shown using two technical lemmas, which are proved by repeated application of the cubical identities in Definition 5.1.5. While we will need to translate the results into the (I, b) notation for degenerate cubes introduced in Proposition 5.1.1, we begin by stating and proving the lemmas using the degeneracy/connection notation so that we can continue using the composition identities. Recall from Equation (5.1.24) that for subsets $\bar{I} = (i_1, \dots, i_k) \subset [n]$ with $|\bar{I}| = k$, and $\bar{J} = (j_1, \dots, j_m) \subset [n-k-1]$, we define

$$s_{\bar{I}} g_{\bar{J}} b := s_{i_k} s_{i_{k-1}} \dots s_{i_1} g_{j_m} g_{j_{l-1}} \dots g_{j_1} b.$$

We begin with the case where $\bar{I} = \emptyset$, so we only have connection degeneracies.

Lemma 6.4.2. *Let $n \in \mathbb{N}$, and $a = g_{[n-1]} e_{02}$. For $\epsilon = (\epsilon_1, \dots, \epsilon_n) \in \mathbb{Z}_2^n$, let*

$$\epsilon^{-1}(0) = \big(i \,:\, \epsilon_i = 0\big)$$

and $m_\epsilon = (n-1) - |\epsilon^{-1}(0)|$. Then, the decomposition function of a is

$$D[g_{[n-1]} e_{02}](\epsilon) = \begin{cases} g_{[n-1]} e_{01} & : \epsilon = \vec{0} \\ s_{\epsilon^{-1}(0)} g_{[m_\epsilon]} e_{02} & : \text{otherwise} \end{cases} \tag{6.4.7}$$

Proof. This lemma will be shown by iterated applications of the composition identites for connections in Equation (5.1.27), repeated here for convenience.

$$g_i(a +_j b) = \begin{cases} g_i a +_{j+1} g_i b & : i < j \\ (g_i a +_{i+1} s_i b) +_i (s_{i+1} b +_{i+1} g_i b) & : i = j \\ g_i a +_j g_i b & : i > j. \end{cases}$$

We begin with the case of $\epsilon = \vec{0}$. This is immediate from an iterated application of

$$g_j(a +_j b) = (g_j a +_{j+1} s_j b) +_j (s_{j+1} b +_{j+1} g_j b)$$

to the first term of the expression, $(g_j a +_{j+1} s_j b)$, and then taking the first element at the end - this corresponds to the element of summand indexed by $\epsilon = \vec{0}$.

The rest will be done by induction on n. The case of $n = 1$ is given by definition of $e_{02} = e_{01} +_1 e_{12}$, which satisfies our formula. Now, suppose the formula given in the lemma holds for n. Using the composition rule for g_i with $i = j$, we find that

$$\begin{aligned} D[g_{[n]}e_{02}] &= g_n\left(g_{[n-1]}e_{02}\right) \\ &= g_n\Big(D[g_{[n-1]}e_{02}]^0 +_n D[g_{[n-1]}e_{02}]^1\Big) \\ &= \Big(g_n D[g_{[n-1]}e_{02}]^0 +_{n+1} s_n D[g_{[n-1]}e_{02}]^1\Big) \\ &\quad +_n \Big(s_{n+1} D[g_{[n-1]}e_{02}]^1 +_{n+1} g_n D[g_{[n-1]}e_{02}]^1\Big) \end{aligned}$$

Here, $D[g_{[n-1]}e_{02}]^p$ denotes the composition of all elements $D[g_{[n-1]}e_{02}](\epsilon)$ such that $\epsilon_n = p$,

$$D[g_{[n-1]}e_{02}]^p = \bigoplus_{i_1 \in \mathbb{Z}_2}{}_1 \cdots \bigoplus_{i_{n-1} \in \mathbb{Z}_2}{}_{n-1} D[a](i_1, \ldots, i_{n-1}, p).$$

All of the compositions in $D[g_{[n-1]}e_{02}]^p$ are $+_k$, where $k < n$. This implies that the g_n, s_n and s_{n+1}, in the last expression above, factor "all the way though" using the composition rules for $i > j$. Now, for $\epsilon \in \mathbb{Z}_2^{n+1}$, we can write down an explicit formula for $D[g_{[n]}e_{02}](\epsilon)$ in terms of the known decomposition $D[g_{[n-1]}e_{02}]$. For $\epsilon = (\epsilon_1, \ldots, \epsilon_{n+1}) \in \mathbb{Z}_2^{n+1}$, define $\hat{\epsilon}, \hat{\epsilon}^1 \in \mathbb{Z}_2^n$ by

$$\begin{aligned} \hat{\epsilon} &= (\epsilon_1, \ldots, \epsilon_{n-1}, \epsilon_n) \\ \hat{\epsilon}^1 &= (\epsilon_1, \ldots, \epsilon_{n-1}, 1). \end{aligned}$$

The decomposition function can be written as

$$D[g_{[n]}e_{02}](\epsilon) = \begin{cases} g_n D[g_{[n-1]}e_{02}](\hat{\epsilon}) & : (\epsilon_n, \epsilon_{n+1}) = (0,0) \\ s_n D[g_{[n-1]}e_{02}](\hat{\epsilon}^1) & : (\epsilon_n, \epsilon_{n+1}) = (0,1) \\ s_{n+1} D[g_{[n-1]}e_{02}](\hat{\epsilon}) & : (\epsilon_n, \epsilon_{n+1}) = (1,0) \\ g_n D[g_{[n-1]}e_{02}](\hat{\epsilon}) & : (\epsilon_n, \epsilon_{n+1}) = (1,1) \end{cases} \tag{6.4.8}$$

Now, we will use the expanded notation for degeneracies since we will need to work with the cubical identities. Suppose that

$$D[g_{[n-1]}e_{02}](\hat{\epsilon}) = s_{i_k} \dots s_{i_1} g_m \dots g_1 e^{12}.$$

Rewriting the formula in the statement of the lemma in terms of this explicit expression, we get

$$D[g_{[n]}e_{02}](\epsilon) = \begin{cases} s_{n+1}s_{i_k} \dots s_{i_1} g_m \dots g_1 e^{12} & : \epsilon_{n+1} = 0 \\ s_{i_k} \dots, s_{i_1} g_{m+1} g_m \dots g_1 e^{12} & : \epsilon_{n+1} = 1 \end{cases}$$

Thus, to complete the lemma, we need to use the inductive formula in Equation 6.4.8 to show the above expression holds.

- $(\epsilon_n, \epsilon_{n+1}) = (0,0)$. Since $\epsilon_n = 0$, we know that $i_k = n$, so using the fact that $g_n s_n = s_{n+1} s_n$, we get

$$\begin{aligned} D[g_{[n]}e_{02}](\epsilon) &= g_n s_n s_{i_{k-1}} \dots, s_{i_1} g_m \dots g_1 e_{12} \\ &= s_{n+1} s_n s_{i_{k-1}} \dots, s_{i_1} g_m \dots g_1 e_{12}. \end{aligned}$$

- $(\epsilon_n, \epsilon_{n+1}) = (0,1)$. This formula is a bit different since we use $\hat{\epsilon}^1$. Note that

$$D[g_{[n-1]}e_{02}](\hat{\epsilon}^1) = s_{i_{k-1}} \dots s_{i_1} g_{m+1} \dots g_1 e_{12}.$$

  Then, using our inductive decomposition formula, we get

$$\begin{aligned} D[g_{[n]}e_{02}](\epsilon) &= s_n D[g_{[n-1]}e^{02}](\hat{\epsilon}^1) \\ &= s_n s_{i_{k-1}} \dots s_{i_1} g_{m+1} \dots g_1 e_{12}. \end{aligned}$$

- $(\epsilon_n, \epsilon_{n+1}) = (1,0)$. This case is immediate.
- $(\epsilon_n, \epsilon_{n+1}) = (1,1)$. Since $\epsilon_n = 1$, we know that $i_k < n$, so

$$\begin{aligned} D[g_{[n]}e_{02}](\epsilon) &= g_n(s_{i_k} \dots s_{i_1} g_m \dots g_1 e_{02}) \\ &= s_{i_k} \dots s_{i_1} g_{l+1} g_l \dots g_1 e_{12} \end{aligned}$$

  where the second equality is by repeated applications of $g_i s_j = s_j g_{i-1}$ for $i > j$. We do this k times, and so the index on n decreases by k to get $n - k = l + 1$, as desired.

□

Next, we consider the case for arbitrary degeneracies.

Lemma 6.4.3. *Let $n \in \mathbb{N}$. Let $\bar{I} \subset [n]$ such that $|\bar{I}| = k < n$, and let $r = n - k - 1$. For $\epsilon \in \mathbb{Z}_2^n$, let $A = \bar{I} \cup \epsilon^{-1}(0)$, and $m = r - |\epsilon^{-1}(0)|$. Then*

$$D[s_{\bar{I}} g_{[r]} e_{02}](\epsilon) = \begin{cases} s_{\bar{I}} g_{[r]} e_{01} & : \epsilon^{-1}(1) \subseteq \bar{I} \\ s_A g_{[m]} e_{12} & : \text{otherwise} \end{cases}$$

Proof. This will be done by induction on k. The base case of $k = 0$ is the result of Lemma 6.4.2. Thus, we assume that the expression in the lemma is true for k, and we will prove the case of $k+1$. Thus, let $\bar{I} = (i_1, \ldots, i_k, p) \subset [n]$, and since $\bar{I}$ is an ordered subset, $p > i_k$. Next, let

$$a = s_{\hat{I}} g_{[r]} e_{02},$$

where $\hat{I} = (i_1, \ldots, i_k)$, so the n-cube that we wish to decompose is

$$s_p a = s_{i_p} s_{i_k} \ldots s_{i_1} g_r \ldots g_1 e_{02}.$$

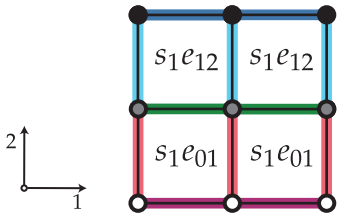
Unlike connections in which $g_i(b +_i c)$ is decomposed into a composition of four cubes, degeneracies $s_i(b +_j c)$ is decomposed as

$$s_i(b +_j c) = \begin{cases} s_i b +_{j+1} s_i c & : i \leq j \\ s_i b +_j s_i c & : i > j \end{cases}$$

Thus, in order to obtain a full decomposition function, we use the fact that $s_p a$ is its own left and right p-unit, and decompose $s_p a$ as

$$s_p a = s_p a +_p s_p a.$$

For example, the decomposition of $s_1(e_{01} +_1 e_{12})$ using our convention is given as follows.



In general, the decomposition function of $s_p a$ is

$$D[s_p a](\epsilon) = s_p D[a](\hat{\epsilon}_p), \tag{6.4.9}$$

where $\epsilon = (\epsilon_1, \ldots, \epsilon_n) \in \mathbb{Z}_2^n$ and $\hat{\epsilon}_p = (\epsilon_1, \ldots, \epsilon_{p-1}, \epsilon_{p+1}, \ldots, \epsilon_n) \in \mathbb{Z}_2^{n-1}$.

If $\epsilon^{-1}(1) \subseteq \bar{I}$, then $\hat{\epsilon}_p^{-1}(1) \subseteq \hat{I}$, so $D[a](\hat{\epsilon}_p) = s_{\hat{I}} g_{[l]} e^{01}$ by the induction hypothesis, and thus $D[s_p a](\epsilon) = s_{\bar{I}} g_{[l]} e^{01}$ by Equation (6.4.9).

Next, suppose that $\epsilon^{-1}(1) \nsubseteq \bar{I}$. Since $p \in \bar{I}$, this implies that $\hat{\epsilon}_p^{-1}(1) \nsubseteq \hat{I}$. Then, $D[a](\hat{\epsilon}_p) = s_A g_{[m]} e^{12}$, where $\hat{A} = \hat{I} \cup \hat{\epsilon}^{-1}(0)$ and $m = l - |\hat{\epsilon}_p^{-1}(0)|$. Thus, by Equation (6.4.9), we have

$$D[s_p a](\epsilon) = s_p s_{\hat{A}} g_{[m]} e^{12}$$

We need to show that this is equal to $s_A g_{[m]} e^{12}$, where $A = \bar{I} \cup \epsilon^{-1}(0)$. In this case, we may have some elements $q \in A$ such that $q > p$; specifically, these come from $\hat{\epsilon}_p^{-1}(0)$. However, the corresponding elements in $\tilde{A}$ must be $q+1$ due to the shifting of indices between ϵ and $\hat{\epsilon}$. By applying $s_i s_j = s_{j+1} s_i$ for $i < j$ to shift the s_p to the right until the degeneracies are in the correct order, we get that $s_p s_{\hat{A}} = s_A$ as desired. □

Finally, we translate Lemma 6.4.3 into the (I, b) notation for degenerate cubes.

Proposition 6.4.2. *Let $I \subset [n]$ such that $I \neq \emptyset$. For $\epsilon = (\epsilon_1, \ldots, \epsilon_n) \in \mathbb{Z}_2^n$, let*

$$\epsilon^{-1}(1) = (i \,:\, \epsilon_i = 1)\,.$$

Then, the decomposition function of the n-cube (I, e_{02}) is

$$D[(I, e_{02})](\epsilon) = \begin{cases} (I, e_{01}) & : I \cap \epsilon^{-1}(1) = \emptyset \\ \left(I \cap \epsilon^{-1}(1), e_{12}\right) & : \text{otherwise.} \end{cases} \tag{6.4.10}$$

Proof. First, we note that given $\bar{I} \subset [n]$ such that $|\bar{I}| = k < n$ and $r = n - k - 1$, we have

$$s_{\bar{I}} g_{[r]} e_{02} = (I, e_{02}),$$

where I is the complement of $\bar{I}$ in $[n]$.

The condition $\epsilon^{-1}(1) \subset \bar{I}$ is equivalent to $I \cap \epsilon^{-1}(1) = \emptyset$ since I is the complement of $\bar{I}$. Next, given $A = \bar{I} \cup \epsilon^{-1}(0)$, the degenerate cube $s_A g_{[m]} e_{12}$ is rewritten as $\left(I \cap \epsilon^{-1}(1), e_{12}\right)$ in the new notation since $I \cap \epsilon^{-1}(1)$ is the complement of A in $[n]$. □

6.4.3 *Chen's Identity*

In this section, we provide the composition law for cubical monomials, and derive Chen's identity as a special case. To begin, we introduce cubical monomials with respect to our modified cubical model of $[0,2]$. Suppose $\gamma^{01}, \gamma^{12} : [0,1] \to \mathbb{R}^N$ and $\gamma^{02} = \gamma^{01} * \gamma^{12} : [0,2] \to \mathbb{R}^N$. We use the formula in Equation (5.3.1) to define the cubical monomials for $L_\bullet$.

Definition 6.4.4. Let $n \in \mathbb{N}$. A level n cubical monomial for $L_\bullet$ is indexed using the ordered sets

$$\mathcal{I} = \big((I_1, e_{\tau_1}), \dots, (I_n, e_{\tau_n})\big)$$
$$P = (p_1, \dots, p_n),$$

where $I_j \subset [n]$ are nonempty subsets, and $\tau_j \in \{01, 12, 02\}$ specifies the nondegenerate cubes. Here, P is a multi-index in $[N]$ which specifies the standard 1-forms dx_p to use. The cubical monomial of γ^{02} with respect to $(\mathcal{I}, P)$ is

$$S_L^{\mathcal{I},P}(\gamma^{02}) \coloneqq \int_{\square^n} \bigwedge_{k=1}^{n} \big(\eta[(I_k, e_{\tau_k})]\big)^* (\gamma^{02})^* (dx_{p_k}) \tag{6.4.11}$$

$$= \int_{\square^n} \bigwedge_{k=1}^{n} \left(\eta_n^{I_k}\right)^* \left((\gamma_{p_k}^{\tau_k})'(t) dt\right). \tag{6.4.12}$$

The subscript in S_L is used to denote the monomials derived using the modified cubical model $L_\bullet$ in order to distinguish them form $S_\square$, the monomials derived using the standard cubical model for $[0,1]$. Note that there are more monomials S_L due to the additional choice of the non-degenerate 1-cells e_τ in the set $\mathcal{I}$. However, when all of the e_{τ_j} are the same in $\mathcal{I}$, the cubical monomials S_L reduce to cubical monomials $S_\square$.

Lemma 6.4.4. *Let $n \in \mathbb{N}$. Suppose*

$$\mathcal{I} = \big((I_1, e_\tau), \dots, (I_n, e_\tau)\big),$$

where $\tau \in \{01, 12, 02\}$. Let $P = (p_1, \dots, p_n)$ be a multi-index valued in $[N]$. Then, if $\tau \in \{01, 12\}$, then

$$S_L^{\mathcal{I},P}(\gamma^{02}) = S_\square^{\mathcal{I},P}(\gamma^\tau), \tag{6.4.13}$$

where $\mathcal{I}$ on the right is interpreted as $\mathcal{I} = (I_1, \dots, I_n)$. If $\tau = 02$, then

$$S_L^{\mathcal{I},P}(\gamma^{02}) = S_\square^{\mathcal{I},P}(\gamma^{02} \circ \phi), \tag{6.4.14}$$

where $\phi : [0,1] \to [0,2]$ is any increasing bijective reparametrization.

Proof. The case of $\tau \in \{01, 12\}$ is immediate from the definition of S_L; in particular Equation (6.4.12). The case of $\tau = 02$ is the same, but we also use reparametrization invariance of the signature. □

In the model $L_\bullet$, the nondegenerate 1-cell e_{02} can be decomposed, and we use this to define the following composition property. Given such an indexing set $\mathcal{I}$, the *decomposition function* of $\mathcal{I}$ is defined as

$$D[\mathcal{I}](\epsilon) = \big(D[(I_1, e_{02})](\epsilon), \dots, D[(I_n, e_{02})]\big)(\epsilon). \tag{6.4.15}$$

Theorem 6.4.1. *Let $n \in \mathbb{N}$, and let*

$$\mathcal{I} = \big((I_1, e_{02}), \dots, (I_n, e_{02})\big)$$
$$P = (p_1, \dots, p_n).$$

Suppose $\gamma^{01}, \gamma^{12} : [0,1] \to \mathbb{R}^N$ and $\gamma^{02} = \gamma^{01} * \gamma^{12} : [0,2] \to \mathbb{R}^N$. Then,

$$S_L^{\mathcal{I},P}(\gamma^{02}) = \sum_{\epsilon \in \mathbb{Z}_2^n} S_L^{D[\mathcal{I}](\epsilon),P}(\gamma^{02}). \tag{6.4.16}$$

Proof. From the definition of S_L in Equation (6.4.11), and using the decomposition of $\square^n$ into subcubes indexed by $\mathbb{Z}_2^n$ defined in Equation (6.4.5), we have

$$\begin{aligned} S_L^{\mathcal{I},P}(\gamma^{02}) &= \int_{\square^n} \bigwedge_{k=1}^n \big(\eta[(I_k, e_{02})]\big)^* (\gamma^{02})^* (dx_{p_k}) \\ &= \sum_{\epsilon \in \mathbb{Z}_2^n} \int_{\square_\epsilon^n} \bigwedge_{k=1}^n \big(\eta[(I_k, e_{02})]\big)^* (\gamma^{02})^* (dx_{p_k}) \end{aligned}$$

Now, restricted to the subcube $\square_\epsilon^n$, the evaluation function

$$\eta[(I_k, e_{02})](t_1, \dots, t_n) = \eta[D[(I_k, e_{02})](\epsilon)](2t_1 - \epsilon_1, \dots, 2t_n - \epsilon_n),$$

by Proposition 6.4.1. Then,

$$\begin{aligned} S_L^{\mathcal{I},P}(\gamma^{02}) &= \sum_{\epsilon \in \mathbb{Z}_2^n} \int_{\square_\epsilon^n} \bigwedge_{k=1}^n \big(\eta[D[(I_k, e_{02})](\epsilon)]\big)^* (\gamma^{02})^* (dx_{p_k}) \\ &= \sum_{\epsilon \in \mathbb{Z}_2^n} \int_{\square^n} \bigwedge_{k=1}^n \big(\eta[D[(I_k, e_{02})](\epsilon)]\big)^* (\gamma^{02})^* (dx_{p_k}) \\ &= \sum_{\epsilon \in \mathbb{Z}_2^n} S_L^{D[\mathcal{I}](\epsilon),P}(\gamma^{02}), \end{aligned}$$

where we performed a change of variables

$$(2t_1 - \epsilon_1, \dots, 2t_n - \epsilon_n) \mapsto (t_1, \dots, t_n)$$

in the second equality to transform the domain of integration from $\square^n_\epsilon$ to $\square^n$. □

This theorem allows us to compute cubical monomials of γ^{02} by using integrals involving γ^{01} and γ^{12}. However, for arbitrary cubical monomials, some of the integrals in the decomposition will involve both γ^{01} and γ^{12}, in contrast with the usual Chen's identity. We will show that in the case where $\mathcal{I}$ is an n-flag, we recover Chen's identity. Let

$$\mathcal{F}^n_\tau := (([1], e_\tau), ([2], e_\tau), \ldots, ([n], e_\tau)), \tag{6.4.17}$$

where $\tau \in \{01, 12, 02\}$. The main step in proving Chen's identity is the following lemma.

Lemma 6.4.5. *Let $n \in \mathbb{N}$, and let $\alpha_k = (0, 0, \ldots, 0_k, 1, 1, \ldots, 1) \in \mathbb{Z}^n_2$, where the subscript denotes the coordinate. In other words, α_k is the element where the first k terms are 0 and all other terms are 1. Then*

$$D[\mathcal{F}^n_{02}](\alpha_k) = \mathcal{F}^k_{01} * \mathcal{F}^{n-k}_{12}. \tag{6.4.18}$$

Furthermore, if $\epsilon \neq \alpha_k$ for any k, then

$$S^{D[\mathcal{F}^n_{02}](\epsilon), P}(\gamma^{02}) = 0 \tag{6.4.19}$$

for all $\gamma^{02} : [0, 2] \to \mathbb{R}^N$.

Proof. In order to prove the first statement, we must consider the decomposition for each element of the flag. From the formula for decomposition functions

$$D[([m], e_{02})](\alpha_k) = \begin{cases} ([m], e_{01}) & : m \leq k \\ ([m-k]+k, e_{12}) & : m > k \end{cases}$$

Putting this together for all m, we get Equation (6.4.18).

Next, suppose that $\epsilon \neq \alpha_k$ for any k. In particular, this implies there exist $j \in [n]$ such that $\epsilon(j) = 1$ and $\epsilon(j+1) = 0$. In this case, we have both $[j] \cap \epsilon^{-1}(1) = [j+1] \cap \epsilon^{-1}(1) \neq \emptyset$. Thus,

$$\begin{aligned} D[([j], e_{02})](\epsilon) &= ([j] \cap \epsilon^{-1}(1), e_{12}) \\ &= ([j+1] \cap \epsilon^{-1}(1), e_{12}) \\ &= D[([j+1], e_{02})](\epsilon). \end{aligned}$$

Therefore, $D[(\mathcal{F}^n_{02}, e_{02})](\epsilon)$ will have two of the same element and therefore any cubical monomial with respect to $D[(\mathcal{F}^n_{02}, e_{02})](\epsilon)$ will be 0. □

Theorem 6.4.2 (Chen's identity). *Let $n \in \mathbb{N}$ and suppose $P = (p_1, \dots, p_n)$ is a multi-index valued in $[N]$. Suppose $\gamma^{01}, \gamma^{12} : [0,1] \to \mathbb{R}^N$ and $\gamma^{02} = \gamma^{01} * \gamma^{12} : [0,2] \to \mathbb{R}^N$. Then,*

$$S_L^{\mathcal{F}^n_{02},P}(\gamma^{02}) = S_\square^{\mathcal{F}^n,P}(\gamma^{01}) + S_\square^{\mathcal{F}^n,P}(\gamma^{12}) + \sum_{k=1}^{n-1} S_\square^{\mathcal{F}^k,(p_1,\dots,p_k)}(\gamma^{01}) S_\square^{\mathcal{F}^{n-k},(p_{k+1},\dots,p_n)}(\gamma^{12}). \tag{6.4.20}$$

Proof. From Theorem 6.4.1 and Lemma 6.4.5, we get

$$S_L^{\mathcal{F}^n_{02},P}(\gamma^{02}) = S_L^{\mathcal{F}^n_{01},P}(\gamma^{02}) + S_L^{\mathcal{F}^n_{01},P}(\gamma^{02}) + \sum_{k=1}^{n-1} S_L^{\mathcal{F}^k_{01} * \mathcal{F}^{n-k}_{12},P}(\gamma^{02}).$$

Next, applying the product structure for cubical monomials in Theorem 5.3.1, we get

$$\begin{aligned}
&\sum_{k=0}^{n} S_L^{\mathcal{F}^k_{01} * \mathcal{F}^{n-k}_{12},P}(\gamma^{02}) \\
&\quad = S_L^{\mathcal{F}^n_{01},P}(\gamma^{02}) + S_L^{\mathcal{F}^n_{01},P}(\gamma^{02}) + \sum_{k=1}^{n-1} S_L^{\mathcal{F}^k_{01},(p_1,\dots,p_k)}(\gamma^{02}) S_L^{\mathcal{F}^{n-k}_{12},(p_{k+1},\dots,p_n)}(\gamma^{02}) \\
&\quad = S_\square^{\mathcal{F}^n,P}(\gamma^{01}) + S_\square^{\mathcal{F}^n,P}(\gamma^{12}) + \sum_{k=1}^{n-1} S_\square^{\mathcal{F}^k,(p_1,\dots,p_k)}(\gamma^{01}) S_\square^{\mathcal{F}^{n-k},(p_{k+1},\dots,p_n)}(\gamma^{12})
\end{aligned}$$

□

7

THE MAPPING SPACE SIGNATURE

In this chapter, we introduce and develop a notion of the *mapping space signature*, a generalization of the path signature to higher dimensional domains. One of the primary motivations for the development of the mapping space signature is the prospect of defining a notion of *rough maps*, a higher dimensional analogue of rough paths, which were briefly introduced in Section 1.4. The theory of rough paths provides a deterministic theory of integration against paths which are highly irregular. Integrating a continuous function $f : [0,1] \to \mathbb{R}$ against a bounded 1-variation path $g : [0,1] \to \mathbb{R}$,

$$\int_0^1 f(x)dg(x), \tag{7.0.1}$$

can be done using the classical Riemann-Stieltjes integral. If we wish to *lower* the regularity of g, then we must compensate by allowing the integrand f to be of *higher* regularity. When f is of bounded p-variation and g is of bounded q-variation, then we can use the Young integral [You36] as long as $1/p + 1/q > 1$ is satisfied. If we require f and g to be of the same regularity, the Young integral is only well defined when both f and g are bounded p-variation, where $p > 2$. In particular, since the second level integral of $\gamma : [0,1] \to \mathbb{R}^N$ with $\gamma(0) = 0$ is given as

$$S^{i,j}(\gamma) = \int_0^1 \gamma_i(t)d\gamma_j(t),$$

the path signature exists only when γ is bounded p-variation with $p < 2$. For integrals with respect to functions of bounded p-variation where $p \geq 2$, classical deterministic theories of integration break down, and one must enrich the paths with additional structure in order to define the desired integrals [LCL07], as discussed in Section 1.4.

In the case of higher dimensional domains, classical integration theories can still be used for integrals with respect to $f, g : \square^d \to \mathbb{R}$. As in the case of paths, if f is continuous and g

is of bounded 1-variation, then the classical multidimensional Riemann-Stieltjes integral can be used to define

$$\int_{\square^d} f(\mathbf{x}) dg(\mathbf{x}). \tag{7.0.2}$$

For the case of $d = 2$, Towghi developed a multidimensional generalization of the Young integral [Tow02] such that if f is of bounded p-variation and g is of bounded q-variation with $1/p + 1/q > 1$, then the integral in Equation (7.0.2) exists. In a related direction, Züst [Züs11] considered integrals of the form

$$\int_{\square^d} f dg_1 \wedge \ldots \wedge dg_d, \tag{7.0.3}$$

where $f, g_1, \ldots, g_d : \square^d \to \mathbb{R}$ are Hölder continuous functions, where f is a-Hölder, and g_i is b_i-Hölder. Züst showed that if $a + b_1 + \ldots + b_d > d$, then the integral is well defined. These integrals of irregular differential forms were further studied by Stepanov and Trevisan [ST20]. Their work reformulated the Züst integral in terms of more general algebraic machinery, extending the algebraic integration theory of Gubinelli [Gub04].

The higher dimensional generalizations of the multidimensional integral discussed thus far do not require any additional information underlying maps, and should be viewed as the multidimensional analogue of Young's theory of integration for paths. Going beyond the scope of these theories, Chouk and Gubinelli [CG14] introduce the notion of a *rough sheet*, a 2-dimensional generalization of rough paths. Their approach is partially successful, and is based on a multidimensional generalization of the algebraic integration theory of Gubinelli [Gub04]. However, their definition of a rough sheet results in a very complicated algebraic structure which is not well understood.

The work in this chapter provides the first steps towards an alternate approach to defining rough sheets and rough maps. In particular, we develop a notion of the *mapping space signature* for smooth maps, and show that it has several properties analogous to the path signature. Our approach is motivated by the observation that the path signature monomials arise as the 0-cochains of Chen's iterated integral cochain model for path spaces. In Chapter 5, we reformulated Chen's mapping space construction [GTZ10] in terms of cubical sets rather than simplicial sets, and in Chapter 6, we showed that the main properties of the path signature can be directly derived from the cochain model.

In this chapter, we will consider the cubical mapping space construction for smooth maps $\mathrm{Diff}(\square^d, \mathbb{R}^N)$, and study the monomials (0-cochains) which arise from this model. In Section 7.1, we define these monomials and consider several explicit examples. We classify the monomials into three categories and focus our attention on one of these classes: the basic monomials. We provide the definition for two different mapping space signatures,

both built from the basic monomials. In the later sections, we find that certain properties are satisfied by one of these signatures but not the other.

In Section 7.2, we discuss the invariance properties of the mapping space signature. Our definition of the monomials is in terms of Jacobian minors of the underlying maps, and thus results in invariance with respect to *Jacobian equivalence classes* of maps. We also discuss permutation invariance (of indices) and reparametrization invariance. We conclude the section by studying maps which are defined as sums of paths. The resulting monomials can be computed using the path signature of the underlying paths, and this leads to a class of paths with trivial mapping space signature.

Next, in Section 7.3, we introduce the shuffle product for the basic monomials, which is a direct generalization of the shuffle product for paths. We discuss two *parametrized* injectivity theorems for the mapping space signature in Section 7.4. When maps in $\mathsf{Diff}(\square^d, \mathbb{R}^N)$ are lifted to maps in $\mathsf{Diff}(\square^d, \mathbb{R}^{d+N})$ by appending their parametrization, we show that two modifications of the signature are injective with respect to the Jacobian equivalence classes in $\mathsf{Diff}(\square^d, \mathbb{R}^N)$, and injective on the entire space of maps. This injectivity result then allows us to show that these two signatures are universal and characteristic with respect to the corresponding space of maps.

Finally, in Section 7.5, we consider the algebraic structure imposed by composition of maps, analogous to Chen's identity for concatenation of paths. However, we find that there is an obstruction to the extension of Chen's identity to higher dimensional domains. While our formula reduces to Chen's identity in the case of $d = 1$, there is an additional integration step required when $d > 1$.

We will begin with the definition of the cubical monomials for both $\mathsf{Diff}(\square^d, \mathbb{R}^N)$ and $\mathsf{Diff}(\square^d, X)$ for an arbitrary smooth manifold X. We will then restrict ourselves to the smooth maps in $\mathsf{Diff}(\square^d, \mathbb{R}^N)$ in order to simplify the notation; though most results in this chapter will still hold for arbitrary manifolds with some modification. While this does not result in a generalization of rough paths, the mapping space signature defined in this chapter is of independent interest. In particular, the smooth hypothesis can easily be weakened to a piecewise smooth hypothesis, which is sufficient for many machine learning applications. We can extend the perspective of the path signature as a feature map for time series data and view the mapping space signature as a feature map for data parametrized over the unit d-cube $\square^d$, such as images, videos, and vector fields. The universal and characteristic properties of the mapping space signature provide strong theoretical motivation for the study of the mapping space signature in data science applications.

7.1 CUBICAL MONOMIALS

In this section, we will define the cubical monomials for $\mathsf{Diff}(\square^d, X)$, where X is an N-dimensional manifold. We begin by defining the cubical structure on $\square^d$, and interpreting the cubical monomials in this context. We classify the cubical monomials into three categories

$$\text{basic monomials} \subset \text{restricted monomials} \subset \text{interior monomials}.$$

The *interior monomials* are all monomials obtained from the mapping space construction using the degeneracies of the single non-degenerate d-cube. While even more monomials exist in the mapping space construction, they use degeneracies of the lower dimensional faces of $\square^d$. Such monomials contain information that is restricted to the boundary of $\square^d$, and because our primary interest is in understanding the entire map $\gamma : \square^d \to X$, we restrict our focus to the monomials coming from the degeneracies of the d-cube.

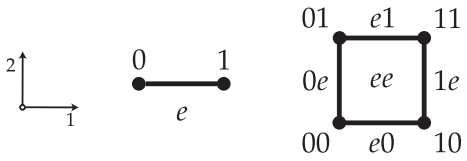
The *restricted monomials* form a subset of the interior monomials such that the domain of integration can be factored into a d-fold product of regions. Then, the *linear basis monomials* further restricts the collection of monomials and forms a subset of the restricted monomials. Any restricted monomial can be written as a linear combination of the linear basis monomials, which motivates the name. This phenomenon is analogous to Theorem 6.2.1, which showed that any cubical monomial for the path space is a linear combination of the monomials indexed by n-flags. Indeed, for the case of $d = 1$, the class of restricted monomials is the same as the class of interior monomials.

Finally, we will use the linear basis monomials to define the *mapping space signature* for $\mathsf{Diff}(\square^d, X)$. The mapping space signature has several properties which are analogous to those of the path signature, and we will discuss these in the following sections of this chapter. Thus, our main focus is on the linear basis monomials; the general and restricted monomials are included for completeness.

7.1.1 *Cubical Model for the d-Cube*

In this section, we will define a cubical set $K^d_\bullet$ modelling the unit d-cube $\square^d$. As in the case of the unit interval in Example 5.1.1 and Example 5.1.2, the definition of the cubical set $K^d_\bullet$ can be given by providing the non-degenerate cubes along with their boundary maps. Let $E = \{0, 1, e\}$. The non-degenerate cubes in $K^d_\bullet$ are indexed by E^d, which is shown for $d = 1$ and $d = 2$ below. We will denote elements of E^d as words $a_1 \dots a_d$, where $a_i \in E$.

The *dimension* of an element $a \in E^d$ is the number of e's in the word a. We will denote the unique non-degenerate d cube by e^d. The face maps can also be easily defined using this notation. Let $a = a_1 \dots a_d \in E^d$ be a non-degenerate cube of dimension k, so there are



k instances of the letter e in the word a. Let $i < k$, and suppose $j \in [d]$ is the index of the i^{th} instance of e in a. Then the face maps for the nondegenerate cubes are defined by

$$d_i^0(a) = a_0 \dots a_{j-1} 0 a_{j+1} \dots a_d$$
$$d_i^1(a) = a_0 \dots a_{j-1} 1 a_{j+1} \dots a_d.$$

Note in particular that $d_i^e(a)$ will be of dimension $k-1$. Furthermore, if $a \in E^d$ is of dimension k, the evaluation map $\eta[a] : \square^k \to \square^d$ can also be easily read off of the notation. All of the e's in the word a represent the dimensions to parametrize, and the 1's and 0's specify the coordinates in those dimensions. For example, given $0e1e \in E^4$, the evaluation map $\eta[0e1e] : \square^2 \to \square^4$ is defined by

$$\eta[0e1e](t_1, t_2) = (0, t_1, 1, t_2).$$

The degeneracies of these cubes can be denoted using the (I, b) notation given in Proposition 5.1.1 and the evaluation map for (I, b) can be computed using its factored form as $\eta[b] \circ \eta_n^I$, where η_n^I is given in Equation (5.1.34). We will primarily be focused on the non-degenerate d-cube e^d, and in this case, the evaluation map $\eta[e^d] : \square^d \to \square^d$ is simply the identity.

7.1.2 *Interior Monomials*

Now that we have a cubical model for $\square^d$, we can define the cubical monomials for $\mathrm{Diff}(\square^d, X)$. Recall from Definition 5.3.1 that a level n cubical monomial is indexed by

$$\mathcal{I} = \big((I_1, b_1), \dots, (I_m, b_m)\big)$$
$$P = (\omega_1, \dots, \omega_m).$$

Each (I_j, b_j) specifies an n-cube in K_n^d, where $b_j \in E^d$ is a non-degenerate cube of dimension q_j, and I_j is a q_j-ordered subset of $[n]$, where $\sum_{j=1}^m q_j = n$. Furthermore, each $\omega_j \in \Omega^{q_j}(X)$. Given $\gamma \in \mathsf{Diff}(\square^d, X)$, the cubical monomial of γ with respect to $(\mathcal{I}, P)$ is

$$M^{\mathcal{I},P}(\gamma) = \int_{\square^n} \bigwedge_{k=1}^{m} \left(\eta[b_k] \circ \eta_n^{I_k}\right)^* (\gamma^* \omega_k).$$

As we mentioned in the introduction, the collection of all monomials for mapping spaces $\mathsf{Diff}(\square^d, \mathbb{R}^N)$ where $d > 1$ is much larger than the case of the path space with $d = 1$, due to the existence of many more non-degenerate cubes. If $b \in E^d$ is a nondegenerate cube of dimension $k < d$, then the evaluation map for a degeneracy (I, b) of b maps onto the boundary of $\square^d$. Thus, the pullback $(\eta[(I, b)])^*(\gamma^* \omega)$ results in a form which only contains information about the boundary. Because we wish to study the entire map γ, including the interior, we begin by considering monomials where $\mathcal{I}$ only contains degeneracies of the top dimensional cube e^d.

In other words, we set $b_j = e^d$ for all j, and we will denote $\mathcal{I} = (I_1, \ldots, I_m)$, where I_j is a d-ordered subset. Note that this places restrictions on the possible levels of monomials considered. In particular, since the dimension of each non-degenerate cube is d, we only consider monomials of level md. This leads to our definition of the interior monomials.

Definition 7.1.1. Let $m \in \mathbb{N}$ and $n = md$. Let

$$\mathcal{I} = (I_1, \ldots, I_m)$$
$$P = (\omega_1, \ldots, \omega_m),$$

where each I_j is a d-ordered subset of $[n]$ and $\omega_j \in \Omega^d(X)$. For $\gamma : \square^d \to X$, the *interior monomial of γ with resepct to $(\mathcal{I}, P)$* is defined by

$$M_{\text{int}}^{\mathcal{I},P}(\gamma) = \int_{\square^n} \bigwedge_{k=1}^{m} \left(\eta_n^{I_k}\right)^* (\gamma^* \omega_k). \tag{7.1.1}$$

In the case where $X = \mathbb{R}^N$, we have a collection of standard d-forms,

$$\omega_P = dx_{p_1} \wedge \ldots \wedge dx_{p_d},$$

which can be described using a subset $P = (p_1, \ldots, p_d) \subset [N]$. For such a subset and a map $\gamma = (\gamma_1, \ldots, \gamma_N) : \square^d \to \mathbb{R}^N$, we let

$$\gamma_P = (\gamma_{p_1}, \ldots, \gamma_{p_d}) \tag{7.1.2}$$

be the projection onto the coordinates specified by P. The pullback of the standard d-form ω_P along the map γ is

$$\gamma^*(\omega_P) = J[\gamma_P]ds_1 \wedge \ldots \wedge ds_d \in \Omega^d(\square^d), \tag{7.1.3}$$

where $J[\gamma_P] : \square^d \to \mathbb{R}$ is the *Jacobian determinant of* γ_P,

$$J[\gamma_P] = \begin{vmatrix} \frac{\partial \gamma_{p_1}}{\partial s_1} & \ldots & \frac{\partial \gamma_{p_1}}{\partial s_d} \\ \vdots & \ddots & \vdots \\ \frac{\partial \gamma_{p_d}}{\partial s_1} & \ldots & \frac{\partial \gamma_{p_d}}{\partial s_d}. \end{vmatrix} \tag{7.1.4}$$

We can give a more precise definition of the interior monomials for $X = \mathbb{R}^N$.

Definition 7.1.2. Let $m \in \mathbb{N}$ and $n = md$. Let

$$\mathcal{I} = (I_1, \ldots, I_m)$$
$$\mathcal{P} = (P_1, \ldots, P_m)$$

where each I_j is a d-ordered subset of $[n]$ and $P_i \subset [N]$ is an ordered subset of length d. For $\gamma : \square^d \to \mathbb{R}^N$, the *interior monomial of* γ *with respect to* $(\mathcal{I}, \mathcal{P})$ is defined by

$$M_{\text{int}}^{\mathcal{I},\mathcal{P}}(\gamma) = \int_{\square^n} \bigwedge_{k=1}^{m} \left(\eta_n^{I_k}\right)^* (J[\gamma_{P_k}]ds_1 \wedge \ldots \wedge ds_d). \tag{7.1.5}$$

Note that we will always use $(s_1, \ldots, s_d) \in \square^d$ to parametrize the domain of the mapping space, and use $(t_1, \ldots, t_n) \in \square^n$ to parametrize the domain of integration.

7.1.3 *Examples*

In order to gain a better understanding of these monomials, we begin with several examples with $X = \mathbb{R}^N$.

Example 7.1.1. We begin with the example where $n = d$, the case of an interior monomial for $\mathsf{Diff}(\square^d, \mathbb{R}^N)$. In this case, both $\mathcal{I} = (I)$ and $\mathcal{P} = (P)$ have a single element. Recall from Definition 5.1.9 that a d-ordered subset I is an ordered subset $I \subset [n]$ equipped with a partition into d nonempty subsets. Because $n = d$, we must take $I = [d]$, and the subsets in the partitions are the singletons $I^{(k)} = \{k\}$. With this choice of I, the evaluation map $\eta_d^I : \square^d \to \square^d$ is the identity.

Now let $P \subset [N]$ be a subset of length d and $\gamma : \square^d \to \mathbb{R}^N$. The interior monomial of γ with respect to $(\mathcal{I}, \mathcal{P})$ is

$$M_{\text{int}}^{\mathcal{I},\mathcal{P}}(\gamma) = \int_{\square^d} J[\gamma_P](t_1, \dots, t_d) dt_1 \dots dt_d.$$

This is the signed d-volume of the map γ_P, the projection of γ onto the coordinates specified by P. This is analogous to the level 1 signature monomials, which provides the signed displacement of a path along a given coordinate.

Example 7.1.2. Next, we consider the case of $n = 2d$, where $\mathcal{I} = (I_1, I_2)$ and $\mathcal{P} = (P_1, P_2)$ each have two elements. In this case, we choose $I_1 = [d]$ partitioned by singletons $I_1^k = \{k\}$, and $I_2 = [d] + d$ also partitioned by singletons $I_2^{(k)} = \{k + d\}$. The evaluation maps $\eta_{2d}^{I_1}, \eta_{2d}^{I_2} : \square^{2d} \to \square^d$ are given by projections on to the first and last d coordinates respectively. Given two subsets $P_1, P_2 \subset [N]$, each of length d, and $\gamma : \square^d \to \mathbb{R}^N$, the interior monomial of γ with respect to $(\mathcal{I}, \mathcal{P})$ is

$$\begin{aligned} M_{\text{int}}^{\mathcal{I},\mathcal{P}}(\gamma) &= \int_{\square^{2d}} J[\gamma_{P_1}](t_1, \dots, t_d) J[\gamma_{P_2}](t_{d+1}, \dots, t_{2d}) dt_1 \dots dt_{2d} \\ &= \left(\int_{\square^d} J[\gamma_{P_1}](t_1, \dots, t_d) dt_1 \dots dt_d \right) \cdot \left(\int_{\square^d} J[\gamma_{P_2}](t_1, \dots, t_d) dt_1 \dots dt_d \right) \end{aligned}$$

In this case, we get a product of level d interior monomials analogous to Example 6.1.1 in the case of $d = 1$. We can extend these types of monomials to level $n = md$ by choosing $I_k = [d] + (k-1)d$ partitioned by singletons to obtain monomials analogous to Example 6.1.2.

Example 7.1.3. Now we consider a specific case for $d = 2$ and $n = 2d = 4$. In this case, we have $\mathcal{I} = (I_1, I_2)$ and $\mathcal{P} = (P_1, P_2)$. We let $I_1 = (1, 2, 3)$ with the partition defined by $I_1^{(1)} = \{1\}$ and $I_1^{(2)} = \{2, 3\}$; and let $I_2 = (1, 2, 3, 4)$ with partition defined by $I_2^{(1)} = \{1, 2\}$ and $I_2^{(2)} = \{3, 4\}$. In this case,

$$(\eta_4^{I_1})^*(J[\gamma_{P_1}] ds_1 \wedge ds_2) = \begin{cases} J[\gamma_{P_1}](t_1, t_2) dt_1 \wedge dt_2 & : t_3 < t_2 \\ J[\gamma_{P_1}](t_1, t_3) dt_1 \wedge dt_3 & : t_2 < t_3 \end{cases}$$

$$(\eta_4^{I_2})^*(J[\gamma_{P_2}] ds_1 \wedge ds_2) = \begin{cases} J[\gamma_{P_2}](t_1, t_3) dt_1 \wedge dt_3 & : t_2 < t_1,\ t_4 < t_3 \\ J[\gamma_{P_2}](t_2, t_3) dt_2 \wedge dt_3 & : t_1 < t_2,\ t_4 < t_3 \\ J[\gamma_{P_2}](t_1, t_4) dt_1 \wedge dt_4 & : t_2 < t_1,\ t_3 < t_4 \\ J[\gamma_{P_2}](t_2, t_4) dt_2 \wedge dt_4 & : t_1 < t_2,\ t_3 < t_4. \end{cases}$$

One can check that the only subsimplex of $\square^4$ such that the wedge product of these two forms is nontrivial is $\Delta^4 = \{0 \leq t_1 < t_2 < t_3 < t_4 \leq 1\}$. The interior monomial of $\gamma : \square^2 \to \mathbb{R}^N$ with respect to $(\mathcal{I}, \mathcal{P})$ is

$$M_{\text{int}}^{\mathcal{I},\mathcal{P}}(\gamma) = \int_{\Delta^4} J[\gamma_{P_1}](t_1, t_3) J[\gamma_{P_2}](t_2, t_4) dt_1 dt_2 dt_3 dt_4. \tag{7.1.6}$$

Note that the ordering of coordinates in the Jacobian determinants is not consistent with the ordering of coordinates in the subsimplex. In fact, due to the fact that the subsets in the partition of I must be *ordered*, there does not exist a choice of $\mathcal{I}$ such that the integral has the form

$$\int_{\Delta^4} J[\gamma_{P_1}](t_1, t_2) J[\gamma_{P_2}](t_3, t_4) dt_1 dt_2 dt_3 dt_4.$$

Nevertheless, Equation (7.1.6) may seem to be an appropriate analogue of the cubical path signature monomials indexed by 2-flags such as in Example 6.1.4 since the integrand is nonzero only on a single subsimplex of $\square^4$. While we can generalize Equation (7.1.6) to higher levels, we argue that this is not the appropriate generalization. Indeed, a key element of the definition of the path signature is its recursive definition in Equation (1.2.11) which is not possible for monomials of this form. For example, it is impossible to define $M_{\text{int}}^{\mathcal{I},\mathcal{P}}(\gamma)$ as an integral of the level 2 monomial from Example 7.1.1.

Example 7.1.4. We consider another example for $d = 2$ and $n = 2d = 4$. In this case, we take $I_1 = (1, 3)$ partitioned by singletons $I_1^{(1)} = \{1\}$ and $I_1^{(2)} = \{3\}$, and $I_2 = (1, 2, 3, 4)$ partitioned by the subsets $I_2^{(1)} = \{1, 2\}$ and $I_2^{(2)} = \{3, 4\}$. By writing out the pullbacks along the evaluation maps and considering the subregion of $\square^4$ with a nonzero integrand, the interior monomial of $\gamma : \square^2 \to \mathbb{R}^N$ with respect to $(\mathcal{I}, \mathcal{P})$ is

$$M_{\text{int}}^{\mathcal{I},\mathcal{P}}(\gamma) = \int_{\Delta^2 \times \Delta^2} J[\gamma_{P_1}](t_1, t_3) J[\gamma_{P_2}](t_2, t_4) dt_1 dt_2 dt_3 dt_4,$$

where $\Delta^2 \times \Delta^2 = \{0 \leq t_1 < t_2 \leq 1\} \times \{0 \leq t_3 < t_4 \leq 1\} \subset \square^4$. As opposed to the previous example, this monomial can be defined as an integral of the monomial from Example 7.1.1. Indeed,

$$\begin{aligned} M_{\text{int}}^{\mathcal{I},\mathcal{P}}(\gamma) &= \int_{(t_2,t_4) \in \square^2} \left(\int_{(t_1,t_3) \in [0,t_2] \times [0,t_4]} J[\gamma_{P_1}](t_1, t_3) dt_1 dt_3 \right) J[\gamma_{P_2}](t_2, t_4) dt_2 dt_4 \\ &= \int_{(t'_1,t'_2) \in \square^2} M_{\text{int}}^{I'_1,P_1}(\gamma)_{t'_1,t'_2} J[\gamma_{P_2}](t'_1, t'_2) dt'_1 dt'_2, \end{aligned}$$

where we have renamed the variables in the second equality. Here, $I_1' = (1,2)$ partitioned by singletons as in Example 7.1.1, and

$$M_{\text{int}}^{I_1',P_1}(\gamma)_{t_1',t_2'} = \int_{(t_1,t_2)\in[0,t_1']\times[0,t_2']} J[\gamma_{P_1}](t_1,t_2)dt_1dt_2$$

is the cubical monomial of γ restricted to $[0,t_1'] \times [0,t_2']$.

7.1.4 *Restricted and Basic Monomials*

In the previous subsection, we considered several explicit examples of interior monomials, and in this subsection, we will classify these monomials into two subcategories: restricted monomials and basic monomials.

From the previous examples, we find that the domain of integration (where the integrand is nonzero) is determined by the choice of $\mathcal{I}$. In most cases, the nonzero domain of integration $D \subset \square^{md}$ for a level $n = md$ interior monomial could be decomposed into a product of d subsets of $\square^m$

$$D = D_1 \times \ldots \times D_d \subset \square^{md}, \tag{7.1.7}$$

where $D_j \subset \square^m$. Note that Example 7.1.3 does not satisfy this property. We will be interested in certain interior monomials whose nonzero domains of integration satisfy this decomposition property.

We can ensure the decomposition of the nonzero domain of integration by restricting the possible d-ordered subsets in $\mathcal{I} = (I_1, \ldots, I_m)$. At level $n = md$, we restrict the subsets in the partition of each I_j such that they satisfy

$$I_j^{(k)} \subset [m] + (k-1)m = \{k(m-1)+1, \ldots, km\}. \tag{7.1.8}$$

Lemma 7.1.1. *Let $\mathcal{I} = (I_1, \ldots, I_m)$ be a collection of d-ordered subsets which satisfy Equation (7.1.8). The nonzero domain of integration of an interior monomial with respect to $\mathcal{I}$ satisfies the decomposition in Equation (7.1.7).*

Proof. Recall that a subsimplex $\Delta_\sigma^{md} \subset \square^{md}$, where $\sigma \in \Sigma_{md}$, results in a trivial integrand when two components from any two of the evaluation maps $\eta_{md}^{I_k}$ take the same value. More specifically, this occurs when two of the maximum functions in the definition of the evaluation map take the same value. Given the restriction in Equation (7.1.8) on the allowed partitions of each I_j, we note the coordinates t_a and t_b of $\square^{md}$ never appear in the same maximum function in any of the evaluation maps if $a \in [m] + (k_1 - 1)m$ and $b \in [m] + (k_2 - 1)m$ for different k_1, k_2. □

Now that we have factored the domain of integration into a product of smaller domains, it is convenient to relabel our coordinates. In particular, we make the change of variables

$$t_{k(m-1)+j} \mapsto t_{j,k}, \tag{7.1.9}$$

where $j \in [m]$ and $k \in [d]$, in the parametrization of $\square^{md}$. Then, in the decomposition $D = D_1 \times \ldots \times D_d \subset \square^{md}$, we use $(t_{1,k}, \ldots, t_{m,k})$ to parametrize D_k. Furthermore, we will denote $\mathcal{I}$ in terms of the subsets of its partitions rather than d-ordered subsets in order to relate it to the cubical path signature monomials. Given a collection $\mathcal{I} = (I_1, \ldots, I_m)$ of d-ordered subsets, we define

$$\mathcal{I}^{(k)} = (I_1^{(k)}, \ldots, I_m^{(k)}).$$

However, due to the change of coordinates in Equation (7.1.9), each $\mathcal{I}^{(k)}$ can be simply described as a subset $\mathcal{I}^{(k)} \subset 2^{[m]}$. These modifications lead to the definition of the restricted monomials.

Definition 7.1.3. Let $m \in \mathbb{N}$ and $n = md$. Let

$$\mathcal{I} = (\mathcal{I}^{(1)}, \ldots, \mathcal{I}^{(d)})$$
$$P = (\omega_1, \ldots, \omega_m),$$

where $\mathcal{I}^{(k)} = (I_1^{(k)}, \ldots, I_m^{(k)}) \subset 2^{[m]}$ such that $\emptyset \notin \mathcal{I}^{(k)}$ and $\omega_j \in \Omega^d(X)$. The *restricted monomial of* $\gamma : \square^d \to X$ *with respect to* $(\mathcal{I}, P)$ is

$$M_{\text{res}}^{\mathcal{I},P} = \int_{\square^{md}} \bigwedge_{j=1}^{m} \gamma^*(\omega_j)(s_{j,1}, \ldots, s_{j,d}), \tag{7.1.10}$$

where

$$s_{j,k} = \max_{i \in I_j^{(k)}} \{t_{i,k}\}. \tag{7.1.11}$$

In the case of $X = \mathbb{R}^N$, we can provide a more explicit definition using standard d-forms and Jacobian determinants.

Definition 7.1.4. Let $m \in \mathbb{N}$ and $n = md$. Let

$$\mathcal{I} = (\mathcal{I}^{(1)}, \ldots, \mathcal{I}^{(d)})$$
$$\mathcal{P} = (P_1, \ldots, P_m),$$

	$S_{\square}^{\mathcal{I},P}$	$M_{\text{res}}^{\mathcal{I},\mathcal{P}}$
level	n	$n = md$
degeneracy index	$\mathcal{I} \subset 2^{[n]}$	$\mathcal{I} = (\mathcal{I}^{(1)}, \dots, \mathcal{I}^{(d)}), \quad \mathcal{I}^{(k)} \subset 2^{[m]}$
forms index	$P = (p_1, \dots, p_n), \quad p_j \in [N]$	$\mathcal{P} = (P_1, \dots, P_m), \quad P_j \subset [N]$
integrand	$\gamma'_{p_j}(s_j)$	$J[\gamma_{P_j}](s_{j,1}, \dots, s_{j,d})$

Table 7.1: Comparison between the cubical path signature monomials and the restricted cubical monomials for mapping spaces.

where $\mathcal{I}^{(k)} = (I_1^{(k)}, \dots, I_m^{(k)}) \subset 2^{[m]}$ such that $\emptyset \notin \mathcal{I}^{(k)}$ and $P_j \subset [N]$ is a subset of length d. The *restricted monomial of* $\gamma : \square^d \to \mathbb{R}^N$ *with respect to* $(\mathcal{I}, \mathcal{P})$ is

$$M_{\text{res}}^{\mathcal{I},\mathcal{P}} = \int_{\square^{md}} \prod_{j=1}^{m} J[\gamma_{P_j}](s_{j,1}, \dots, s_{j,d}) d\mathbf{s} \tag{7.1.12}$$

where $s_{j,k}$ is defined in Equation (7.1.11).

Note the similarity between this definition of the restricted monomials $M_{\text{res}}^{\mathcal{I},\mathcal{P}}$ for $\mathsf{Diff}(\square^d, \mathbb{R}^N)$ and Definition 6.1.1 of the cubical monomials $S_{\square}^{\mathcal{I},P}$ for $\mathsf{Diff}(\square^1, \mathbb{R}^N)$.

The first observation to make in this table is that the integer m in the restricted monomials $M_{\text{res}}^{\mathcal{I},\mathcal{P}}$ coincides more closely with the level n in the cubical path signature monomials $S_{\square}^{cI,P}$. Then, rather than a single collection of subsets $\mathcal{I} \subset 2^{[n]}$ in the definition of $S_{\square}^{\mathcal{I},P}$, we use d collections of subsets $\mathcal{I}^{(k)} \subset 2^{[m]}$ in $\mathcal{M}_{\text{res}}^{\mathcal{I},\mathcal{P}}$. Instead of choosing n 1-forms indexed using integers $p_j \in [N]$, we choose m d-forms indexed using subsets $P_j \subset [N]$. Finally, in both cases, the integrand consists of the Jacobian determinant of γ projected onto the coordinate p_j and the coordinates specified by P_j, using the maximum coordinates s_j or $s_{j,k}$. Thus, we can think of $M_{\text{res}}^{\mathcal{I},\mathcal{P}}$ as the analogue of $S_{\square}^{\mathcal{I},P}$ for $d > 1$.

For the cubical path signature monomials, we considered a further restriction to the case where $\mathcal{I} = \mathcal{F}^n$, the n-flag. We can consider the analogous degeneracy index for the mapping space monomials. In particular, suppose we take

$$\mathcal{I} = (\mathcal{F}^m, \dots, \mathcal{F}^m).$$

In this case, the restricted monomial is

$$M_{\text{res}}^{\mathcal{I},\mathcal{P}}(\gamma) = \int_{\Delta^m \times \dots \times \Delta^m} \prod_{k=1}^{m} J[\gamma_{P_j}](t_{j,1}, \dots, t_{j,d}) d\mathbf{t}, \tag{7.1.13}$$

an expression that generalizes the simplicial path signature monomials. However, in the higher dimensional setting, there are more monomials of this form that we must consider.

In particular, we will allow each flag $\mathcal{F}^m$ to be permuted. Let $\pi_1, \dots, \pi_d \in \Sigma_m$. The action of Σ_m on $\mathcal{F}^m$ is given by permuting the elements of the ordered set. Explicitly, the action is defined as

$$\pi_k \mathcal{F}^m = \big([\pi_k(1)], [\pi_k(2)], \dots, [\pi_k(m)]\big).$$

Then, we take

$$\mathcal{I} = (\pi_1 \mathcal{F}^m, \dots, \pi_d \mathcal{F}^m),$$

Then, the restricted monomial of $\gamma : \square^d \to \mathbb{R}^N$ with respect to $(\mathcal{I}, \mathcal{P})$ is

$$M_{\text{res}}^{\mathcal{I},\mathcal{P}}(\gamma) = \int_{\Delta^m \times \dots \Delta^m} \prod_{j=1}^{m} J[\gamma_{P_j}](t_{\pi_1(j),1}, t_{\pi_2(j),2}, \dots, t_{\pi_d(j),d}) d\mathbf{t}. \tag{7.1.14}$$

By performing the change of variables

$$t_{\tau_k(j),k} \mapsto t_{j,k},$$

we can rewrite the integral as

$$M_{\text{res}}^{\mathcal{I},\mathcal{P}}(\gamma) = \int_{\Delta^m_{\pi_1^{-1}} \times \dots \Delta^m_{\pi_k^{-1}}} \prod_{k=1}^{m} J[\gamma_{P_j}](t_{j,1}, \dots, t_{j,d}) d\mathbf{t}. \tag{7.1.15}$$

These monomials are the *basic monomials*, and we will index these monomials using the permutations π_k rather than the degeneracy index $\mathcal{I}$. Given the permutations $\Pi = (\pi_1, \dots, \pi_d) \in \Sigma_m^d$, we define the domain

$$D_{m,d}(\Pi) \coloneqq \Delta^m_{\pi_1} \times \Delta^m_{\pi_2} \times \dots \times \Delta^m_{\pi_d}. \tag{7.1.16}$$

Definition 7.1.5. Let $m \in \mathbb{N}$ and $n = md$. Let

$$\Pi = (\pi_1, \dots, \pi_d) \in \Sigma_m^d$$
$$P = (\omega_1, \dots, \omega_m),$$

where $\omega_j \in \Omega^d(X)$. The *basic monomial of* $\gamma : \square^d \to X$ *with respect to* (Π, P) is

$$M_{\text{bas}}^{\mathcal{I},P} = \int_{D_{m,d}(\Pi)} \bigwedge_{j=1}^{m} \gamma^*(\omega_j)(t_{j,1}, \dots, t_{j,d}). \tag{7.1.17}$$

Note that if we were to write the restricted monomial in Equation (7.1.15) as a basic monomial, we would use Π^{-1} as the index. Once again, the case of $X = \mathbb{R}^N$ can be written more explicitly using Jacobian determiants.

Definition 7.1.6. Let $m \in \mathbb{N}$ and $n = md$. Let

$$\Pi = (\pi_1, \dots, \pi_d) \in \Sigma_m^d$$
$$\mathcal{P} = (P_1, \dots, P_m),$$

where $P_j \subset [N]$ is a subset of length d. The *basic monomial of* $\gamma : \square^d \to \mathbb{R}^N$ *with respect to* $(\Pi, \mathcal{P})$ is

$$M_{\text{bas}}^{\Pi,\mathcal{P}} = \int_{D_{m,d}(\Pi)} \prod_{j=1}^{m} J[\gamma_{P_j}](t_{j,1}, \dots, t_{j,d}). \tag{7.1.18}$$

We will use the basic monomials as the generalization of the simplicial monomials for paths, and we will use them to define the mapping space signature in the next section. Before moving on, we show that any restricted monomial can be written as a linear combination of basic monomials, providing the mapping space analogue of Theorem 6.2.1.

The proof of this fact follows that of Theorem 6.2.1; the main difference is the domain of integration. In the case of $d = 1$, we began by considering the nonzero domains $\Delta_\sigma^n \subset \square^n$, where $\sigma \in \Sigma_n$. In the current case of $d > 1$, we must consider the collection of subregions $D_{m,d}(\Pi)$, where $\Pi \in \Sigma_m^d$ such that the integrand of the restricted monomial is zero. Note that the subregion $D_{m,d}(\Pi)$ is the analogue of Δ_σ^n for mapping spaces: for each fixed k, the order of the variables $\{t_{j,k}\}_{j=1}^m$ is fixed, and thus all maximum functions are fixed projection maps within $D_{m,d}(\Pi)$. The exact projection maps can be determined by a generalization of the maximal coordinates in Definition 6.2.2.

Definition 7.1.7. Let $m \in \mathbb{N}$ and $n = md$. Let

$$\mathcal{I} = (\mathcal{I}^{(1)}, \dots, \mathcal{I}^{(d)})$$

where $\mathcal{I}^{(k)} = (I_1^{(k)}, \dots, I_m^{(k)}) \subset 2^{[m]}$ such that $\emptyset \notin \mathcal{I}^{(k)}$, and let $\Pi = (\pi_1, \dots, \pi_d) \in \Sigma_m^d$. The *maximal coordinates with respect to* $\mathcal{I}$ *and* Π is

$$\mu(\mathcal{I}, \Pi) \coloneqq \big(\mu(\mathcal{I}^{(1)}, \pi_1), \dots, \mu(\mathcal{I}^{(d)}, \pi_d)\big), \tag{7.1.19}$$

where each $\mu(\mathcal{I}^{(k)}, \pi_k)$ is defined in Equation (6.2.2).

When $\mathcal{I}$ and Π are clear, we will denote $\mu(\mathcal{I}, \Pi)$ by $\mu = (\mu_1, \dots, \mu_d)$ to simplify notation. Given $\mathcal{I}$ and Π, the $\mathbf{s}$ variables on $D_{d,m}(\Pi)$ can be written explicitly as

$$s_{j,k} = t_{\mu_k(j),k}.$$

Analogous to the case of $d = 1$, if two maximum coordinates $s_{j,k}$ and $s_{j',k}$ take the same value on $D_{m,d}(\Pi)$, then the integrand in this domain will be 0 due to a $dt \wedge dt$ term. Therefore, in order for the integrand to be nontrivial on $D_{m,d}(\Pi)$, each of the maximal coordinates $\mu(\mathcal{I}^{(k)}, \pi_k)$ must contain distinct elements.

Definition 7.1.8. Let $m \in \mathbb{N}$ and $n = md$. Let

$$\mathcal{I} = (\mathcal{I}^{(1)}, \dots, \mathcal{I}^{(d)})$$

where $\mathcal{I}^{(k)} = (I_1^{(k)}, \dots, I_m^{(k)}) \subset 2^{[m]}$ such that $\emptyset \notin \mathcal{I}^{(k)}$, and let $\Pi = (\pi_1, \dots, \pi_k) \in \Sigma_m^d$. The subregion $D_{d,m} \subset \square^{dm}$ is a *nontrivial domain* if each $\mu(\mathcal{I}^{(k)}, \pi_k)$ contains distinct elements, and thus $\mu(\mathcal{I}, \Pi) \in \Sigma_m^d$.

The collection of $\Pi \in \Sigma_m^d$ such that $D_{m,d}(\Pi)$ is a nonzero domain of integration can be enumerated using products of maximal chains. Let $\mathcal{I} = (\mathcal{I}^{(1)}, \dots, \mathcal{I}^{(d)})$ such that each $\mathcal{I}^{(k)} \subset 2^{[m]}$ is an ordered subset of length m. Then, recall from Definition 6.2.5 that $\mathcal{M}(\mathcal{I}^{(k)})$ is the set of maximal chains in $U(\mathcal{I}^{(k)})$. Then, since $\mathcal{I}$ is a collection of these subsets, we define

$$\mathcal{M}(\mathcal{I}) \coloneqq \mathcal{M}(\mathcal{I}^{(1)}) \times \ldots \times \mathcal{M}(\mathcal{I}^{(d)}). \tag{7.1.20}$$

Each maximal chain in $\mathcal{M}(\mathcal{I}(k))$ corresponds to a permutation $\sigma \in \Sigma_m$. Therefore, we can view $\mathcal{M}(\mathcal{I})$ as a subset of Σ_m^d.

Lemma 7.1.2. *Let*

$$\mathcal{I} = (\mathcal{I}^{(1)}, \dots, \mathcal{I}^{(m)})$$
$$P = (\omega_1, \dots, \omega_d)$$

such that each $\mathcal{I}^{(k)} \subset 2^{[m]}$ is an ordered subset of length m and $\omega_j \in \Omega^d(X)$. Let $\Pi = (\pi_1, \dots, \pi_d) \in \Sigma_m^d$. There exists a map $\gamma : \square^d \to X$ such that the integrand of $M_{\text{res}}^{\mathcal{I},P}(\gamma)$,

$$\bigwedge_{j=1}^{m} \gamma^*(\omega_j)(s_{j,1}, \dots, s_{j,d}), \quad s_{j,k} = \max_{i \in I_j^{(k)}} \{t_{i,k}\},$$

is nonzero on $D_{d,m}(\Pi)$ if and only if $\Pi \in \mathcal{M}(\mathcal{I})$.

Proof. The integrand is generically nonzero if and only if each $s_{j,k}$ is mapped to a distinct $t_{j',k}$; otherwise we would have a $dt_{j,k} \wedge dt_{j,k} = 0$ term in the integrand. In other words, the integrand is generically nonzero if each μ_k contains distinct elements, and thus $\mu_k \in \Sigma_m$ for all k. From Lemma 6.2.2 and Lemma 6.2.3, $\mu_k \in \Sigma_m$ if and only if $\pi_k \in \mathcal{M}(\mathcal{I}^{(k)})$. Thus, $D_{d,m}(\Pi)$ is a nonzero domain of integration if and only if $\Pi \in \mathcal{M}(\mathcal{I})$. □

Using this lemma, we provide an expression for restricted monomials in terms of basic monomials.

Theorem 7.1.1. *Let $m \in \mathbb{N}$ and $n = md$. Let*

$$\mathcal{I} = (\mathcal{I}^{(1)}, \dots, \mathcal{I}^{(d)})$$
$$P = (\omega_1, \dots, \omega_m),$$

where $\mathcal{I}^{(k)} = (I_1^{(k)}, \dots, I_m^{(k)}) \subset 2^{[m]}$ such that $\emptyset \notin \mathcal{I}^{(k)}$ and $\omega_j \in \Omega^d(X)$. Then, for $\gamma : \square^d \to X$, we have

$$M_{\text{res}}^{\mathcal{I},P}(\gamma) = \sum_{\sigma=(\sigma_1,\dots,\sigma_d)\in\mathcal{M}(\mathcal{I})} M_{\text{bas}}^{\Pi(\mathcal{I},\sigma)^{-1}\circ\sigma,P}(\gamma), \tag{7.1.21}$$

where

$$\Pi(\mathcal{I},\sigma) = \big(\mu(\mathcal{I}^{(1)},\sigma_1), \dots, \mu(\mathcal{I}^{(d)},\sigma_d)\big). \tag{7.1.22}$$

Proof. From the previous lemma, the nontrivial domains of $M_{\text{res}}^{\mathcal{I},P}$ are $D_{m,d}(\sigma)$, where $\sigma = (\sigma_1, \dots, \sigma_d) \in \mathcal{M}(\mathcal{I})$. Let $\mu_k = \mu(\mathcal{I}^{(k)}, \sigma_k) \in \Sigma_m$. Then, restricted to $D_{m,d}(\sigma)$, the restricted monomial is

$$\int_{D_{m,d}(\sigma)} \bigwedge_{j=1}^{m} \gamma^* \omega_j(t_{\mu_1(j),1}, \dots, t_{\mu_d(j),d}).$$

By performing the change of variables

$$t_{\mu_k(j),k} \mapsto t_{j,k},$$

the integral becomes

$$\int_{D_{m,d}(\Pi(\mathcal{I},\sigma)^{-1}\circ\sigma)} \bigwedge_{j=1}^{m} \gamma^* \omega_j(t_{j,1}, \dots, t_{j,d}) = M_{\text{bas}}^{\Pi(\mathcal{I},\sigma)^{-1}\circ\sigma,P}(\gamma).$$

By summing over all $\sigma \in \mathcal{M}(\mathcal{I})$, we obtain the result. □

7.1.5 *The Mapping Space Signature*

In this section, we will define two notions of the mapping space signature. First, we define the mapping space signature in terms of all the basic monomials, motivated by the perspective that the basic monomials are the analogues of the simplicial path signature monomials described in the previous section. This first definition of the mapping space signature satisfies invariance properties, discussed in Section 7.2, which are analogous to those for the path signature. We show in Section 7.3 that this mapping space signature satisfies a shuffle product. Furthermore, we prove an injectivity theorem in Section 7.4 when the map $\gamma : \square^d \to X$ is lifted to include its parametrization $\tilde{\gamma} : \square^d \to \square^d \times X$, allowing us to show that the signature is a universal and characteristic feature map.

As opposed to the case of $d = 1$, maps in $\mathsf{Diff}(\square^d, X)$ have d distinct compositions, one for each dimension. In general, the relationship between the composition of maps and the algebraic structure of the mapping space signature is much more delicate in the higher dimensional case. A direct analogue of Chen's identity does not hold for the mapping space signature. This leads us to define a second notion of the signature, where the permutation index Π is fixed to the identity. In this case, we formulate a modified version of Chen's identity in Section 7.5.

We begin with the mapping space signature using all of the basic monomials. In the formulation of the signature for a manifold X, we must first choose a finite set of d-forms,

$$W = \{\omega_1, \dots, \omega_r\} \subset \Omega^d(X), \tag{7.1.23}$$

which contains the allowed forms to take in the forms index $\mathcal{P}$.

Definition 7.1.9. Let X be an N-dimensional manifold, and let $W = \{\omega_1, \dots, \omega_r\} \subset \Omega^d(X)$ be a fixed subset of d-forms. Let $H_{d,W}$ be normed vector space spanned by the orthonormal basis vectors $e_{\Pi,P}$, where $\Pi \in \Sigma^d_m$ and $P \in W^m$. Then, let $\mathcal{H}_{d,W}$ be the Hilbert space defined by the completion of $H_{d,W}$. The *mapping space signature for* $\mathsf{Diff}(\square^d, X)$ *with respect to* W is a map

$$M_{\mathrm{bas},W} : \mathsf{Diff}(\square^d, X) \to \mathcal{H}_{d,W}$$

defined by

$$M_{\mathrm{bas},W}(\gamma) := 1 + \sum_{(\Pi,P)} M^{\Pi,P}_{\mathrm{bas}}(\gamma)\, e_{\Pi,P}, \tag{7.1.24}$$

where the sum is taken over all allowed pairs (Π, P).

In the case of $X = \mathbb{R}^N$, our definitions of the interior monomials are already restricted to the standard d-forms.

Definition 7.1.10. Let $H_{d,N}$ be the normed vector space spanned by the orthonormal basis vectors $e_{\Pi,P}$, where $\Pi \in \Sigma_m^d$, and $\mathcal{P} = (P_1, \ldots, P_m)$, where each P_j is a multi-index in $[N]$. Then, let $\mathcal{H}_{d,N}$ be the Hilbert space defined as the completion of $H_{d,N}$. The *mapping space signature for* $\mathsf{Diff}(\square^d, \mathbb{R}^N)$ is the map

$$M_{\mathsf{bas}} : \mathsf{Diff}(\square^d, \mathbb{R}^N) \to \mathcal{H}_{d,N}$$

defined by

$$M_{\mathsf{bas}}(\gamma) := 1 + \sum_{(\Pi,\mathcal{P})} M_{\mathsf{bas}}^{\Pi,\mathcal{P}}(\gamma)\, e_{\Pi,\mathcal{P}}, \tag{7.1.25}$$

where the sum is taken over all allowed pairs $(\Pi, \mathcal{P})$.

We now introduce the second notion of the mapping space signature, where we restrict the permutation index $\Pi \in \Sigma_m^d$ to be the identity element $\Pi = \mathrm{id}$. In this case, the codomain of the signature can be described using the tensor algebra $T((\mathbb{R}^K))$ for some K.

Definition 7.1.11. Let X be an N-dimensional manifold, and let $W = \{\omega_1, \ldots, \omega_r\} \subset \Omega^d(X)$ be a fixed subset of d-forms. Let $T((\mathbb{R}^r))$ be the power series tensor algebra, with basis elements $e_P = e_{p_1} \otimes \ldots, e_{p_m}$ given by a multi-index $P = (p_1, \ldots, p_m)$ in $[r]$. The *identity mapping space signature for* $\mathsf{Diff}(\square^d, X)$ *with respect to* W is the map

$$M_{\mathsf{id}} : \mathsf{Diff}(\square^d, \mathbb{R}^N) \to T((\mathbb{R}^r)),$$

defined by

$$M_{\mathsf{id}}(\gamma) := 1 + \sum_{m=1}^{\infty} \sum_{|P|=m} M_{\mathsf{bas}}^{\mathrm{id},P}(\gamma)\, e_P. \tag{7.1.26}$$

Once again, in the case of $X = \mathbb{R}^N$, we simply choose W to be the collection of standard d-forms.

Definition 7.1.12. Let $T((\mathbb{R}^{\binom{N}{d}}))$ be the power series tensor algebra with basis elements $e_{\mathcal{P}} = e_{P_1} \otimes \ldots \otimes e_{P_m}$, given by $\mathcal{P} = (P_1, \ldots, P_m)$, where each P_j is a multi-index in $[N]$ of length d. The *identity mapping space signature for* $\mathsf{Diff}(\square^d, \mathbb{R}^N)$ is the map

$$M_{\mathsf{id}} : \mathsf{Diff}(\square^d, \mathbb{R}^N) \to T((\mathbb{R}^{\binom{N}{d}}))$$

defined by

$$M_{\mathsf{id}}(\gamma) := 1 + \sum_{m=1}^{\infty} \sum_{\mathcal{P}=(P_1,\ldots,P_m)} M_{\mathsf{bas}}^{\mathrm{id},\mathcal{P}}(\gamma)\, e_{\mathcal{P}}, \tag{7.1.27}$$

where the second sum is taken over all $\mathcal{P}$ consisting of m subsets $P_j \subset [N]$ of length d.

We have stated the definition of the various notions of monomials and signatures for both arbitrary manifolds and the special case of $X = \mathbb{R}^N$, but we will begin to exclusively state results for $X = \mathbb{R}^N$. While most of the results in the remainder of the chapter hold in both cases, we choose to work with $X = \mathbb{R}^N$ due to its direct analogy with the ordinary path signature, and to reduce repetition.

An important aspect of the definition of the path signature is the fact that it can be defined recursively. We show that a similar definition holds for the basic monomials indexed by the identity permutation, as well as the identity mapping space signature.

Proposition 7.1.1. *Let $\mathcal{P} = (P_1)$, where P_1 is a multi-index in $[N]$ of length d. For $\mathbf{a} = (a_1, \dots, a_d) \in \square^d$ and $\mathbf{b} = (b_1, \dots, b_d) \in \square^d$, define*

$$M_{\text{bas}}^{\text{id},\mathcal{P}}(\gamma)_{\mathbf{a},\mathbf{b}} \coloneqq \int_{\square^d(\mathbf{a},\mathbf{b})} J[\gamma_{P_1}](t_1, \dots, t_d) d\mathbf{t},$$

where

$$\square^d(\mathbf{a}, \mathbf{b}) \coloneqq \prod_{k=1}^{d} [a_k, b_k]. \tag{7.1.28}$$

Then, for $\mathcal{P} = (P_1, \dots, P_m)$, where each $P_j \subset [N]$ of length d, let $\overline{\mathcal{P}} = (P_1, \dots, P_{m-1})$, and we recursively define

$$M_{\text{bas}}^{\text{id},\mathcal{P}}(\gamma)_{\mathbf{a},\mathbf{b}} \coloneqq \int_{\square^d(\mathbf{a},\mathbf{b})} M_{\text{bas}}^{\text{id},\overline{\mathcal{P}}}(\gamma)_{\mathbf{a},\mathbf{t}} J[\gamma_{P_m}](t_1, \dots, t_d) d\mathbf{t}. \tag{7.1.29}$$

Then,

$$M_{\text{bas}}^{\text{id},\mathcal{P}}(\gamma)_{\mathbf{0},\mathbf{1}} = M_{\text{bas}}^{\text{id},\mathcal{P}}(\gamma). \tag{7.1.30}$$

Proof. We obtain the proof by recursively expanding the definition of $M_{\text{bas}}^{\text{id},\mathcal{P}}(\gamma)_{\mathbf{0},\mathbf{1}}$ given in Equation (7.1.29). Let $\mathbf{t}_j = (t_{j,1}, \dots, t_{j,d})$. Then

$$\begin{aligned}
M_{\text{bas}}^{\text{id},\mathcal{P}}(\gamma)_{\mathbf{0},\mathbf{1}} &= \int_{\mathbf{t}_m \in \square^d} M_{\text{bas}}^{\text{id},(P_1,\dots,P_{m-1})}(\gamma)_{\mathbf{0},\mathbf{t}_m} J[\gamma_{P_m}](\mathbf{t}_m) d\mathbf{t}_m \\
&= \int_{\mathbf{t}_m \in \square^d} \int_{\mathbf{t}_{m-1} \in \square^d(\mathbf{0},\mathbf{t}_m)} M_{\text{bas}}^{\text{id},(P_1,\dots,P_{m-2})}(\gamma)_{\mathbf{0},\mathbf{t}_{m-1}} J[\gamma_{P_{m-1}}](\mathbf{t}_{m-1}) d\mathbf{t}_{m-1} d\mathbf{t}_m \\
&= \int_{\mathbf{t}_m \in \square^d} \int_{\mathbf{t}_{m-1} \in \square^d(\mathbf{0},\mathbf{t}_m)} \dots \int_{\mathbf{t}_1 \in \square^d(\mathbf{0},\mathbf{t}_2)} \prod_{j=1}^{m} J[\gamma_{P_j}](\mathbf{t}_j) d\mathbf{t}_1 \dots d\mathbf{t}_m.
\end{aligned}$$

Then, since the domain of integration can be rewritten as

$$\square^d \times \square^d(0, \mathbf{t}_m) \times \ldots \square^d(0, \mathbf{t}_2) = \Delta^m \times \ldots \times \Delta^m = D_{d,m}(\mathrm{id}),$$

we have $M_{\mathrm{bas}}^{\mathrm{id},\mathcal{P}}(\gamma)_{\mathbf{0},\mathbf{1}} = M_{\mathrm{bas}}^{\mathrm{id},\mathcal{P}}(\gamma)$.

□

Corollary 7.1.1. *For* $\mathbf{a}, \mathbf{b} \in \square^d$ *and* $\gamma \in \mathsf{Diff}(\square^d, \mathbb{R}^N)$*, let*

$$M_{\mathrm{id}}(\gamma)_{\mathbf{a},\mathbf{b}} \coloneqq 1 + \sum_{m=1}^{\infty} \sum_{\mathcal{P}=(P_1,\ldots,P_m)} M_{\mathrm{bas}}^{\mathrm{id},\mathcal{P}}(\gamma)_{\mathbf{a},\mathbf{b}}\, e_{\mathcal{P}}, \tag{7.1.31}$$

and let

$$\begin{aligned} J[\gamma] : \square^d &\to \mathbb{R}^{\binom{N}{d}} \\ \mathbf{t} &\mapsto \left(J[\gamma_P](\mathbf{t})\right)_{P\subset[N],|P|=d}. \end{aligned} \tag{7.1.32}$$

Then,

$$M_{\mathrm{id}}(\gamma) = \int_{\square^d} M_{\mathrm{id}}(\gamma)_{\mathbf{0},\mathbf{t}} \otimes J[\gamma](\mathbf{t}) d\mathbf{t}. \tag{7.1.33}$$

Proof. This is immediate from the previous proposition. □

7.2 INVARIANCE PROPERTIES

In this section, we discuss invariance properties of the mapping space signature. We begin with Jacobian invariance, which reduces to translation invariance for the path signature. Next, we discuss permutation invariance with respect to the indexing subsets, and reparametrization invariance in which each coordinate of $\square^d$ is reparametrized independently. Finally, we discuss a special case of maps which have trivial mapping space signature.

7.2.1 *Jacobian Invariance*

Because the mapping space signature is defined in terms of Jacobian determinants,

Definition 7.2.1. Given two maps $\alpha, \beta \in \mathsf{Diff}(\square^d, \mathbb{R}^N)$, we say that α and β are *Jacobian equivalent*, denoted $\alpha \sim_J \beta$ if

$$J[\alpha_P](\mathbf{t}) = J[\beta_P](\mathbf{t}) \tag{7.2.1}$$

for all $P \subset [N]$ such that $|P| = d$ and all $\mathbf{t} \in \square^d$. Furthermore, we will denote the Jacobian equivalence classes of maps by

$$\widetilde{\mathsf{Diff}}(\square^d, \mathbb{R}^N) := \mathsf{Diff}(\square^d, \mathbb{R}^N) / \sim_J . \tag{7.2.2}$$

Example 7.2.1. We provide a simple 1-parameter family of maps which are all Jacobian equivalent. Let $a \in \mathbb{R}$ such that $a \neq 0$. Let $\gamma^a : \square^2 \to \mathbb{R}^3$ be given by

$$\gamma^a(t_1, t_2) = \big(at_1, \frac{1}{a}t_2, -at_1 + \frac{1}{a}t_2\big).$$

The Jacobian of this map is

$$d\gamma^a(t_1, t_2) = \begin{pmatrix} a & 0 & -a \\ 0 & \frac{1}{a} & \frac{1}{a} \end{pmatrix}.$$

Then, the determinants of the three Jacobian minors are

$$J[\gamma^a_{1,2}](t_1, t_2) = J[\gamma^a_{2,3}](t_1, t_2) = J[\gamma^a_{1,3}](t_1, t_2) = 1$$

for all $(t_1, t_2) \in \square^2$. Thus, $\gamma^a \sim_J \gamma^b$ for any $a, b \in \mathbb{R} - \{0\}$.

The following proposition is clear from the definition of the basic monomials.

Proposition 7.2.1. *Suppose $\alpha, \beta \in \mathsf{Diff}(\square^d, \mathbb{R}^N)$ such that $\alpha \sim_J \beta$. Then*

$$M_{\text{bas}}(\alpha) = M_{\text{bas}}(\beta).$$

Remark 7.2.1. In the case of $d = 1$, if $\alpha, \beta \in [0,1] \to \mathbb{R}^N$, then $\alpha \sim_J \beta$ if and only if $\alpha_i'(t) = \beta_i'(t)$ for all $i \in [N]$ and $t \in [0,1]$. In other words, $\alpha \sim_J \beta$ if and only if $\alpha = \beta + c$ for some constant $c \in \mathbb{R}^N$. Thus, Jacobian invariance of the mapping space signature reduces to translation invariance of the path signature.

7.2.2 *Permutation Invariance*

Next, we consider invariance of the signature with respect to permutations of the indices in the basic monomials.

Proposition 7.2.2. *Let $\Pi = (\pi_1, \ldots, \pi_d) \in \Sigma_m^d$ and $\mathcal{P} = (P_1, \ldots, P_m)$ where $P_j \subset [N]$ such that $|P_j| = d$. Let $\sigma \in \Sigma_m$ and define the permutation actions*

$$\sigma\Pi = (\sigma\pi_1, \ldots, \sigma\pi_d)$$
$$\sigma\mathcal{P} = (P_{\sigma(1)}, \ldots, P_{\sigma(m)}).$$

Then for any $\gamma \in \mathsf{Diff}(\square^d, \mathbb{R}^N)$,

$$M_{\text{bas}}^{\Pi,\sigma\mathcal{P}}(\gamma) = M_{\text{bas}}^{\sigma\Pi,\mathcal{P}}(\gamma). \tag{7.2.3}$$

Proof. This will be proved by expanding the definition and performing a change of variables. Indeed

$$M_{\text{bas}}^{\Pi,\sigma\mathcal{P}}(\gamma) = \int_{D_{d,m}(\Pi)} \prod_{j=1}^{m} J[\gamma_{P_{\sigma(j)}}](t_{j,1}, \ldots, t_{j,d}) d\mathbf{t}.$$

We perform the change of variables

$$t_{j,k} \mapsto t_{\sigma(j),k}.$$

Under this transformation, the domain of integration becomes $D_{d,m}(\sigma\Pi)$, and we obtain the desired result. □

This result implies that in practice, we do not need to compute all of the basic monomials in order to obtain the mapping space signature. Indeed, we can restrict ourselves to the basic monomials where the permutation index $\Pi = (\pi_1, \ldots, \pi_d) \in \Sigma_m^d$ has one component fixed to the identity; for example $\pi_1 = \text{id}$. Indeed, we can compute the basic monomial for an arbitrary Π as

$$M_{\text{bas}}^{\Pi,\mathcal{P}}(\gamma) = M_{\text{bas}}^{\pi_1^{-1}\Pi,\pi_1\mathcal{P}}(\gamma).$$

Remark 7.2.2. In the case of $d = 1$, this implies that the permutation index can always be set to the identity, and thus all of the basic monomials can be written as integrals over the standard n simplex Δ^n. This explains the absence of a permutation index for the path signature monomials.

7.2.3 *Reparametrization Invariance*

Next we will consider reparametrization invariance, one of the fundamental properties of the path signature. We find that the mapping space signature is invariant under reparametrizations in which each coordinate of $\square^d$ is reparametrized independently.

Proposition 7.2.3. *Let* $\phi_k : [0,1] \to [0,1]$ *be smooth, monotone increasing bijections for* $k = 1, \ldots, d$ *and let* $\phi = (\phi_1, \ldots, \phi_d) : \square^d \to \square^d$. *For any* $\Pi \in \Sigma_m^d$, $\mathcal{P} = (P_1, \ldots, P_m)$ *such that* $P_j \subset [N]$ *and* $|P_j| = d$, *and* $\gamma \in \mathsf{Diff}(\square^d, \mathbb{R}^N)$,

$$M_{\text{bas}}^{\Pi,\mathcal{P}}(\gamma \circ \phi) = M_{\text{bas}}^{\Pi,\mathcal{P}}(\gamma). \tag{7.2.4}$$

Proof. First, we note that for any $P \subset [N]$ such that $|P| = d$, the chain rule gives us

$$\begin{aligned} J[(\gamma \circ \phi)_P](t_1, \ldots, t_d) &= J[\gamma_P](\phi_1(t_1), \ldots, \phi_d(t_d)) J[\phi](t_1, \ldots, t_d) \\ &= J[\gamma_P](\phi_1(t_1), \ldots, \phi_d(t_d)) \phi_1'(t_1) \ldots \phi_d'(t_d). \end{aligned}$$

Then, applying this to $M_{\text{bas}}^{\Pi,\mathcal{P}}(\gamma \circ \phi)$, we get

$$\begin{aligned} M_{\text{bas}}^{\Pi,\mathcal{P}}(\gamma \circ \phi) &= \int_{D_{d,m}(\Pi)} \prod_{j=1}^{m} J[(\gamma \circ \phi)_{P_j}](t_{j,1}, \ldots, t_{j,d}) d\mathbf{t} \\ &= \int_{D_{d,m}(\Pi)} \prod_{j=1}^{m} J[\gamma_{P_j}](\phi_1(t_{j,1}), \ldots, \phi_d(t_{j,d})) \phi_1'(t_{j,1}) \ldots \phi_d'(t_{j,d}) d\mathbf{t}. \end{aligned}$$

Next, we make the change of variables

$$\begin{aligned} \phi_k(t_{j,k}) &\mapsto t_{j,k} \\ \phi_k'(t_{j,k}) dt &\mapsto dt_{j,k}. \end{aligned}$$

Because each ϕ_k is a monotone increasing bijection, each simplex $\Delta_{\pi_k}^m$ is preserved under this transformation:

$$t_{\pi_k(1),k} < \ldots < t_{\pi_k(m),k} \quad \Longrightarrow \quad \phi_k^{-1}(t_{\pi_k(1),k}) < \ldots < \phi_k^{-1}(t_{\pi_k(d),k}).$$

Then, under this transformation, the above integral becomes

$$\begin{aligned} M_{\text{bas}}^{\Pi,\mathcal{P}}(\gamma \circ \phi) &= \int_{D_{d,m}(\Pi)} \prod_{j=1}^{m} J[\gamma_{P_j}](t_{j,1}, \ldots, t_{j,d}) d\mathbf{t} \\ &= M_{\text{bas}}^{\Pi,\mathcal{P}}(\gamma). \end{aligned}$$

□

7.2.4 *Maps with Trivial Signature*

In this section, we consider maps $\gamma \in \mathsf{Diff}(\square^d, \mathbb{R}^N)$ defined as a sum of d paths. For such maps, we show that the basic monomials can be expressed as a linear combination of the simplicial path signature monomials, which leads to a class of maps which have trivial signature.

Let $g_1, \dots, g_d : [0,1] \to \mathbb{R}^N$ be a collection of paths on $\mathbb{R}^N$, where each path is written component-wise as $g_k = (g_{1,k}, \dots, g_{N,k})$. Note that k is fixed in the second index. We consider the map

$$\begin{aligned} \gamma : \square^d &\to \mathbb{R}^N \\ (t_1, \dots, t_d) &\mapsto \sum_{k=1}^{d} g_k(t_k). \end{aligned}$$

This can be written component wise as $\gamma = (\gamma_1, \dots, \gamma_N)$, where

$$\begin{aligned} \gamma_j : \square^d &\to \mathbb{R} \\ (t_1, \dots, t_d) &\mapsto \sum_{k=1}^{d} g_{j,k}(t_k). \end{aligned}$$

Given a subset $P \subset [N]$ of length d, which we treat as a function $P : [d] \to [N]$, the Jacobian determinant is

$$J[\gamma_P](t_1, \dots, t_d) = \sum_{\sigma \in \Sigma_d} (-1)^{\operatorname{sgn}(\sigma)} \prod_{i=1}^{d} g'_{P(i),\sigma(i)}(t_{\sigma(i)}). \tag{7.2.5}$$

Now, we can compute the basic monomials for such maps as follows. Suppose $\Pi = (\pi_1, \dots, \pi_d) \in \Sigma_m^d$ and $\mathcal{P} = (P_1, \dots, P_m)$, where each $P_j \subset [N]$ is of length d. Then,

$$\begin{aligned} M_{\text{bas}}^{\Pi,\mathcal{P}}(\gamma) &= \int_{D_{d,m}(\Pi)} \prod_{j=1}^{m} J[\gamma_{P_j}](t_{j,1}, \dots, t_{j,d}) d\mathbf{t} \\ &= \int_{D_{d,m}(\Pi)} \prod_{j=1}^{m} \left(\sum_{\sigma_j \in \Sigma_d} (-1)^{\operatorname{sgn}(\sigma_j)} \prod_{i=1}^{d} g'_{P_j(i),\sigma_j(i)}(t_{j,\sigma(i)}) \right) d\mathbf{t} \\ &= \int_{D_{d,m}(\Pi)} \left(\sum_{\sigma=(\sigma_1,\dots,\sigma_m) \in \Sigma_d^m} (-1)^{\operatorname{sgn}(\sigma)} \prod_{j=1}^{m} \prod_{i=1}^{d} g'_{P_j(i),\sigma_j(i)}(t_{j,\sigma_j(i)}) \right) d\mathbf{t} \\ &= \sum_{\sigma=(\sigma_1,\dots,\sigma_n) \in \Sigma_d^n} (-1)^{\operatorname{sgn}(\sigma)} \int_{D_{d,m}(\Pi)} \prod_{i=1}^{d} \prod_{j=1}^{m} g'_{P_j(i),\sigma_j(i)}(t_{j,\sigma_j(i)}) d\mathbf{t} \end{aligned}$$

At this point, we break the inner integral up into a product of d integrals. This is done by using the structure of the domain of integration

$$D_{d,n}(\Pi) = \Delta^n_{\pi_1} \times \ldots \times \Delta^n_{\pi_d}$$

where $\Delta^n_{\pi_j}$ is parametrized by $\mathbf{t}_k = (t_{1,k}, \dots, t_{m,k})$ (note that k is the second index). Therefore, we write the double product in the integrand as a product of functions of $t_{j,k}$ for fixed k. Given $\sigma = (\sigma_1, \dots, \sigma_m) \in \Sigma_d^m$ and some $k \in [d]$, for every j, there exists a unique i such that $\sigma_j(i) = k$. Thus, we re-index the double product in terms of k rather than i:

$$\prod_{i=1}^{d}\prod_{j=1}^{m} g'_{P_j(i),\sigma_j(i)}(t_{j,\sigma_j(i)}) = \prod_{k=1}^{d}\prod_{j=1}^{m} g'_{P_j(\sigma_j^{-1}(k)),k}(t_{j,k}).$$

Continuing along with the above manipulations, we get

$$\begin{aligned} M^{\Pi,\mathcal{P}}_{\text{bas}}(\gamma) &= \sum_{\sigma=(\sigma_1,\dots,\sigma_m)\in\Sigma_d^m} (-1)^{\text{sgn}(\sigma)} \int_{\mathbf{t}\in D_{d,n}(\Pi)} \prod_{k=1}^{d}\prod_{j=1}^{m} g'_{P_j(\sigma_j^{-1}(k)),k}(t_{j,k}) d\mathbf{t} \\ &= \sum_{\sigma=(\sigma_1,\dots,\sigma_m)\in\Sigma_d^m} (-1)^{\text{sgn}(\sigma)} \prod_{k=1}^{d} \int_{\mathbf{t}_k\in\Delta^m_{\pi_k}} \prod_{j=1}^{m} g'_{P_j(\sigma_j^{-1}(k)),k}(t_{j,k}) d\mathbf{t}_k \\ &= \sum_{\sigma=(\sigma_1,\dots,\sigma_m)\in\Sigma_d^m} (-1)^{\text{sgn}(\sigma)} \prod_{k=1}^{d} S^{Q(\Pi,\mathcal{P},\sigma,k)}(g_k), \end{aligned}$$

Here, $Q(\Pi, \mathcal{P}, \sigma, k)$ is a multi-index of length m valued in $[N]$, defined as follows. Suppose $\Pi = (\pi_1, \dots, \pi_d) \in \Sigma_m^d$, $\sigma = (\sigma_1, \dots, \sigma_m) \in \Sigma_d^m$, $\mathcal{P} = (P_1, \dots, P_m)$, where each $P_j \subset [N]$ is of length d, and $k \in [d]$. Then,

$$Q(\Pi, \mathcal{P}, \sigma, k) = \pi_k^{-1} \cdot \left(P_1(\sigma_1^{-1}(k)), \dots, P_m(\sigma_m^{-1}(k))\right), \tag{7.2.6}$$

where $\left(P_1(\sigma_1^{-1}(k)), \dots, P_m(\sigma_m^{-1}(k))\right)$ is a multi-index of length n valued in $[N]$, and $\pi_k^{-1} \in \Sigma_m$ acts by permutation of elements. Thus, we have proved the following.

Proposition 7.2.4. *Let $g_1, \dots, g_d : [0,1] \to \mathbb{R}^N$ be a collection of paths on $\mathbb{R}^N$. Let $\gamma : \square^d \to \mathbb{R}^N$ be defined by*

$$\gamma(t_1, \dots, t_d) := \sum_{k=1}^{d} g_k(t_k). \tag{7.2.7}$$

Then,

$$M^{\Pi,\mathcal{P}}_{\text{bas}}(\gamma) = \sum_{\sigma\in\Sigma_d^m} (-1)^{\text{sgn}(\sigma)} \prod_{k=1}^{d} S_\Delta^{Q(\Pi,\mathcal{P},\sigma,k)}(g_k), \tag{7.2.8}$$

where $Q(\Pi, \mathcal{P}, \sigma, k)$ is given in Equation (7.2.6).

An immediate corollary is the following.

Corollary 7.2.1. *Let $g_1, \dots, g_d : [0,1] \to \mathbb{R}^N$ be a collection of paths on $\mathbb{R}^N$. Let $\gamma : \square^d \to \mathbb{R}^N$ be defined by Equation (7.2.7). Furthermore, suppose at least one of the paths g_k is tree-like. Then,*

$$M_{\text{bas}}^{\Pi,\mathcal{P}}(\gamma) = 0$$

for all Π and $\mathcal{P}$.

Proof. Without loss of generality, suppose g_1 is tree-like. Because g_1 is tree-like, its path signature is trivial, so for any multi-index P valued in $[N]$, $S_\Delta^P(g_1) = 0$. Then, by Equation (7.2.8), $M_{\text{bas}}^{\Pi,\mathcal{P}}(\gamma) = 0$ for any Π and $\mathcal{P}$. □

7.3 SHUFFLE PRODUCT

In this section, we consider the product structure of the restricted and basic monomials. The product structure for restricted monomials is immediate from Theorem 3.3.3.

Theorem 7.3.1. *Let $m_1, m_2 \in \mathbb{N}$. Let*

$$\mathcal{I} = (\mathcal{I}^{(1)}, \dots, \mathcal{I}^{(d)}), \quad \mathcal{J} = (\mathcal{J}^{(1)}, \dots, \mathcal{J}^{(d)})$$
$$\mathcal{P} = (P_1, \dots, P_{m_1}), \quad \mathcal{Q} = (Q_1, \dots, Q_{m_2}),$$

where $\mathcal{I}^{(k)} = (I_1^{(k)}, \dots, I_{m_1}^{(k)}) \subset 2^{[m_1]}$ and $\mathcal{J}^{(k)} = (J_1^{(k)}, \dots, J_{m_2}^{(k)}) \subset 2^{[m_2]}$such that $\emptyset \notin \mathcal{I}^{(k)}, \mathcal{J}^{(k)}$ for any k and $P_j, Q_j \subset [N]$ are subsets of length d. Let

$$\mathcal{I} * \mathcal{J} = \left(\mathcal{I}^{(1)} * \mathcal{J}^{(1)}, \dots, \mathcal{I}^{(d)} * \mathcal{J}^{(d)}\right)$$
$$\mathcal{P} * \mathcal{Q} = (P_1, \dots, P_{m_1}, Q_1, \dots, Q_{m_2}),$$

where

$$\mathcal{I}^{(k)} * \mathcal{J}^{(k)} = \left(I_1^{(k)}, \dots, I_{m_1}^{(k)}, J_1^{(k)} + m_1, \dots, J_{m_2}^{(k)} + m_1\right).$$

Then, for any $\gamma \in \mathsf{Diff}(\square^d, \mathbb{R}^N)$,

$$M_{\text{res}}^{\mathcal{I},\mathcal{P}}(\gamma) M_{\text{res}}^{\mathcal{J},\mathcal{Q}}(\gamma) = M_{\text{res}}^{\mathcal{I}*\mathcal{J},\mathcal{P}*\mathcal{Q}}(\gamma). \tag{7.3.1}$$

The product structure of the basic monomials provide the more familiar shuffle product.

Theorem 7.3.2. *Let $m_1, m_2 \in \mathbb{N}$. Let*

$$\begin{aligned} \Pi_1 &= (\pi_{1,1}, \dots, \pi_{1,d}) \in \Sigma_{m_1}^d, & \Pi_2 &= (\pi_{2,1}, \dots, \pi_{2,d}) \in \Sigma_{m_2}^d \\ \mathcal{P}_1 &= (P_1, \dots, P_{m_1}), & \mathcal{P}_2 &= (P_{m_1+1}, \dots, P_{m_1+m_2}), \end{aligned}$$

where $P_j \subset [N]$ are subsets of length d for $j = 1, \dots, m_1 + m_2$. Let

$$\begin{aligned} \Pi &= (\pi_1, \dots, \pi_d) \in \Sigma_{m_1+m_2}^d \\ \mathcal{P} &= (P_1, \dots, P_{m_1+m_2}), \end{aligned}$$

where $\pi_k = \pi_{1,k} * (\pi_{2,k} + m_1) \in \Sigma_{m_1+m_2}$. Then, for any $\gamma \in \mathsf{Diff}(\square^d, \mathbb{R}^N)$,

$$M_{\mathrm{bas}}^{\Pi_1,\mathcal{P}_1}(\gamma) M_{\mathrm{bas}}^{\Pi_2,\mathcal{P}_2}(\gamma) = \sum_{\sigma=(\sigma_1,\dots,\sigma_d) \in Sh(m_1,m_2)^d} M_{\mathrm{bas}}^{\sigma\Pi,\mathcal{P}}(\gamma), \tag{7.3.2}$$

where $Sh(m_1, m_2) \subset \Sigma_{m_1+m_2}$ is the set of (m_1, m_2)-shuffles.

Proof. Writing out the product of the two basic monomials, we have

$$M_{\mathrm{bas}}^{\Pi_1,\mathcal{P}_1}(\gamma) M_{\mathrm{bas}}^{\Pi_2,\mathcal{P}_2}(\gamma) = \int_{D_{d,m_1}(\Pi_1) \times D_{d,m_2}(\Pi_2)} \prod_{j=1}^{m_1+m_2} J[\gamma_{P_j}](t_{j,1}, \dots, t_{j,d}) d\mathbf{t}, \tag{7.3.3}$$

where $(t_{j,k})_{\substack{j=1,\dots,m_1 \\ k=1,\dots,d}} \in D_{d,m_1}(\Pi_1)$ and $(t_{j,k})_{\substack{j=m_1+1,\dots,m_1+m_2 \\ k=1,\dots,d}} \in D_{d,m_2}(\Pi_2)$. The domain of integration is

$$\begin{aligned} D_{d,m_1}(\Pi_1) \times D_{d,m_2}(\Pi_2) &= \left(\prod_{k=1}^d \Delta_{\pi_{1,k}}^{m_1} \right) \times \left(\prod_{k=1}^d \Delta_{\pi_{2,k}}^{m_2} \right) \\ &= \prod_{k=1}^d \left(\Delta_{\pi_{1,k}}^{m_1} \times \Delta_{\pi_{2,k}}^{m_2} \right). \end{aligned}$$

Then, using the shuffle decomposition of $\Delta_{\pi_{1,k}}^{m_1} \times \Delta_{\pi_{2,k}}^{m_2}$ into the disjoint union

$$\Delta_{\pi_{1,k}}^{m_1} \times \Delta_{\pi_{2,k}}^{m_2} = \coprod_{\sigma_k \in Sh(m_1,m_2)} \Delta_{\sigma_k \pi_k}^{m_1+m_2},$$

we get

$$\begin{aligned} D_{d,m_1}(\Pi_1) \times D_{d,m_2}(\Pi_2) &= \coprod_{\sigma=(\sigma_1,\dots,\sigma_d)\in Sh(m_1,m_2)^d} \prod_{k=1}^{d} \Delta^{m_1+m_2}_{\sigma_k\pi_k} \\ &= \coprod_{\sigma=(\sigma_1,\dots,\sigma_d)\in Sh(m_1,m_2)^d} D_{d,m_1+m_2}(\sigma\Pi). \end{aligned}$$

This decomposition of the domain of integration allows us to rewrite the integral in Equation (7.3.3) as a sum of integrals over the domain $D_{d,m_1+m_2}(\sigma\Pi)$,

$$\begin{aligned} M^{\Pi_1,\mathcal{P}_1}_{\text{bas}}(\gamma) M^{\Pi_2,\mathcal{P}_2}_{\text{bas}}(\gamma) &= \sum_{\sigma\in Sh(m_1,m_2)^d} \int_{D_{d,m_1+m_2}(\sigma\Pi)} \prod_{j=1}^{m_1+m_2} J[\gamma_{P_j}](t_{j,1},\dots,t_{j,d}) d\mathbf{t} \\ &= \sum_{\sigma\in Sh(m_1,m_2)^d} M^{\sigma\Pi,\mathcal{P}}_{\text{bas}}(\gamma). \end{aligned}$$

□

This provides the basic monomials with a shuffle product structure, analogous to that of the simplicial signature monomials. It is also important to note that the shuffle product is not closed for the basic monomials restricted to $\Pi = \text{id}$. Indeed, the shuffle product for basic monomials gives us

$$M^{\text{id},\mathcal{P}_1}_{\text{bas}}(\gamma) M^{\text{id},\mathcal{P}_2}_{\text{bas}}(\gamma) = \sum_{\sigma\in Sh(m_1,m_2)^d} M^{\sigma,\mathcal{P}}_{\text{bas}}(\gamma),$$

where $\sigma \in Sh(m_1,m_2)^d \subset \Sigma^d_{m_1+m_2}$ is not restricted to the identity. In the case of $d = 1$, permutation invariance from Proposition 7.2.2 allows us to move the σ to the forms index $\mathcal{P}$. However, in the case of $d > 1$, this approach is not possible since $\sigma = (\sigma_1,\dots,\sigma_d) \in \Sigma^d_m$ can consist of d independent permutations.

7.4 PARAMETRIZED INJECTIVITY THEOREM

In this section, we consider two injectivity theorems for the mapping space signature. Both of these theorems require that we lift maps $\gamma \in \mathsf{Diff}(\square^d, \mathbb{R}^N)$ to a map $\overline{\gamma} \in \mathsf{Diff}(\square^d, \mathbb{R}^{d+N})$ by appending the parametrization, and using the mapping space signature for $\mathsf{Diff}(\square^d, \mathbb{R}^{d+N})$. The difference between the two theorems comes from the choice of forms we allow from the additional d coordinates. The first will yield an injective map on Jacobian equivalence classes of maps $\widetilde{\mathsf{Diff}}(\square^d, \mathbb{R}^N)$, while the second will yield an injective map for all smooth maps $\mathsf{Diff}(\square^d, \mathbb{R}^N)$.

7.4.1 *Jacobian Equivalence Classes*

We begin with the inclusion

$$\begin{aligned} \iota : \mathsf{Diff}(\square^d, \mathbb{R}^N) &\to \mathsf{Diff}(\square^d, \mathbb{R}^{d+N}) \\ \gamma(\mathbf{t}) &\mapsto (\mathbf{t}, \gamma(\mathbf{t})). \end{aligned} \tag{7.4.1}$$

We will denote the coordinates of $\mathbb{R}^{d+N}$ by $(u_1, \ldots, u_d, x_1, \ldots, x_N)$. Recall that in the definition of the basic monomials, we denote the standard d form $dx_{p_1} \wedge \ldots \wedge dx_{p_d} \in \Omega^d(\mathbb{R}^N)$ by the subset $P = (p_1, \ldots, p_d) \subset [N]$. We will continue to use the same notation for the standard d-forms on the x coordinates of $\mathbb{R}^N$. In addition, we will also allow the d-form $du_1 \wedge \ldots \wedge du_d$ in our signature, and we denote this form by V. Let $\widetilde{\mathcal{H}}_{d,N}$ be the Hilbert space given by the completion of the normed vector space with orthonormal basis vectors $e_{\Pi,\mathcal{P}}$, where $\Pi \in \Sigma_m^d$ and $\mathcal{P} = (P_1, \ldots, P_m)$, where P_j is either a subset of $[N]$ of length d, or $P_j = V$. This is equivalent to the Hilbert space defined in Definition 7.1.9 given our allowed choice of forms.

Then, let $m \in \mathbb{N}$, $\Pi \in \Sigma_m^d$, and $\mathcal{P} = (P_1, \ldots, P_m)$, where P_j is either a subset of $[N]$ of length d, or $P_j = V$. We define the parametrized mapping space signature signature $\widetilde{M}_{\text{bas}} : \mathsf{Diff}(\square^d, \mathbb{R}^N) \to \widetilde{\mathcal{H}}_{d,N}$ by

$$\widetilde{M}_{\text{bas}}(\gamma) := 1 + \sum_{(\Pi,\mathcal{P})} M_{\text{bas}}^{\Pi,\mathcal{P}}(\iota(\gamma))\, e_{\Pi,\mathcal{P}}, \tag{7.4.2}$$

for all allowed pairs $(\Pi, \mathcal{P})$. The map $\iota(\gamma) : \square^d \to \mathbb{R}^{d+N}$ is simply the identity map on the u-variables in $\mathbb{R}^{d+N}$. Thus, $J[\iota(\gamma)_V](\mathbf{t}) = 1$ for all $\mathbf{t} \in \square^d$. As an example of this parametrized signature, consider $\Pi \subset \Sigma_m^d$, and $\mathcal{P} = (V, \ldots, V, P)$, where $P \subset [N]$ of length d. Then, the parametrized mapping space signature of $\gamma \in \mathsf{Diff}(\square^d, \mathbb{R}^N)$ is

$$\widetilde{M}_{\text{bas}}^{\Pi,\mathcal{P}}(\gamma) = \int_{D_{d,m}(\Pi)} J[\gamma_P](t_{m,1}, \ldots, t_{m,d}) d\mathbf{t}. \tag{7.4.3}$$

It is clear that the parametrized mapping space signature is invariant under Jacobian equivalence, and descends to a map $\widetilde{M}_{\text{bas}} : \widetilde{\mathsf{Diff}}(\square^d, \mathbb{R}^N) \to \widetilde{\mathcal{H}}_{d,N}$. We show that $\widetilde{M}_{\text{bas}}$ is injective on $\widetilde{\mathsf{Diff}}(\square^d, \mathbb{R}^N)$. The main idea behind the proof is to show that for any exponent vector $\mathbf{c} = (c_1, \ldots, c_d) \in \mathbb{N}^d$ and any $P \subset [N]$ of length d, there exists a linear combination of the parametrized basic monomials $\widetilde{M}_{\text{bas}}^{\Pi,\mathcal{P}}$ which is equal to

$$\int_{\square^d} \mathbf{t}^{\mathbf{c}} J[\gamma_P](\mathbf{t}) d\mathbf{t}$$

for all $\gamma \in \mathsf{Diff}(\square^d, \mathbb{R}^N)$, where $\mathbf{t}^{\mathbf{c}} := t_1^{c_1} \ldots t_d^{c_d}$. In other words, the following lemma holds.

Lemma 7.4.1. *For any $\mathbf{c} \in \mathbb{N}^d$, and any $P \subset [N]$ of length d, there exists an element $\ell_{\mathbf{c},P} \in \widetilde{\mathcal{H}}^*_{d,N}$ such that*

$$\langle \ell_{\mathbf{c},P}, \widetilde{M}_{\text{bas}}(\gamma) \rangle = \int_{\square^d} \mathbf{t}^{\mathbf{c}} J[\gamma_P](\mathbf{t}) d\mathbf{t}. \tag{7.4.4}$$

Before we prove this lemma, we will show that is used to prove the injectivity theorem. This will require a second lemma.

Lemma 7.4.2. *Suppose $g : \square^d \to \mathbb{R}$ is a smooth function. Then, if*

$$\int_{\square^d} \mathbf{t}^{\mathbf{c}} g(\mathbf{t}) d\mathbf{t} = 0 \tag{7.4.5}$$

for all $\mathbf{c} \in \mathbb{N}^d$, then $g = 0$.

Proof. Let $\mathbb{R}[t_1, \ldots, t_d]$ denote the space of polynomials in d variables viewed as functions $Q : \square^d \to \mathbb{R}$. By the Stone-Weierstrass theorem, $\mathbb{R}[t_1, \ldots, t_d]$ is dense in $C(\square^d, \mathbb{R})$, the space of continuous functions under the uniform topology. Next, we treat $\mu_g = g(\mathbf{t})d\mathbf{t}$ as a finite Borel measure on $\square^d$, where

$$\mu_g(\mathbf{t}^{\mathbf{c}}) = \int_{\square^d} \mathbf{t}^{\mathbf{c}} g(\mathbf{t}) d\mathbf{t}.$$

Then, since $\mathbb{R}[t_1, \ldots, t_d]$ is dense in $C(\square^d, \mathbb{R})$, and $\mu_g(\mathbf{t}^{\mathbf{c}}) = 0$ for all $\mathbf{c} \in \mathbb{N}^d$, we have $\mu_g(f) = 0$ for all $f \in C(\square^d, \mathbb{R})$. The space of finite Borel measures on $\square^d$ is the dual space of $C(\square^d, \mathbb{R})$, and the fact that $\mu_g(f) = 0$ for all $f \in C(\square^d, \mathbb{R})$ implies that $\mu_g = 0$. Thus, $g = 0$. □

Using these two lemmas, we can prove the parametrized injectivity theorem.

Theorem 7.4.1. *The parametrized mapping space signature*

$$\widetilde{M}_{\text{bas}} : \widetilde{\mathsf{Diff}}(\square^d, \mathbb{R}^N) \to \widetilde{\mathcal{H}}_{d,N}$$

is injective.

Proof. Suppose $\alpha, \beta \in \mathsf{Diff}(\square^d, \mathbb{R}^N)$ such that $\widetilde{M}_{\text{bas}}(\alpha) = \widetilde{M}_{\text{bas}}(\beta)$. Let $\mathbf{c} \in \mathbb{N}^d$ and $P \subset [N]$ of length d. Then by Lemma 7.4.1, there exists an element $\ell_{\mathbf{c},P} \in \widetilde{\mathcal{H}}^*_{d,N}$ such that $\langle \ell_{\mathbf{c},P}, \widetilde{M}_{\text{bas}}(\gamma) \rangle = \int_{\square^d} \mathbf{t}^{\mathbf{c}} J[\gamma_P](\mathbf{t}) d\mathbf{t}$ for any $\gamma \in \mathsf{Diff}(\square^d, \mathbb{R}^N)$. Then, since $\widetilde{M}_{\text{bas}}(\alpha) = \widetilde{M}_{\text{bas}}(\beta)$, we have

$$\langle \ell_{\mathbf{c},P}, \widetilde{M}_{\text{bas}}(\alpha) - \widetilde{M}_{\text{bas}}(\beta) \rangle = \int_{\square^d} \mathbf{t}^{\mathbf{c}} \left(J[\alpha_P](\mathbf{t}) - J[\beta_P](\mathbf{t}) \right) d\mathbf{t} = 0$$

for all $\mathbf{c} \in \mathbb{N}^d$. By Lemma 7.4.2, this implies that

$$J[\alpha_P](\mathbf{t}) - J[\beta_P](\mathbf{t}) = 0,$$

which holds for all P. Therefore, $\alpha \sim_J \beta$ and $\widetilde{M}_{\text{bas}}$ is injective on $\widetilde{\text{Diff}}(\square^d, \mathbb{R}^N)$. □

Finally, it remains to prove Lemma 7.4.1.

Proof of Lemma 7.4.1. Fix $\mathbf{c} \in \mathbb{N}^d$ and $P \subset [N]$ of length d. The goal of this proof is to find a linear combination of basic monomials such that

$$\sum_{i=1}^{r} \lambda_i \widetilde{M}_{\text{bas}}^{\Pi_i, \mathcal{P}}(\gamma) = \int_{\square^d} \mathbf{t}^{\mathbf{c}} J[\gamma_P](\mathbf{t}) d\mathbf{t}, \tag{7.4.6}$$

where $\lambda_i \in \mathbb{R}$, $\Pi_i \in \Sigma_m^d$, and $\mathcal{P} = (V, \dots, V, P)$. Note that the forms index $\mathcal{P}$ is fixed in this sum.

Given this particular $\mathcal{P}$, the basic monomial is

$$\widetilde{M}_{\text{bas}}^{\Pi, \mathcal{P}} = \int_{D_{d,m}(\Pi)} J[\gamma_P](t_{m,1}, \dots, t_{m,d}) d\mathbf{t}. \tag{7.4.7}$$

This is an integral over $D_{d,m}(\Pi)$, parametrized by the variables $(t_{j,k})_{\substack{j=1,\dots,m \\ k=1,\dots d}}$, and thus we can integrate out all of the $t_{j,k}$ with $j < m$. Each component π_k of Π specifies the order of the variables $t_{1,k}, \dots, t_{m,k}$, and since there is no $t_{j,k}$ dependence for $j < m$, the integral only depends on $\pi_k^{-1}(m)$. Thus, we let

$$B = (b_1, \dots, b_d) = \left(\pi_1^{-1}(m), \dots, \pi_d^{-1}(m)\right), \tag{7.4.8}$$

which we treat as a multi-index in $[m]$ of length d. For $\mathbf{t}_m = (t_{m,1}, \dots, t_{m,d})$, we let

$$E_{d,m}(B)(\mathbf{t}_m) = \prod_{k=1}^{d} \Delta^{b_k - 1}(0, t_{m,k}) \times \Delta^{m - b_k}(t_{m,k}, 1), \tag{7.4.9}$$

where

$$\Delta^k(a, b) = \{a \leq t_1 < \dots < t_k < b\}.$$

We let $\hat{\mathbf{t}} = (t_{j,k})_{\substack{j=1,\dots,m-1 \\ k=1,\dots,d}}$ parametrize $E_{d,m}(B)(\mathbf{t}_m)$. Then, the integral in Equation (7.4.7) can be rewritten as

$$\widetilde{M}_{\text{bas}}^{\Pi, \mathcal{P}} = \int_{\mathbf{t}_m \in \square^d} \left(\int_{\hat{\mathbf{t}} \in E_{d,m}(B)(\mathbf{t}_m)} d\hat{\mathbf{t}} \right) J[\gamma_P](\mathbf{t}_m) d\mathbf{t}_m.$$

The inner integral is simply the volume of the region $E_{d,m}(B)(t_m)$, which we can compute by using the fact that the volume of the k-simplex $\Delta^k(a,b)$ is

$$\int_{\mathbf{t}\in\Delta^k(a,b)} d\mathbf{t} = \frac{(b-a)^k}{k!}.$$

Then, using the definition in Equation (7.4.9), the volume of $E_{d,m}(B)(\mathbf{t}_m)$ is given by

$$Q_{m,b}(\mathbf{t}) := \int_{\hat{\mathbf{t}}\in E_{d,m}(B)(\mathbf{t})} d\hat{\mathbf{t}} = \prod_{k=1}^{d} \frac{t_k^{b_k-1}}{(b_k-1)!}\frac{(1-t_k)^{m-b_k}}{(m-b_k)!}. \tag{7.4.10}$$

Therefore, our original problem of finding a linear combination of basic monomials in Equation (7.4.6) is reduced to finding some $m \in \mathbb{N}$ and subsets $B_i \subset [m]$ of length d such that a linear combination of the polynomials $Q_{m,B_i}(\mathbf{t}) \in \mathbb{R}[t_1, \dots t_d]$ satisfy

$$\sum_{i=1}^{r} \lambda_i Q_{m,B_i}(\mathbf{t}) = \mathbf{t}^{\mathbf{c}}, \tag{7.4.11}$$

for some $\lambda_i \in \mathbb{R}$. The choice of $m \in \mathbb{N}$ is straightforward: we will choose $m = \max_k\{c_k\} + 1$. Now, we must find the subsets B_i and coefficients λ_i.

We can further reduce the problem using the fact that $Q_{m,B}(\mathbf{t})$ is defined in factored form. Let

$$q_{m,b}(t) := \frac{t^{b-1}}{(b-1)!}\frac{(1-t)^{m-b}}{(m-b)!}, \tag{7.4.12}$$

so that given $B = (b_1, \dots, b_d) \subset [m]$, we have

$$Q_{m,B}(\mathbf{t}) = \prod_{k=1}^{d} q_{m,b_k}(t_k).$$

Suppose for each $k \in [d]$ there exists a linear combination

$$\sum_{i=1}^{r_k} \lambda_{k,i} q_{m,b_{k,i}}(t_k) = t^{c_k}, \tag{7.4.13}$$

where $\lambda_{k,i} \in \mathbb{R}$ and $b_{k,i} \in [m]$. Then,

$$\begin{aligned}\mathbf{t}^{\mathbf{c}} &= \prod_{k=1}^{d} \sum_{i=1}^{r_k} \lambda_{k,i} q_{m,b_{k,i}}(t_k) \\ &= \sum_{i_1=1}^{r_1} \cdots \sum_{i_d=1}^{r_d} (\lambda_{1,i_1} \dots \lambda_{d,i_d}) \left(\prod_{k=1}^{d} q_{m,b_{k,i_k}}(t_k) \right) \\ &= \sum_{i_1=1}^{r_1} \cdots \sum_{i_d=1}^{r_d} (\lambda_{1,i_1} \dots \lambda_{d,i_d}) Q_{m,(b_{1,i_1},\dots,b_{d,i_d})}(\mathbf{t}),\end{aligned}$$

providing the desired linear combination.

Thus, it suffices to show that for any $m \in \mathbb{N}$, and $c \in \mathbb{N}$ such that $c \leq m-1$ there exists a linear combination

$$\sum_{i=1}^{r} \lambda_i q_{b_i}(t) = t^c. \tag{7.4.14}$$

We fix $m \in \mathbb{N}$, and prove this by induction on c. The base case will be $c = m-1$, which is straightforward since

$$q_{m,m}(t) = \frac{t^{m-1}}{(m-1)!}.$$

Now, we assume the linear combination in Equation (7.4.14) holds down to c, and we will show the case of $c-1$. Consider

$$q_{m,c}(t) = \frac{t^{c-1}}{(c-1)!} \frac{(1-t)^{m-c}}{(m-c)!},$$

which contains a t^{c-1} term, as well as monomials of degree strictly greater than $c-1$. By the induction hypothesis, each of these monomials can be written as a linear combination of the $q_{m,b}(t)$. Thus, Equation (7.4.14) holds for any $m, c \in \mathbb{N}$ such that $c \leq m-1$.

□

7.4.2 *Smooth Maps*

In the parametrized mapping space signature considered in the previous section, we restricted the permissible forms on $\mathbb{R}^{d+N}$ to be either the standard d-volume form in the $\mathbb{R}^d$ factor or a standard d-form in the $\mathbb{R}^N$ factor. In this section, we will show that by using a different collection of standard d-forms on $\mathbb{R}^{d+N}$, the mapping space signature is injective on the space of all smooth maps in $\mathsf{Diff}(\square^d, \mathbb{R}^N)$.

In particular, consider the d-form

$$\nu_{i,k} \coloneqq du_1 \wedge \ldots \wedge du_{k-1} \wedge du_{k+1} \wedge \ldots \wedge du_d \wedge dx_j \in \Omega^d(\mathbb{R}^{d+N}). \tag{7.4.15}$$

Given a map $\gamma = (\gamma_1, \ldots, \gamma_N) \in \mathsf{Diff}(\square^d, \mathbb{R}^N)$ and its parametrized lift $\overline{\gamma} \coloneqq \iota(\gamma) \in \mathsf{Diff}(\square^d, \mathbb{R}^{d+N})$, the pullback of $\nu_{i,k}$ along $\overline{\gamma}$ is

$$\overline{\gamma}^*(\nu_{i,k})(\mathbf{t}) = \frac{\partial \gamma_i}{\partial t_k}(\mathbf{t}) \tag{7.4.16}$$

for all $\mathbf{t} \in \square^d$. We define a variant of the parametrized mapping space signature as follows. We continue to let $V = du_1 \wedge \ldots \wedge du_d$ denote the standard d-volume form in the $\mathbb{R}^d$ factor of $\mathbb{R}^{d+N}$. Then, let

$$W = \{V\} \cup \{\nu_{i,k}\}_{\substack{i \in [m] \\ k \in [d]}} \tag{7.4.17}$$

be the set of permissible forms. Let $\overline{\mathcal{H}}_{d,N} = \mathcal{H}_{d,W}$ be the Hilbert space from Definition 7.1.9, and define

$$\overline{M}_{\mathrm{bas}} : \mathsf{Diff}(\square^d, \mathbb{R}^N) \to \overline{\mathcal{H}}_{d,N}$$

to be $\overline{M}_{\mathrm{bas}} = M_{\mathrm{bas},W} \circ \iota$, where W is the set of permissible forms in Equation (7.4.17), $M_{\mathrm{bas},W}$ is the mapping space signature with respect to W, and ι is the parametrized lift in Equation (7.4.1). The injectivity theorem can then be proved in a similar manner as above.

Lemma 7.4.3. *For any $\mathbf{c} \in \mathbb{N}^d$, $i \in [N]$ and $k \in [d]$, there exists an element $\ell_{\mathbf{c},i,k} \in \overline{\mathcal{H}}^*_{d,N}$ such that*

$$\langle \ell_{\mathbf{c},i,k}, \overline{M}_{\mathrm{bas}}(\gamma) \rangle = \int_{\square^d} \mathbf{t}^{\mathbf{c}} \frac{\partial \gamma_i}{\partial t_k}(\mathbf{t}) d\mathbf{t}. \tag{7.4.18}$$

Proof. Fix $\mathbf{c} \in \mathbb{N}^d$, $i \in [N]$ and $k \in [d]$. The aim is to find a linear combination of basic monomials such that

$$\sum_{i=1}^{r} \lambda_i \overline{M}_{\mathrm{bas}}^{\Pi_i, \mathcal{P}}(\gamma) = \int_{\square^d} \mathbf{t}^{\mathbf{c}} \frac{\partial \gamma_j}{\partial t_k}(\mathbf{t}) d\mathbf{t}, \tag{7.4.19}$$

where $\lambda_i \in \mathbb{R}$, $\Pi_i \in \Sigma_m^d$, and $\mathcal{P} = (V, \ldots, V, \nu_{i,k})$ in which there are $m-1$ instances of V. Once again, the forms index $\mathcal{P}$ is fixed in the linear combination. Given $\Pi \in \Sigma_m^d$, the parametrized basic monomial with respect to this choice of $\mathcal{P}$ is

$$\overline{M}_{\text{bas}}^{\Pi,\mathcal{P}}(\gamma) = \int_{D_{d,m}(\Pi)} \frac{\partial \gamma_i}{\partial t_k}(t_{m,1}, \ldots, t_{m,d}) d\mathbf{t},$$

which has the same form as the basic monomial in Equation (7.4.7). Thus, we can use the same procedure as in Lemma 7.4.1 to show the existence of the desired linear combination. □

This can then be used to prove the second variant of the parametrized injectivity theorem, which proceeds in the same manner as Theorem 7.4.1

Theorem 7.4.2. *The parametrized mapping space signature*

$$\overline{M}_{\text{bas}} : \overline{\mathsf{Diff}}(\square^d, \mathbb{R}^N) \to \overline{\mathcal{H}}_{d,N}$$

is injective up to translation.

Proof. Suppose $\alpha, \beta \in \mathsf{Diff}(\square^d, \mathbb{R}^N)$ such that $\overline{M}_{\text{bas}}(\alpha) = \overline{M}_{\text{bas}}(\beta)$. Let $\mathbf{c} \in \mathbb{N}^d$, $i \in [N]$, and $k \in [d]$. By Lemma 7.4.3, there exists an element $\ell_{\mathbf{c},i,k} \in \overline{\mathcal{H}}_{d,N}^*$ such that

$$\langle \ell_{\mathbf{c},i,k}, \overline{M}_{\text{bas}}(\gamma) \rangle = \int_{\square^d} \mathbf{t}^{\mathbf{c}} \frac{\partial \gamma_i}{\partial t_k}(\mathbf{t}) d\mathbf{t}$$

for any $\gamma \in \mathsf{Diff}(\square^d, \mathbb{R}^N)$. Because $\overline{M}_{\text{bas}}(\alpha) = \overline{M}_{\text{bas}}(\beta)$, we have

$$\langle \ell_{\mathbf{c},i,k}, \overline{M}_{\text{bas}}(\alpha) - \overline{M}_{\text{bas}}(\beta) \rangle = \int_{\square^d} \mathbf{t}^{\mathbf{c}} \left(\frac{\partial \alpha_i}{\partial t_k}(\mathbf{t}) - \frac{\partial \beta_i}{\partial t_k}(\mathbf{t}) \right) d\mathbf{t}$$

for all $\mathbf{c} \in \mathbb{N}^d$. By Lemma 7.4.2, this implies that

$$\frac{\partial \alpha_i}{\partial t_k}(\mathbf{t}) = \frac{\partial \beta_i}{\partial t_k}(\mathbf{t})$$

for all $i \in [N]$, $k \in [d]$, and $\mathbf{t} \in \square^d$. Because all derivatives of α and β are equivalent, we have $\alpha(\mathbf{t}) = \beta(\mathbf{t}) + \mathbf{a}$, where $\mathbf{a} \in \mathbb{R}^N$. □

7.5 COMPOSITION OF MAPS

In this section, we consider the generalization of Chen's identity to higher dimensions. One of the primary reasons for using cubical domains as the higher dimensional generalization

of paths (rather than simplices, for example) is the existence of natural composition or concatenation operations. In order to define composition in the smooth setting, we must consider maps with a *stationary boundary*. Chen [Che77] called paths with this condition *paths which start and end smoothly*.

Definition 7.5.1. Let $\gamma \in \mathsf{Diff}(\square^d, \mathbb{R}^N)$. The map γ has a *stationary boundary* if there exists some $\epsilon \in (0, 1/2)$ such that

$$\begin{aligned} \gamma(t_1, \dots, t_k, \dots, t_d) &= \gamma(t_1, \dots, 0, \dots, t_d) \\ \gamma(t_1, \dots, 1 - t_k, \dots, t_d) &= \gamma(t_1, \dots, 1, \dots, t_d) \end{aligned}$$

if $t_k < \epsilon$ holds for all $k \in [d]$.

Lemma 7.5.1. *Let $\gamma \in \mathsf{Diff}(\square^d, \mathbb{R}^N)$. There exists a smooth reparametrization $\phi : \square^d \to \square^d$ such that $\gamma \circ \phi$ has a stationary boundary.*

Proof. Let $\epsilon \in (0, 1/2)$. Suppose $f : [0,1] \to [0,1]$ is a smooth bump function supported on $[\epsilon, 1-\epsilon]$ such that $\int_0^1 f(t)dt = 1$. Then, let

$$g(t) = \int_0^t f(t')dt',$$

which has the property that $g(t) = 0$ when $t < \epsilon$ and $g(t) = 1$ when $t > 1 - \epsilon$. Let $\phi(t_1, \dots, t_d) = (g(t_1), \dots, g(t_d))$. Then $\gamma \circ \phi$ has a stationary boundary. □

We can now define composition for maps with stationary boundary.

Definition 7.5.2. Let $\alpha, \beta \in \mathsf{Diff}(\square^d, \mathbb{R}^N)$ be maps with stationary boundaries and let $k \in [d]$. If

$$\alpha(t_1, \dots, t_{k-1}, 1, t_{k+1}, \dots, t_d) = \beta(t_1, \dots, t_{k-1}, 0, t_{k+1}, \dots, t_d)$$

for all $(t_1, \dots, t_{k-1}, t_{k+1}, \dots, t_d) \in \square^{d-1}$, then we say that α and β are k-composable. Given k-composable maps α and β, the *k-composition* $\alpha *_k \beta \in \mathsf{Diff}(\square^d, \mathbb{R}^N)$ is defined as

$$\alpha *_k \beta(t_1, \dots, t_d) := \begin{cases} \alpha(t_1, \dots, 2t_k, \dots, t_d) & : t_k \in [0, 1/2) \\ \beta(t_1, \dots, 2t_k - 1, \dots, t_d) & : t_k \in [1/2, 1] \end{cases} \tag{7.5.1}$$

The k-composable condition ensures that the k-composition is continuous, and the requirement that α and β have stationary boundaries ensures that the resulting k-composition is smooth. By Lemma 7.5.1, any map $\gamma \in \mathsf{Diff}(\square^d, \mathbb{R}^N)$ can be reparametrized into one with stationary boundary. Furthermore, by Proposition 7.2.3, the mapping space signature is invariant under this reparametrization. Thus from this point forward, we will omit

the stationary boundary condition when discussing compositions, but we will tacitly reparametrize all maps as necessary.

In Section 6.4.3, we used the formal composition rules for cubical sets to derive Chen's identity for the cubical path signature monomials. To summarize the derivation, we showed Chen's identity in two steps.

1. Using a modified cubical model $L_\bullet$ for the interval $[0,2]$ with three non-degenerate 1-cubes e_{01}, e_{12} and e_{02} such that $e_{02} = e_{01} +_1 e_{12}$, we showed that any cubical monomial for $L_\bullet$ where the degeneracy index $\mathcal{I}$ only contains degeneracies of e_{02} can be decomposed into a sum of cubical monomials using the degeneracies of e_{01} and e_{12}. This was shown in Theorem 6.4.1.

2. For the special case of cubical monomials indexed by flags, which were shown to be equivalent to the simplicial path signature monomials in Proposition 6.2.1, each term of the decomposition from the first step can be further decomposed as a product of a monomial using only e_{01} degeneracies and a monomial using only e_{12} degeneracies.

The first step can theoretically be accomplished for the mapping space signature, although the manipulations of the formal composition rules for cubical sets become much more complex due to the larger variety of degeneracies for the non-degenerate d-cubes, as well as the increase in the number of possible compositions for these non-degenerate d-cubes. Moreover, we find that the second step is not possible in higher dimensions: the basic monomials of a composite map $\alpha *_k \beta$ *cannot* be expressed as a product of the basic monomials of α and the basic monomials of β.

We illustrate this obstruction with an example. We will consider $d = 2$ for maps from the domain $\square^2$, and level $m = 2$. Suppose $\alpha, \beta \in \mathsf{Diff}(\square^2, \mathbb{R}^N)$ are 1-composable, and let $\gamma = \alpha *_1 \beta$ be their 1-composition. We let the permutation index be the identity $\mathrm{id} \in \Sigma_2^2$ and $\mathcal{P} = (P_1, P_2)$, where $P_1, P_2 \subset [N]$ of length 2. Then,

$$M_{\mathrm{bas}}^{\mathrm{id},\mathcal{P}}(\gamma) = \int_{(\mathbf{s},\mathbf{t})\in\Delta^2\times\Delta^2} J[\gamma_{P_1}](s_1,t_1)J[\gamma_{P_2}](s_2,t_2)d\mathbf{s}d\mathbf{t},$$

where we use $\mathbf{s}$ and $\mathbf{t}$ variables rather than the double indexed $t_{j,k}$ in order to better distinguish the variables $s_1 < s_2$ in the first Δ^2 and $t_1 < t_2$ in the second Δ^2. In order to express the integral in terms of α and β, we decompose the first Δ^2 as

$$\Delta^2 = \Delta^2(0,1/2) \cup ([0,1/2] \times [1/2,1]) \cup \Delta^2(1/2,1).$$

Using the definition of the 1-composition in Equation (7.5.1), the basic monomial above can be expressed as

$$\begin{aligned} M_{\text{bas}}^{\text{id},\mathcal{P}}(\gamma) &= \int_{(\mathbf{s},\mathbf{t})\in\Delta^2(0,1/2)\times\Delta^2} J[\gamma_{P_1}](s_1,t_1)J[\gamma_{P_2}](s_2,t_2)d\mathbf{s}d\mathbf{t} \\ &\quad + \int_{(\mathbf{s},\mathbf{t})\in([0,1/2]\times[1/2,1])\times\Delta^2} J[\gamma_{P_1}](s_1,t_1)J[\gamma_{P_2}](s_2,t_2)d\mathbf{s}d\mathbf{t} \\ &\quad + \int_{(\mathbf{s},\mathbf{t})\in\Delta^2(1/2,1)\times\Delta^2} J[\gamma_{P_1}](s_1,t_1)J[\gamma_{P_2}](s_2,t_2)d\mathbf{s}d\mathbf{t} \\ &= \int_{(\mathbf{s},\mathbf{t})\in\Delta^2\times\Delta^2} J[\alpha_{P_1}](s_1,t_1)J[\alpha_{P_2}](s_2,t_2)d\mathbf{s}d\mathbf{t} \\ &\quad + \int_{(\mathbf{s},\mathbf{t})\in\square^2\times\Delta^2} J[\alpha_{P_1}](s_1,t_1)J[\beta_{P_2}](s_2,t_2)d\mathbf{s}d\mathbf{t} \\ &\quad + \int_{\Delta^2\times\Delta^2} J[\beta_{P_1}](s_1,t_1)J[\beta_{P_2}](s_2,t_2)d\mathbf{s}d\mathbf{t} \end{aligned}$$

We can see that the first and third integrals in this decomposition correspond to basic monomials for α and β, and thus we have

$$M_{\text{bas}}^{\text{id},\mathcal{P}}(\gamma) = M_{\text{bas}}^{\text{id},\mathcal{P}}(\alpha) + \int_{(\mathbf{s},\mathbf{t})\in\square^2\times\Delta^2} J[\alpha_{P_1}](s_1,t_1)J[\beta_{P_2}](s_2,t_2)d\mathbf{s}d\mathbf{t} + M_{\text{bas}}^{\text{id},\mathcal{P}}(\beta). \quad (7.5.2)$$

The obstruction to a formula such as Chen's identity is the remaining integral in the above expression. Due to the nontrivial relationship $t_1 < t_2$ in the domain of integration, we cannot express this integral as a product of a monomial of α and a monomial of β. However, we can express it as the *integral* of a product of monomials. By expressing the domain of integration explicitly, the integral can be written as

$$\begin{aligned} &\int_{(\mathbf{s},\mathbf{t})\in\square^2\times\Delta^2} J[\alpha_{P_1}](s_1,t_1)J[\beta_{P_2}](s_2,t_2)d\mathbf{s}d\mathbf{t} \quad (7.5.3) \\ &\qquad = \int_{t_1=0}^1 \left(\int_{s_1=0}^1 J[\alpha_{P_1}](s_1,t_1)ds_1\right)\left(\int_{s_2=0}^1\int_{t_2=t_1}^1 J[\beta_{P_2}](s_2,t_2)ds_2dt_2\right)dt_1. \end{aligned}$$

This integral can be expressed in terms of basic monomials of α and β restricted to subdomains, which was previous defined in Proposition 7.1.1. From the definition in the Proposition, we can express

$$M_{\text{bas}}^{\text{id},\mathcal{P}}(\gamma)_{\mathbf{a},\mathbf{b}} := \int_{D_{d,m}(\text{id})(\mathbf{a},\mathbf{b})} \prod_{j=1}^m J[\gamma_{P_j}](t_{j,1},\ldots,t_{j,d})d\mathbf{t}, \quad (7.5.4)$$

where

$$D_{d,m}(\mathrm{id})(\mathbf{a},\mathbf{b}) := \prod_{k=1}^{d} \Delta^m(a_k, b_k). \tag{7.5.5}$$

In other words, $M_{\mathrm{bas}}^{\mathrm{id},\mathcal{P}}(\gamma)_{\mathbf{a},\mathbf{b}}$ is the basic monomial of γ restricted to

$$[a_1, b_1] \times \ldots \times [a_d, b_d].$$

Note that we can take derivatives of this monomial with respect to the elements in $\mathbf{a}$ and $\mathbf{b}$. Returning to our example in terms of α and β, consider $\mathbf{a} = (0,0)$ and $\mathbf{b} = (1,t)$. Then the derivative with respect to t is

$$\begin{aligned}\frac{\partial}{\partial t} M_{\mathrm{bas}}^{\mathrm{id},(P_1)}(\alpha)_{\mathbf{a},\mathbf{b}} &= \frac{\partial}{\partial t} \int_{s_1=0}^{1} \int_{t_1=0}^{t} J[\alpha_{P_1}](s_1,t_1) ds_1 dt_1 \\ &= \int_{s_1=0} J[\alpha_{P_1}](s_1,t) ds_1.\end{aligned}$$

Finally, we can express the integral in Equation (7.5.3) as

$$\begin{aligned}&\int_{(\mathbf{s},\mathbf{t})\in\square^2\times\Delta^2} J[\alpha_{P_1}](s_1,t_1) J[\beta_{P_2}](s_2,t_2) d\mathbf{s} d\mathbf{t} \\ &\qquad = \int_0^1 \frac{\partial}{\partial t} M_{\mathrm{bas}}^{\mathrm{id},(P_1)}(\alpha)_{(0,0),(1,t)} \cdot M_{\mathrm{bas}}^{\mathrm{id},(P_2)}(\beta)_{(0,t),(1,1)} dt.\end{aligned}$$

Thus, the basic monomial of the composition $\alpha *_1 \beta)$ is

$$\begin{aligned}&M_{\mathrm{bas}}^{\mathrm{id},\mathcal{P}}(\alpha *_1 \beta) \\ &\qquad = M_{\mathrm{bas}}^{\mathrm{id},\mathcal{P}}(\alpha) + \int_0^1 \frac{\partial}{\partial t} M_{\mathrm{bas}}^{\mathrm{id},(P_1)}(\alpha)_{(0,0),(1,t)} \cdot M_{\mathrm{bas}}^{\mathrm{id},(P_2)}(\beta)_{(0,t),(1,1)} dt + M_{\mathrm{bas}}^{\mathrm{id},\mathcal{P}}(\beta).\end{aligned} \tag{7.5.6}$$

This composition identity can be established for any basic monomial where the permutation index is the identity $\mathrm{id} \in \Sigma_m^d$.

Theorem 7.5.1. *Let $\alpha, \beta \in \mathsf{Diff}(\square^d, \mathbb{R}^N)$ such that α and β are i-composable for some $i \in [d]$. Let $\mathcal{P} = (P_1, \ldots, P_m)$ where each $P_j \subset [N]$ is of length d, and let*

$$\underline{\mathcal{P}}_r = (P_1, \ldots, P_j) \tag{7.5.7}$$

$$\overline{\mathcal{P}}_r = (P_{m-r+1}, \ldots, P_m) \tag{7.5.8}$$

denote the first r and last r elements of $\mathcal{P}$ respectively. Then, we will use

$$\hat{\mathbf{t}}_i = (t_1, \ldots, t_{i-1}, t_{i+1}, \ldots, t_d) \in \square^{d-1}$$

to denote an element of $\square^{d-1}$ *with the* t_i *coordinate omitted. Let*

$$\hat{\mathbf{t}}_i^0 = (t_1, \ldots, t_{i-1}, 0, t_{i+1}, \ldots, t_d) \in \square^d$$
$$\hat{\mathbf{t}}_i^1 = (t_1, \ldots, t_{i-1}, 1, t_{i+1}, \ldots, t_d) \in \square^d,$$

denote the the corresponding coordinates on the 0 *or* 1 *face of* $\square^d$ *in the i-direction. The all zeros and all ones vector are denoted by* $\mathbf{0} = (0, \ldots, 0) \in \square^d$ *and* $\mathbf{1} = (1, \ldots, 1) \in \square^d$ *respectively. Finally, let*

$$\frac{\partial}{\partial \hat{\mathbf{t}}_i} := \frac{\partial}{\partial t_1} \cdots \frac{\partial}{\partial t_{i-1}} \frac{\partial}{\partial t_{i+1}} \cdots \frac{\partial}{\partial t_d}.$$

Then, we have

$$M_{\text{bas}}^{\text{id},\mathcal{P}}(\alpha *_i \beta) = M_{\text{bas}}^{\text{id},\mathcal{P}}(\alpha) + M_{\text{bas}}^{\text{id},\mathcal{P}}(\beta) + \sum_{r=1}^{m-1} \int_{\hat{\mathbf{t}}_i \in \square^{d-1}} \frac{\partial}{\partial \hat{\mathbf{t}}_i} M_{\text{bas}}^{\text{id},\underline{\mathcal{P}}_r}(\alpha)_{\mathbf{0},\hat{\mathbf{t}}_i^1} \cdot M_{\text{bas}}^{\text{id},\overline{\mathcal{P}}_{m-r}}(\beta)_{\hat{\mathbf{t}}_i^0,\mathbf{1}} \, d\hat{\mathbf{t}}_i. \tag{7.5.9}$$

Proof. We begin with the following decomposition of the standard m-simplex

$$\Delta^m = \coprod_{r=0}^{m} \Delta^r(0, 1/2) \times \Delta^{m-r}(1/2, 1). \tag{7.5.10}$$

The definition of $M_{\text{bas}}^{\text{id},\mathcal{P}}(\alpha *_k \beta)$ is an integral over

$$D_{d,m}(\text{id}) = \prod_{k=1}^{d} \Delta^m,$$

and we apply the decomposition in Equation (7.5.10) to the i^{th} Δ^m in $D_{d,m}(\text{id})$. By the definition of i-composition in Equation (7.5.1), this allows us to express the Jacobians in the integrand in terms of α and β individually, after a change of variables which takes

$$\Delta^r(0, 1/2) \times \Delta^{m-r}(1/2, 1) \mapsto \Delta^r \times \Delta^{m-r}.$$

We will use

$$\mathbb{D}_{d,m}^{i,r} := (\Delta^m)^{i-1} \times (\Delta^r \times \Delta^{m-r}) \times (\Delta^m)^{d-i}, \tag{7.5.11}$$

to denote the resulting domain of integration. We will continue to use $t_{j,k}$ as the coordinates of $\mathbb{D}_{d,m}^{i,r}$, where the $(t_{j,i})_{j \in [m]}$ variables are the coordinates for $(\Delta^r \times \Delta^{m-r})$.

In particular, we have

$$M^{\mathrm{id},\mathcal{P}}_{\mathrm{bas}}(\alpha *_i \beta) = M^{\mathrm{id},\mathcal{P}}_{\mathrm{bas}}(\alpha) + M^{\mathrm{id},\mathcal{P}}_{\mathrm{bas}}(\beta) + \sum_{r=1}^{m-1} \int_{\mathbb{D}^{i,r}_{d,m}} \prod_{j=1}^{r} J[\alpha_{P_j}](\mathbf{t}_j) \cdot \prod_{j=r+1}^{m} J[\beta_{P_j}](\mathbf{t}_j) d\mathbf{t}. \tag{7.5.12}$$

Here, we use $\mathbf{t}_j = (t_{j,1}, \ldots, t_{j,d}) \in \square^d$ to supress the notation. The integrals in the summand can be rewritten by changing the order of integration. In particular, we integrate the $t_{r,1}, \ldots, t_{r,i-1}, t_{r,i+1}, \ldots, t_{r,d}$ at the end. In the following decomposition, given $\mathbf{a} = (a_1, \ldots, a_d), \mathbf{b} = (b_1, \ldots, b_d) \in \square^d$, the domain $D_{d,m}(\mathbf{a}, \mathbf{b})$ is defined as $D_{d,m}(\mathrm{id})(\mathbf{a}, \mathbf{b})$ in Equation (7.5.5), and we let

$$D^i_{d,m}(\mathbf{a}, \mathbf{b}) \coloneqq (\Delta^{m-1})^{i-1} \times \Delta^m \times (\Delta^{m-1})^{d-i}.$$

In order to preserve the labelling of the coordinates, we let

$$\hat{\mathbf{t}}_i = (t_{r,1}, \ldots, t_{r,i-1}, t_{r,i+1}, \ldots, t_{r,i})$$

in the following expressions. By changing the order of integration, the integrals in the summand above can be expressed as

$$\int_{D^{i,r}_{d,m}} \prod_{j=1}^{r} J[\alpha_{P_j}](\mathbf{t}_j) \cdot \prod_{j=r+1}^{m} J[\beta_{P_j}](\mathbf{t}_j) d\mathbf{t} = \int_{\hat{\mathbf{t}}_i \in \square^{d-1}} \left(\int_{D^i_{d,r}(\mathbf{0}, \hat{\mathbf{t}}^1_i)} \prod_{j=1}^{r} J[\alpha_{P_j}](\mathbf{t}_j) d\mathbf{t} \right) \cdot \left(\int_{D_{d,m-r}(\hat{\mathbf{t}}^0_i, \mathbf{1})} \prod_{j=r+1}^{m} J[\beta_{P_j}](\mathbf{t}_j) d\mathbf{t} \right) d\hat{\mathbf{t}}_i.$$

By definition, the second integral in the integrand is

$$M^{\mathrm{id},\overline{\mathcal{P}}_{m-r}}_{\mathrm{bas}}(\beta)_{\hat{\mathbf{t}}^0_i, \mathbf{1}} = \int_{D_{d,m-r}(\hat{\mathbf{t}}^0_i, \mathbf{1})} \prod_{j=r+1}^{m} J[\beta_{P_j}](\mathbf{t}_j) d\mathbf{t}.$$

Furthermore, it can easily be checked by iterated application of the fundamental theorem of calculus that

$$\frac{\partial}{\partial \hat{\mathbf{t}}_i} M^{\mathrm{id},\underline{\mathcal{P}}_r}_{\mathrm{bas}}(\alpha)_{\mathbf{0}, \hat{\mathbf{t}}^1_i} = \frac{\partial}{\partial \hat{\mathbf{t}}_i} \int_{D_{d,r}(\mathbf{0}, \hat{\mathbf{t}}^1_i)} \prod_{j=1}^{r} J[\alpha_{P_j}](\mathbf{t}_j) d\mathbf{t} = \int_{D^i_{d,r}(\mathbf{0}, \hat{\mathbf{t}}^1_i)} \prod_{j=1}^{r} J[\alpha_{P_j}](\mathbf{t}_j) d\mathbf{t},$$

which is the first integral in the integral. Thus, we can express

$$\int_{D^{i,r}_{d,m}} \prod_{j=1}^{r} J[\alpha_{P_j}](\mathbf{t}_j) \cdot \prod_{j=r+1}^{m} J[\beta_{P_j}](\mathbf{t}_j) d\mathbf{t} = \int_{\hat{\mathbf{t}}_i \in \square^{d-1}} \frac{\partial}{\partial \hat{\mathbf{t}}_i} M^{\mathrm{id},\mathcal{P}_r}_{\mathrm{bas}}(\alpha)_{\mathbf{0},\hat{\mathbf{t}}^1_i} \cdot M^{\mathrm{id},\overline{\mathcal{P}}_{m-r}}_{\mathrm{bas}}(\beta)_{\hat{\mathbf{t}}^0_i,\mathbf{1}} d\hat{\mathbf{t}}_i.$$

Finally, substituing this integral back into Equation (7.5.12), we obtain the desired composition formula. □

BIBLIOGRAPHY

[Ada+17] H. Adams, T. Emerson, M. Kirby, R. Neville, C. Peterson, P. Shipman, S. Chepushtanova, E. Hanson, F. Motta, and L. Ziegelmeier. "Persistence Images: A Stable Vector Representation of Persistent Homology." In: *J. Mach. Learn. Res.* 18.8 (2017), pp. 1–35.

[AB15] M. Alexandrino and R. Bettiol. *Lie Groups and Geometric Aspects of Isometric Actions*. Springer International Publishing, 2015.

[AFS19] C. Améndola, P. Friz, and B. Sturmfels. "Varieties of Signature Tensors." In: *Forum Math. Sigma* 7 (2019), e10.

[And72] D. W. Anderson. "A Generalization of the Eilenberg-Moore Spectral Sequence." EN. In: *Bull. Amer. Math. Soc.* 78.5 (Sept. 1972), pp. 784–786.

[Ant02] R. Antolini. "Geometric Realisations of Cubical Sets with Connections, and Classifying Spaces of Categories." en. In: *Appl. Categ. Structures* 10.5 (2002), pp. 481–494.

[Aro50] N. Aronszajn. "Theory of Reproducing Kernels." In: *Trans. Amer. Math. Soc.* 68.3 (1950), pp. 337–404.

[Bac+20] J. Backhoff-Veraguas, D. Bartl, M. Beiglböck, and M. Eder. "All Adapted Topologies Are Equal." en. In: *Probab. Theory Related Fields* 178.3 (Dec. 2020), pp. 1125–1172.

[BH11] J. C. Baez and A. E. Hoffnung. "Convenient Categories of Smooth Spaces." In: *Trans. Amer. Math. Soc.* 363.11 (2011), pp. 5789–5825.

[BG18] Y. Baryshnikov and R. Ghrist. "Stokes's Theorem, Data, and the Polar Ice Caps." In: *Amer. Math. Monthly* 125.9 (Oct. 2018), pp. 830–834.

[BS16] Y. Baryshnikov and E. Schlafly. "Cyclicity in Multivariate Time Series and Applications to Functional MRI Data." In: *2016 IEEE 55th Conference on Decision and Control (CDC)*. Dec. 2016, pp. 1625–1630.

[BG91] M. Bendersky and S. Gitler. "The Cohomology of Certain Function Spaces." In: *Trans. Amer. Math. Soc.* 326.1 (1991), pp. 423–440.

[Bha+19] D. Bhaskar, A. Manhart, J. Milzman, J. T. Nardini, K. M. Storey, C. M. Topaz, and L. Ziegelmeier. "Analyzing Collective Motion with Machine Learning and Topology." In: *Chaos* 29.12 (2019), p. 123125.

[BMA12] V. Bloom, D. Makris, and V. Argyriou. "G3D: A Gaming Action Dataset and Real Time Action Recognition Evaluation Framework." In: *2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. June 2012, pp. 7–12.

[Boe+15] H. Boedihardjo, X. Geng, T. Lyons, and D. Yang. "Note on the Signatures of Rough Paths in a Banach Space." In: *arXiv:1510.04172 [math]* (Oct. 2015). arXiv: 1510.04172 [math].

[Boe+16] H. Boedihardjo, X. Geng, T. Lyons, and D. Yang. "The Signature of a Rough Path: Uniqueness." en. In: *Adv. Math.* 293 (Apr. 2016), pp. 720–737.

[BLO20] P. Bonnier, C. Liu, and H. Oberhauser. "Adapted Topologies and Higher Rank Signatures." In: *arXiv:2005.08897 [math]* (2020). arXiv: 2005.08897 [math].

[Bor+06] K. M. Borgwardt, A. Gretton, M. J. Rasch, H.-P. Kriegel, B. Schölkopf, and A. J. Smola. "Integrating Structured Biological Data by Kernel Maximum Mean Discrepancy." en. In: *Bioinformatics* 22.14 (July 2006), e49–e57.

[BS77] R. Bott and G. Segal. "The Cohomology of the Vector Fields on a Manifold." en. In: *Topology* 16.4 (Jan. 1977), pp. 285–298.

[BHS11] R. Brown, P. Higgins, and R. Sivera. *Nonabelian Algebraic Topology: Filtered Spaces, Crossed Complexes, Cubical Homotopy Groupoids*. EMS Series of Lectures in Mathematics. European Mathematical Society, 2011.

[BH81] R. Brown and P. J. Higgins. "On the Algebra of Cubes." en. In: *J. Pure Appl. Algebra* 21.3 (June 1981), pp. 233–260.

[Bub15] P. Bubenik. "Statistical Topological Data Analysis Using Persistence Landscapes." In: *J. Mach. Learn. Res.* 16 (2015), pp. 77–102.

[BE20a] P. Bubenik and A. Elchesen. "Universality of Persistence Diagrams and the Bottleneck and Wasserstein Distances." In: *arXiv:1912.02563 [cs, math]* (Sept. 2020). arXiv: 1912.02563 [cs, math].

[BE20b] P. Bubenik and A. Elchesen. "Virtual Persistence Diagrams, Signed Measures, and Wasserstein Distance." In: *arXiv:2012.10514 [math]* (Dec. 2020). arXiv: 2012.10514 [math].

[BW19] P. Bubenik and A. Wagner. "Embeddings of Persistence Diagrams into Hilbert Spaces." In: *arXiv:1905.05604 [cs, math, stat]* (May 2019). arXiv: 1905.05604 [cs, math, stat].

[BdS15] P. Bubenik, V. de Silva, and J. Scott. "Metrics for Generalized Persistence Modules." en. In: *Found. Comput. Math.* 15.6 (Dec. 2015), pp. 1501–1531.

[Bur15] A. Y. Burtscher. "Length Structures on Manifolds with Continuous Riemannian Metrics." en. In: *New York J. Math.* 21 (2015), pp. 273–296.

[CB19] M. Carrière and U. Bauer. "On the Metric Distortion of Embedding Persistence Diagrams into Separable Hilbert Spaces." In: *35th International Symposium on Computational Geometry (SoCG 2019)*. Ed. by G. Barequet and Y. Wang. Vol. 129. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2019, 21:1–21:15.

[CCO17] M. Carrière, M. Cuturi, and S. Oudot. "Sliced Wasserstein Kernel for Persistence Diagrams." en. In: *Proceedings of the 34th International Conference on Machine Learning*. July 2017, pp. 664–673.

[Cas+21] T. Cass, J. Foster, T. Lyons, C. Salvi, and W. Yang. "Computing the Untruncated Signature Kernel as the Solution of a Goursat Problem." In: *arXiv:2006.14794 [cs, math]* (Jan. 2021). arXiv: 2006.14794 [cs, math].

[CLT19] E. Celledoni, P. E. Lystad, and N. Tapia. "Signatures in Shape Analysis: An Efficient Approach to Motion Identification." en. In: *Geometric Science of Information*. Ed. by F. Nielsen and F. Barbaresco. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2019, pp. 21–30.

[Cha+17] J. Chang, N. Duffield, H. Ni, and W. Xu. "Signature Inversion for Monotone Paths." EN. In: *Electron. Commun. Probab.* 22 (2017).

[Cha+16] F. Chazal, V. de Silva, M. Glisse, and S. Oudot. *The Structure and Stability of Persistence Modules*. en. Springer Briefs in Mathematics. Springer International Publishing, 2016.

[Che54] K.-T. Chen. "Iterated Integrals and Exponential Homomorphisms†." In: *Proc. Lond. Math. Soc.* s3-4.1 (1954), pp. 502–512.

[Che57] K.-T. Chen. "Integration of Paths, Geometric Invariants and a Generalized Baker- Hausdorff Formula." In: *Ann. of Math.* 65.1 (1957), pp. 163–178.

[Che58] K.-T. Chen. "Integration of Paths – a Faithful Representation of Paths by Noncommutative Formal Power Series." In: *Trans. Amer. Math. Soc.* 89.2 (1958), pp. 395–407.

[Che73] K.-T. Chen. "Iterated Integrals of Differential Forms and Loop Space Homology." In: *Ann. of Math.* 97.2 (1973), pp. 217–246.

[Che77] K.-T. Chen. "Iterated Path Integrals." EN. In: *Bull. Amer. Math. Soc.* 83.5 (1977), pp. 831–879.

[CK16] I. Chevyrev and A. Kormilitzin. "A Primer on the Signature Method in Machine Learning." In: *arXiv:1603.03788 [cs, stat]* (2016). arXiv: 1603.03788 [cs, stat].

[CNO20] I. Chevyrev, V. Nanda, and H. Oberhauser. "Persistence Paths and Signature Features in Topological Data Analysis." In: *IEEE Trans. Pattern Anal. Mach. Intell.* 42.1 (2020), pp. 192–202.

[CO18] I. Chevyrev and H. Oberhauser. "Signature Moments to Characterize Laws of Stochastic Processes." In: *arXiv:1810.10971 [math, stat]* (2018). arXiv: 1810.10971 [math, stat].

[CG98] V. V. Chistyakov and O. E. Galkin. "On Maps of Bounded P-Variation with P>1." en. In: *Positivity* 2.1 (Mar. 1998), pp. 19–45.

[CG14] K. Chouk and M. Gubinelli. "Rough Sheets." In: *arXiv:1406.7748 [math]* (2014). arXiv: 1406.7748 [math].

[CEM06] D. Cohen-Steiner, H. Edelsbrunner, and D. Morozov. "Vines and Vineyards by Updating Persistence in Linear Time." In: *Proceedings of the 22nd Annual Symposium on Computational Geometry*. SCG '06. New York, NY, USA: ACM, 2006, pp. 119–126.

[CDW16] M. Cúth, M. Doucha, and P. Wojtaszczyk. "On the Structure of Lipschitz-Free Spaces." en. In: *Proc. Amer. Math. Soc.* 144.9 (2016), pp. 3833–3846.

[CKK17] M. Cúth, O. F. K. Kalenda, and P. Kaplický. "Isometric Representation of Lipschitz-Free Spaces over Convex Domains in Finite Dimensional Spaces." en. In: *Mathematika* 63.2 (2017/ed), pp. 538–552.

[D'O+06] M. R. D'Orsogna, Y. L. Chuang, A. L. Bertozzi, and L. S. Chayes. "Self-Propelled Particles with Soft-Core Interactions: Patterns, Stability, and Collapse." en. In: *Phys. Rev. Lett.* 96.10 (Mar. 2006).

[DS87] P. Diaconis and M. Shahshahani. "The Subgroup Algorithm for Generating Uniform Random Variables." en. In: *Probab. Engrg. Inform. Sci.* 1.1 (Jan. 1987), pp. 15–32.

[DET20] J. Diehl, K. Ebrahimi-Fard, and N. Tapia. "Time-Warping Invariants of Multidimensional Time Series." en. In: *Acta Appl. Math.* 170.1 (Dec. 2020), pp. 265–290.

[DR19] J. Diehl and J. Reizenstein. "Invariants of Multidimensional Time Series Based on Their Iterated-Integral Signature." en. In: *Acta Appl. Math.* 164.1 (Dec. 2019), pp. 83–122.

[DL21] V. Divol and T. Lacombe. "Understanding the Topology and the Geometry of the Space of Persistence Diagrams via Optimal Partial Transport." en. In: *J. Appl. Comput. Topol.* 5.1 (Mar. 2021), pp. 1–53.

[DP19] V. Divol and W. Polonik. "On the Choice of Weight Functions for Linear Representations of Persistence Diagrams." en. In: *J. Appl. Comput. Topol.* 3.3 (Sept. 2019), pp. 249–283.

[Dru+97] H. Drucker, C. J. C. Burges, L. Kaufman, A. Smola, and V. Vapnik. "Support Vector Regression Machines." In: *Advances in Neural Information Processing Systems*. Ed. by M. C. Mozer, M. Jordan, and T. Petsche. Vol. 9. MIT Press, 1997, pp. 155–161.

[EH08] H. Edelsbrunner and J. Harer. "Persistent Homology - a Survey." In: *Surveys on Discrete and Computational Geometry*. Vol. 453. 2008, pp. 257–282.

[Fer21] A. Fermanian. "Embedding and Learning with Signatures." en. In: *Comput. Statist. Data Anal.* 157 (May 2021), p. 107148.

[FG10] A. Figalli and N. Gigli. "A New Transportation Distance between Non-Negative Measures, with Applications to Gradients Flows with Dirichlet Boundary Conditions." en. In: *J. Math. Pures Appl.* 94.2 (2010), pp. 107–130.

[FH20] P. K. Friz and M. Hairer. *A Course on Rough Paths: With an Introduction to Regularity Structures*. en. Second. Universitext. Springer International Publishing, 2020.

[FV10] P. K. Friz and N. B. Victoir. *Multidimensional Stochastic Processes as Rough Paths: Theory and Applications*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2010.

[FLS11] K. Fukumizu, G. Lanckriet, and B. K. Sriperumbudur. "Learning in Hilbert vs. Banach Spaces: A Measure Embedding Viewpoint." In: *Advances in Neural Information Processing Systems*. Ed. by J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, and K. Q. Weinberger. Vol. 24. Curran Associates, Inc., 2011, pp. 1773–1781.

[Ghr08] R. Ghrist. "Barcodes: The Persistent Topology of Data." en-US. In: *Bull. Amer. Math. Soc.* 45.1 (2008), pp. 61–75.

[Gil71] R. Giles. "A Generalization of the Strict Topology." In: *Trans. Amer. Math. Soc.* 161 (1971), pp. 467–474.

[GTZ10] G. Ginot, T. Tradler, and M. Zeinalian. "A Chen Model for Mapping Spaces and the Surface Product." en. In: *Ann. Sci. Éc. Norm. Supér.* 43.5 (2010), pp. 811–881.

[GL20] C. Giusti and D. Lee. "Iterated Integrals and Population Time Series Analysis." en. In: *Topological Data Analysis*. Abel Symposia. Cham: Springer International Publishing, 2020, pp. 219–246.

[GJ12] P. G. Goerss and J. F. Jardine. *Simplicial Homotopy Theory*. en. Birkhäuser, Dec. 2012.

[Gre+12] A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Schölkopf, and A. Smola. "A Kernel Two-Sample Test." In: *J. Mach. Learn. Res.* 13 (Mar. 2012), pp. 723–773.

[Gre+09] A. Gretton, K. Fukumizu, Z. Harchaoui, and B. K. Sriperumbudur. "A Fast, Consistent Kernel Two-Sample Test." In: *Advances in Neural Information Processing Systems*. Ed. by Y. Bengio, D. Schuurmans, J. D. Lafferty, C. K. I. Williams, and A. Culotta. Curran Associates, Inc., 2009, pp. 673–681.

[Gub04] M Gubinelli. "Controlling Rough Paths." en. In: *J. Funct. Anal.* 216.1 (2004), pp. 86–140.

[Gyu+13] L. G. Gyurkó, T. Lyons, M. Kontkowski, and J. Field. "Extracting Information from the Signature of a Financial Data Stream." In: *arXiv:1307.7244 [q-fin]* (July 2013). arXiv: 1307.7244 [q-fin].

[HL10] B. Hambly and T. Lyons. "Uniqueness for the Signature of a Path of Bounded Variation and the Reduced Path Group." In: *Ann. of Math.* 171.1 (2010), pp. 109–167.

[HG17] G. Henselman and R. Ghrist. "Matroid Filtrations and Computational Persistent Homology." In: *arXiv:1606.00199 [math]* (Oct. 2017). arXiv: 1606.00199 [math].

[Hua+17] Z. Huang, C. Wan, T. Probst, and L. Van Gool. "Deep Learning on Lie Groups for Skeleton-Based Action Recognition." In: *arXiv:1612.05877 [cs]* (Apr. 2017). arXiv: 1612.05877 [cs].

[Jak12] W. Jakob. "Numerically Stable Sampling of the von Mises Fisher Distribution on S$\hat{2}$ (and Other Tricks)." In: (2012).

[Jar02] J. F. Jardine. "Cubical Homotopy Theory: A Beginning." In: *preprint* (2002).

[Kal19] S. Kališnik. "Tropical Coordinates on the Space of Persistence Barcodes." en. In: *Found. Comput. Math.* 19.1 (Feb. 2019), pp. 101–129.

[KM16] F. A. Khasawneh and E. Munch. "Chatter Detection in Turning Using Persistent Homology." en. In: *Mech. Syst. Signal Process.* 70-71 (Mar. 2016), pp. 527–541.

[Kid+19] P. Kidger, P. Bonnier, I. Perez Arribas, C. Salvi, and T. Lyons. "Deep Signature Transforms." In: *Advances in Neural Information Processing Systems*. Vol. 32. Curran Associates, Inc., 2019, pp. 3099–3109.

[KM20] W. Kim and F. Mémoli. "Spatiotemporal Persistent Homology for Dynamic Metric Spaces." en. In: *Discrete Comput. Geom.* (Feb. 2020).

[KO19] F. J. Kiraly and H. Oberhauser. "Kernels for Sequentially Ordered Data." In: *J. Mach. Learn. Res.* 20.31 (2019), pp. 1–45.

[KPD90] S. Knerr, L. Personnaz, and G. Dreyfus. "Single-Layer Learning Revisited: A Stepwise Procedure for Building and Training a Neural Network." en. In: *Neurocomputing*. Ed. by F. F. Soulié and J. Hérault. NATO ASI Series. Berlin, Heidelberg: Springer, 1990, pp. 41–50.

[Kwi+15] R. Kwitt, S. Huber, M. Niethammer, W. Lin, and U. Bauer. "Statistical Topological Data Analysis - a Kernel Perspective." In: *Advances in Neural Information Processing Systems*. Ed. by C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett. Curran Associates, Inc., 2015, pp. 3070–3078.

[LD17] E. Le Donne. "A Primer on Carnot Groups: Homogenous Groups, Carnot-Carathéodory Spaces, and Regularity of Their Isometries." In: *Anal. Geom. Metr. Spaces* 5.1 (2017), p. 116.

[LG20] D. Lee and R. Ghrist. "Path Signatures on Lie Groups." en. In: *arXiv:2007.06633 [cs, math, stat]* (2020). arXiv: 2007.06633 [cs, math, stat].

[Lee03] J. M. Lee. *Introduction to Smooth Manifolds*. en. Graduate Texts in Mathematics. New York: Springer-Verlag, 2003.

[Li+19] Y. Li, T. Guo, X. Liu, and R. Xia. "Skeleton-Based Action Recognition with Lie Group and Deep Neural Networks." In: *2019 IEEE 4th International Conference on Signal and Image Processing (ICSIP)*. July 2019, pp. 26–30.

[LO14] C. Litterer and H. Oberhauser. "On a Chen–Fliess Approximation for Diffusion Functionals." en. In: *Monatsh. Math.* 175.4 (Dec. 2014), pp. 577–593.

[Lyo98] T. Lyons. "Differential Equations Driven by Rough Signals." eng. In: *Rev. Mat. Iberoam.* 14.2 (1998), pp. 215–310.

[Lyo14] T. Lyons. "Rough Paths, Signatures and the Modelling of Functions on Streams." In: *arXiv:1405.4537 [math, q-fin, stat]* (2014). arXiv: 1405.4537 [math, q-fin, stat].

[LCL07] T. Lyons, M. Caruana, and T. Lévy. *Differential Equations Driven by Rough Paths*. en. Éc. Été Probab. St.-Flour. Berlin Heidelberg: Springer-Verlag, 2007.

[LNO14] T. Lyons, H. Ni, and H. Oberhauser. "A Feature Set for Streams and an Application to High-Frequency Financial Tick Data." In: *BigDataScience '14*. 2014.

[LQ07] T. Lyons and Z. Qian. *System Control and Rough Paths*. English. Oxford: Clarendon, 2007.

[LX17] T. Lyons and W. Xu. "Hyperbolic Development and Inversion of Signature." en. In: *J. Funct. Anal.* 272.7 (Apr. 2017), pp. 2933–2955.

[May92] J. P. May. *Simplicial Objects in Algebraic Topology*. en. University of Chicago Press, 1992.

[MMH11] Y. Mileyko, S. Mukherjee, and J. Harer. "Probability Measures on the Space of Persistence Diagrams." en. In: *Inverse Problems* 27.12 (2011), p. 124007.

[Moo+19] P. Moore, T. Lyons, J. Gallacher, and Alzheimer's Disease Neuroimaging Initiative. "Using Path Signatures to Predict a Diagnosis of Alzheimer's Disease." eng. In: *PloS One* 14.9 (2019), e0222212.

[Mor+21] J. Morrill, A. Fermanian, P. Kidger, and T. Lyons. "A Generalised Signature Method for Multivariate Time Series Feature Extraction." In: *arXiv:2006.00873 [cs, stat]* (Feb. 2021). arXiv: 2006.00873 [cs, stat].

[Mun13] E. Munch. "Applications of Persistent Homology to Time Varying Systems." PhD Thesis. Duke University, 2013.

[Mun+15] E. Munch, K. Turner, P. Bendich, S. Mukherjee, J. Mattingly, and J. Harer. "Probabilistic Fréchet Means for Time Varying Persistence Diagrams." EN. In: *Electron. J. Stat.* 9.1 (2015), pp. 1173–1204.

[NJG12] N. H. P. Nguyen, E. Jankowski, and S. C. Glotzer. "Thermal and Athermal Three-Dimensional Swarms of Self-Propelled Particles." In: *Phys. Rev. E* 86.1 (July 2012), p. 011136.

[OO19] S. Ostrovska and M. Ostrovskii. "Generalized Transportation Cost Spaces." In: *Mediterr. J. Math.* 16.6 (Dec. 2019), p. 157. arXiv: 1902.10334.

[Ott+17] N. Otter, M. A. Porter, U. Tillmann, P. Grindrod, and H. A. Harrington. "A Roadmap for the Computation of Persistent Homology." en. In: *EPJ Data Sci.* 6.1 (Dec. 2017), pp. 1–38.

[PT03] F. Patras and J.-C. Thomas. "Cochain Algebras of Mapping Spaces and Finite Group Actions." In: *Topology Appl.* 128.2 (2003), pp. 189–207.

[Per+15] J. A. Perea, A. Deckard, S. B. Haase, and J. Harer. "SW1PerS: Sliding Windows and 1-Persistence Scoring; Discovering Periodicity in Gene Expression Time Series Data." In: *BMC Bioinform,* 16.1 (Aug. 2015), p. 257.

[PH15] J. A. Perea and J. Harer. "Sliding Windows and Persistence: An Application of Topological Methods to Signal Analysis." en. In: *Found. Comput. Math.* 15.3 (June 2015), pp. 799–838.

[PA+18] I. Perez Arribas, G. M. Goodwin, J. R. Geddes, T. Lyons, and K. E. A. Saunders. "A Signature-Based Machine Learning Model for Distinguishing Bipolar Disorder and Borderline Personality Disorder." eng. In: *Transl. Psychiatry* 8.1 (Dec. 2018), p. 274.

[PSS19] M. Pfeffer, A. Seigal, and B. Sturmfels. "Learning Paths from Signature Tensors." In: *SIAM J. Matrix Anal. Appl.* 40.2 (2019), pp. 394–416.

[PRT19] B. Piccoli, F. Rossi, and M. Tournus. "A Wasserstein Norm for Signed Measures, with Application to Nonlocal Transport Equation with Source Term." In: *arXiv:1910.05105 [math]* (Oct. 2019). arXiv: 1910.05105 [math].

[PWB19] P. Probst, M. N. Wright, and A.-L. Boulesteix. "Hyperparameters and Tuning Strategies for Random Forest." en. In: *Data Min. Knowl. Discov.* 9.3 (2019), e1301.

[Rei+15] J. Reininghaus, S. Huber, U. Bauer, and R. Kwitt. "A Stable Multi-Scale Kernel for Topological Machine Learning." In: *2015 IEEE Conference on Computer Vision and Pattern Recognition*. June 2015, pp. 4741–4748.

[RWF18] M. Rhif, H. Wannous, and I. R. Farah. "Action Recognition from 3D Skeleton Sequences Using Deep Networks on Lie Group Features." In: *2018 24th International Conference on Pattern Recognition (ICPR)*. Aug. 2018, pp. 3427–3432.

[Rie+20] B. Rieck, T. Yates, C. Bock, K. Borgwardt, G. Wolf, N. Turk-Browne, and S. Krishnaswamy. "Uncovering the Topology of Time-Varying fMRI Data Using Cubical Persistence." In: *Advances in Neural Information Processing Systems*. Ed. by H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin. Vol. 33. Curran Associates, Inc., 2020, pp. 6900–6912.

[Sch19a] K. Schlegel. "When Is There a Representer Theorem? Reflexive Banach Spaces." In: *arXiv:1809.10284 [cs, math, stat]* (May 2019). arXiv: 1809.10284 [cs, math, stat].

[Sch19b] K. Schlegel. "When Is There a Representer Theorem?" en. In: *J. Global Optim.* 74.2 (June 2019), pp. 401–415.

[Sch20] K. Schlegel. "Approximate Representer Theorems in Non-Reflexive Banach Spaces." en. In: *Proceedings of the 31st International Conference on Algorithmic Learning Theory*. PMLR, Jan. 2020, pp. 827–844.

[SS02] B. Schölkopf and A. Smola. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Adaptive Computation and Machine Learning. MIT Press, 2002.

[SHS01] B. Schölkopf, R. Herbrich, and A. J. Smola. "A Generalized Representer Theorem." en. In: *Computational Learning Theory*. Ed. by D. Helmbold and B. Williamson. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2001, pp. 416–426.

[Sel04] J. Selig. "Lie Groups and Lie Algebras in Robotics." en. In: *Computational Noncommutative Algebra and Applications*. Ed. by J. Byrnes. NATO Science Series II: Mathematics, Physics and Chemistry. Dordrecht: Springer Netherlands, 2004, pp. 101–125.

[Sha+20] S. Shahsavarani, I. T. Abraham, B. J. Zimmerman, Y. M. Baryshnikov, and F. T. Husain. "Comparing Cyclicity Analysis With Pre-Established Functional Connectivity Methods to Identify Individuals and Subject Groups Using Resting State fMRI." English. In: *Front. Comput. Neurosci.* 13 (2020).

[SS18] C.-J. Simon-Gabriel and B. Schölkopf. "Kernel Distribution Embeddings: Universal Kernels, Characteristic Kernels and Kernel Metrics on Distributions." In: *J. Mach. Learn. Res.* 19.44 (2018), pp. 1–29.

[ST20] E. Stepanov and D. Trevisan. "Towards Geometric Integration of Rough Differential Forms." In: *arXiv:2001.06446 [math]* (Jan. 2020). arXiv: 2001.06446 [math].

[Stro0] S. H. Strogatz. "From Kuramoto to Crawford: Exploring the Onset of Synchronization in Populations of Coupled Oscillators." en. In: *Phys. D* 143.1 (Sept. 2000), pp. 1–20.

[TZH15] C. M. Topaz, L. Ziegelmeier, and T. Halverson. "Topological Data Analysis of Biological Aggregation Models." en. In: *PLoS One* 10.5 (2015), e0126383.

[Tow02] N. Towghi. "Multidimensional Extension of L. C. Young's Inequality." In: *J. Ineq. Pure Appl. Math.* 3.2 (2002).

[TP18] C. J. Tralie and J. A. Perea. "(Quasi)Periodicity Quantification in Video Data, Using Topology." In: *SIAM J. Imaging Sci.* 11.2 (Jan. 2018), pp. 1049–1077.

[TS20] K. Turner and G. Spreemann. "Same but Different: Distance Correlations between Topological Summaries." en. In: *Topological Data Analysis*. Ed. by N. A. Baas, G. E. Carlsson, G. Quick, M. Szymik, and M. Thaule. Abel Symposia. Cham: Springer International Publishing, 2020, pp. 459–490.

[VAC14] R. Vemulapalli, F. Arrate, and R. Chellappa. "Human Action Recognition by Representing 3D Skeletons as Points in a Lie Group." In: *2014 IEEE Conference on Computer Vision and Pattern Recognition*. June 2014, pp. 588–595.

[VC16] R. Vemulapalli and R. Chellappa. "Rolling Rotations for Recognizing Human Actions from 3D Skeletal Data." In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2016, pp. 4471–4479.

[Vil09] C. Villani. *Optimal Transport: Old and New*. en. Grundlehren Der Mathematischen Wissenschaften. Berlin Heidelberg: Springer-Verlag, 2009.

[Wag21] A. Wagner. "Nonembeddability of Persistence Diagrams with p >2 Wasserstein Metric." en. In: *Proceedings of the American Mathematical Society* (2021).

[Wea18] N. Weaver. *Lipschitz Algebras*. Second. World Scientific, 2018. eprint: https://www.worldscientific.com/doi/pdf/10.1142/9911.

[Xia+20] L. Xian, H. Adams, C. M. Topaz, and L. Ziegelmeier. "Capturing Dynamics of Time-Varying Data via Topology." In: *arXiv:2010.05780 [cs, math, stat]* (Oct. 2020). arXiv: 2010.05780 [cs, math, stat].

[YJL16] W. Yang, L. Jin, and M. Liu. "DeepWriterID: An End-to-End Online Text-Independent Writer Identification System." In: *IEEE Intelligent Systems* 31.2 (Mar. 2016), pp. 45–53.

[Yan+19] W. Yang, T. Lyons, H. Ni, C. Schmid, and L. Jin. "Developing the Path Signature Methodology and Its Application to Landmark-Based Human Action Recognition." In: *arXiv:1707.03993 [cs]* (Dec. 2019). arXiv: 1707.03993 [cs].

[Yoo+16] J. Yoo, E. Y. Kim, Y. M. Ahn, and J. C. Ye. "Topological Persistence Vineyard for Dynamic Functional Brain Connectivity during Resting and Gaming Stages." en. In: *J. Neurosci. Methods* 267 (July 2016), pp. 1–13.

[You36] L. C. Young. "An Inequality of the Hölder Type, Connected with Stieltjes Integration." In: *Acta Mathematica* 67 (Jan. 1936), pp. 251–282.

[ZXZ09] H. Zhang, Y. Xu, and J. Zhang. "Reproducing Kernel Banach Spaces for Machine Learning." In: *J. Mach. Learn. Res.* 10.95 (2009), pp. 2741–2775.

[Zim+18] B. J. Zimmerman, I. Abraham, S. A. Schmidt, Y. Baryshnikov, and F. T. Husain. "Dissociating Tinnitus Patients from Healthy Controls Using Resting-State Cyclicity Analysis and Clustering." In: *Network Neurosci.* (Apr. 2018), pp. 1–23.

[ZC05] A. Zomorodian and G. Carlsson. "Computing Persistent Homology." In: *Discrete Comput. Geom.* 33.2 (Feb. 2005), pp. 249–274.

[Züs11] R. Züst. "Integration of Hölder Forms and Currents in Snowflake Spaces." en. In: *Calc. Var. Partial Differential Equations* 40.1 (Jan. 2011), pp. 99–124.