arXiv:1406.7833v4 [math.CA] 20 Jun 2016

Inverting the signature of a path

Terry J. Lyons
University of Oxford

Weijun Xu
University of Warwick

June 21, 2016

Abstract

The aim of this article is to develop an explicit procedure that enables one to reconstruct any $\mathcal{C}^1$ path (at natural parametrization) from its signature. We also explicitly quantify the distance between the reconstructed path and the original path in terms of the number of terms in the signature that are used for the construction and the modulus of continuity of the derivative of the path. A key ingredient in the construction is the use of a procedure of symmetrization that separates the behavior of the path at small and large scales.

Key words: signature, inversion, symmetrization.

AMS Classification: 70G, 93A.

1 Introduction

1.1 The signature of a path

Given an integer $d \geq 1$, a d-dimensional path γ is a continuous function mapping a closed interval $[0, T]$ into $\mathbb{R}^d$. Throughout the article, we equip $\mathbb{R}^d$ with the ℓ^1 norm. The length of γ under this norm is then

$$\|\gamma\| := \sup_{\mathcal{P}} \sum_j |\gamma_{u_j} - \gamma_{u_{j-1}}|,$$

where the supremum is taken over all partitions of the interval $[0, T]$, and $|\mathbf{x}| := \sum_{i=1}^d |x_i|$ for $\mathbf{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d$. We say γ is a path of *finite length* if $\|\gamma\| < +\infty$.

Given two paths $\alpha : [0, S] \to \mathbb{R}^d$ and $\beta : [0, T] \to \mathbb{R}^d$, their *concatenation* $\alpha * \beta$ is a new path defined on the interval $[0, S + T]$ by

$$\alpha * \beta(u) := \begin{cases} \alpha(u), u \in [0, S] \\ \beta(u - S) + \alpha(S) - \beta(0), u \in [S, S + T] \end{cases}. \tag{1.1}$$

If γ has finite length, then its derivative $\dot{\gamma}$ exists for almost every $t \in [0, T]$. We can re-parametrize γ in the fixed time interval $[0, 1]$ in such a way that

$$|\dot{\gamma}_t| = \sum_{i=1}^d |\dot{\gamma}_t^i| \equiv L,$$

where $\dot{\gamma}^i$'s are the components of $\dot{\gamma}$, and L is the length of γ under ℓ^1 norm. We call such a parametrization the *natural parametrization* of γ.

Remark 1.1. The notion of natural parametrization used here is slightly different from the standard one in the literature, as we parametrize γ in the unit interval rather than $[0, L]$. As a consequence, the constant speed is L instead of 1. The reason of using this parametrization is that later on we will compare paths with different lengths, so it will be convenient for us to parametrize all of them in the same interval.

If γ has finite length, then one can define a sequence of iterated integrals of it, called the *signature* of γ. We first introduce the notion of words before giving a precise definition of signature.

Let $\{e_1, \dots, e_d\}$ denote the standard basis of $\mathbb{R}^d$. For every integer $n \geq 0$, a word of length n is an ordered sequence of n letters from the set $\{e_1, \dots, e_d\}$ (with repetition allowed), and we use $|w|$ to denote the length of w. For two words $w_1 = e_{i_1} \cdots e_{i_n}$ and $w_2 = e_{j_1} \dots e_{j_m}$, their *concatenation* $w_1 * w_2$ is a word of length $n + m$ given by

$$w_1 * w_2 = e_{i_1} \dots e_{i_n} e_{j_1} \dots e_{j_m}.$$

We also use $\emptyset$ to denote the empty word, which is the unique word of length 0. The signature of a finite length path can now be defined as follows.

Definition 1.2. *Let $\gamma : [0, T] \to \mathbb{R}^d$ be a path of finite length. For every integer n and every word $w = e_{i_1} \cdots e_{i_n}$, define*

$$C_\gamma(w) := \int_{0<u_1<\cdots<u_n<T} d\gamma_{u_1}^{i_1} \cdots d\gamma_{u_n}^{i_n},$$

where γ^i is the component of γ in the direction e_i. The signature of γ is the formal power series

$$X(\gamma) = \sum_{n=0}^{+\infty} \sum_{|w|=n} C_\gamma(w) w,$$

where we have set $C_\gamma(\emptyset) = 1$.

The signature of γ is a definite integral over a fixed time interval where γ is defined. Re-parametrizing γ does not change its signature. One reason to look at the signatures is that they contain important information about the paths. For example, the first level coefficients $\{C_\gamma(w) : |w| = 1\}$ reproduce the increment of the path, and the second level collection $\{C_\gamma(w) : |w| = 2\}$ represents the area enclosed by the projections of the path on the $e_i - e_j$ planes.

The study of the signature dates back to K.T.Chen in 1950's. In a series of papers ([Che54], [Che57], [Che58]), he showed that the map $\gamma \mapsto X(\gamma)$ is a homomorphism from the monoid of paths with concatenation to the tensor algebra over $\mathbb{R}^d$, and proved that two piecewise regular paths without backtracks have the same signatures if and only if they differ by a re-parametrization and translation. This uniqueness result, modulo tree-like equivalence, was extended to all finite length paths (in [HL10]) and finally to all geometric rough paths[1] (in [BGLY16]). However, all the proofs there are non-constructive, and a natural question that arises after these works is that how one can reconstruct a path from its signature. In this article, we consider the inversion problem when γ is $\mathcal{C}^1$ at natural parametrization, which automatically implies that γ has finite length and does not have any backtracks in its trajectory.

[1]See [BGLY16] or [Lyo98] for the definition of the signature of a rough path.

Remark 1.3. In the case $d = 1$, all back-forth movements of the path are cancelled out and only the total increment counts. The signature X takes the form $\exp(ae_1)$, where $a = \gamma_T - \gamma_0$, and the reduced path with signature X is simply the straight line going from 0 to a. The inversion problem is then interesting only when $d \geq 2$, where evolutions of the path are in general non-commutative.

The key to the reconstruction problem in $d \geq 2$ is then to recover these non-commutative evolutions of the path in their correct orders based on the signature. To appreciate the difficulty of the problem, we note from Definition 1.2 that each term $C_\gamma(w)$ represents some global effect of the path γ, and a priori there is no obvious way of recovering local information of the path from these global representations.

A naive approach to the reconstruction problem can be to try to reproduce the path from the interpretation of each term in the signature, as the explanation of the meanings of $\{C_\gamma(w) : |w| \leq 2\}$ above. However, these intuitive interpretations break down when $|w|$ gets large, and it is essentially impossible to proceed this way to recover much finer information of the path beyond the increment and area. Thus, certain operations on signatures that can reveal local information of the path will be necessary for the reconstruction.

There have been recent attempts to the reconstruction. In establishing the uniqueness of signature for Brownian motion sample paths, Le Jan and Qian ([LJQ13]) constructed polygonal approximations to the Brownian paths by using the information in their signatures only. This approximation scheme has been further extended to diffusions ([GQ16]) and a large class of Gaussian processes ([BG15]). These works all used a fixed approximation scheme that relies on the law of the underlying random paths which gives certain almost sure non-degeneracy property of the sample paths. However, this non-degeneracy property is not available for all deterministic rough paths, so one would not expect any fixed approximation scheme to work in the deterministic setting without further assumptions on the paths. This obstacle was overcome by Geng in [Gen15], where he allowed flexibility in the scheme used depending on the signature of the path in consideration, and gave a construction of any deterministic geometric rough path from its signature.

However, all the approximations schemes used in those works are indirect – they require a sophisticated transform of the signature via smooth one-forms which involves various limiting processes. In addition, even for approximation at any fixed scale and in the setting of smooth paths, these constructions require the use of the *whole signature sequence* rather than any truncation of it. Thus, it is hard to turn them into an effective algorithm.

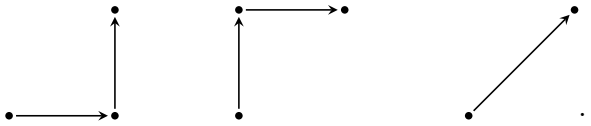
In the recent article [LX15], by using a construction from hyperbolic geometry, the authors gave an explicit inversion scheme together with stability properties for piecewise linear paths. But this scheme makes an essential use of the piecewise linearity of the path, and it is not clear how it can be extended to more general situations.

1.2 Main result and strategy

The main goal of this article is to develop an explicit and implementable procedure that enables one to reconstruct any $\mathcal{C}^1$ path (at natural parametrization) from its signature, together with detailed stability estimates of the procedure.

As mentioned in the previous subsection (Remark 1.3), the key to establish an effective reconstruction algorithm is to recover the non-commutative evolution of

the path in its correct order. For example, the following three paths in $\mathbb{R}^2$ have the same increments. Their signatures agree on level $|w| = 1$, but start to differ when $|w| \geq 2$:



Let x and y denote the standard basis of $\mathbb{R}^2$, then the signatures of the above three paths are the formal series $e^x e^y$, $e^y e^x$ and e^{x+y}, respectively.

As one can see, the order of the evolution of the paths are captured in the signatures through the ordered letters that consist the words w's (see Definition 1.2 above). Our main result is that, by symmetrizing the signatures at high level ($|w|$ large), we can average out the non-commutativities of the path at small scales but still preserve the order of its evolution at larger scales. Thus, for any given k, we produce from the *truncated signature* a piecewise linear path that approximates the original path at scale $\frac{1}{k}$, and the error of the approximation can be explicitly quantified in terms of the modulus of continuity of the derivative of the path. The main theorem could be loosely stated as follows.

Theorem 1.4. *Let γ be a $\mathcal{C}^1$ path in $\mathbb{R}^d$ at natural parametrization, and assume we know its signature $X = \{C_\gamma(w) : |w| \geq 0\}$. For large enough k (depending on the dimension and the path, but will be quantified explicitly below), by using the terms in the truncated signature up to level $|w| = 2dk^3 \log k + k$, we construct a piecewise linear path $\tilde{\gamma}$ with k linear pieces such that when both γ and $\tilde{\gamma}$ are at natural parametrization with respect to ℓ^1 norm, for every $t \in [0,1]$ where $\dot{\tilde{\gamma}}_t$ is defined, we have*

$$|\dot{\gamma}_t - \dot{\tilde{\gamma}}_t| < C\eta_k, \tag{1.2}$$

where $C > 0$ depends on the path γ only, and $\eta_k \to 0$ as $k \to +\infty$ with rate depending on the modulus of continuity of $\dot{\gamma}$.

Remark 1.5. A precise statement of this main result, including the value of the constant C, can be found in Theorem 4.17 for $d = 2$ and in Theorem 5.1 for higher dimensions. The definition of η_k is in (1.4) for $d = 2$ and in (5.2) for general d. As for the decay of η_k, if $\gamma \in \mathcal{C}^{1,\alpha}$ for some $\alpha \in (0,1]$, then $\eta_k = \mathcal{O}(k^{-\frac{\alpha^2}{2}})$ (see (1.4) and Remark 4.1 for more details about the rate). But we expect that both this error bound and the level of truncation $|w| \sim k^3 \log k$ in the signature can be improved.

Remark 1.6. Although many arguments below made an explicit use of the continuity of $\dot{\gamma}$, we expect most of them still hold (with a different proportionality constant) if $\dot{\gamma}$ is piecewise continuous, and its discontinuities have jumps less than π. The only statement that essentially relies on the continuity of $\dot{\gamma}$ is the existence of the word w^* with certain non-degeneracy properties (in Theorem 3.4). This word is used in the symmetrization procedure (see Section 2) mainly for technical convenience – one would not have the closed form expression (2.4) for symmetrized signatures without the use of the word to separate blocks in symmetrization. It will be very interesting, and also convenient for practice, if one could prove the validity and stability of the symmetrization procedure without using this word to separate blocks. This would give an inversion scheme for piecewise $\mathcal{C}^1$ paths and possibly all finite length paths.

On the other hand, however, our procedure does make an essential use of the finite length nature of γ. Any extension of the procedure to rough path together with detailed quantitative characterization of its stability would require new ingredients.

We now briefly explain the main steps in the construction of such a piecewise linear path. Since there is no essential difference between dimension 2 and higher, we will mainly focus on the 2-dimensional case, and give a description of the procedure in higher dimensions with a brief explanation in Section 5.

Each line segment in a piecewise linear path is determined by its direction (a unit vector in $\mathbb{R}^2$) and length. Thus, for each integer k and each $j = 1, \dots, k$, our aim is to find a 2-dimensional unit vector (in ℓ^1 norm)

$$\theta_j = \big(a_j^x \rho_j, a_j^y (1 - \rho_j)\big),$$

where $\rho_j \in [0, 1]$ and $a_j^x, a_j^y \in \{\pm 1\}$, and $\tilde{L} \in \mathbb{R}^+$ such that

$$\sup_{1 \le j \le k} \sup_{u \in [\frac{j-1}{k}, \frac{j}{k}]} |\dot{\gamma}_u - \tilde{L}\theta_j| < C\eta_k.$$

Thus, if we let $\tilde{\gamma}$ to be the path that are concatenations of the line segments $\frac{\tilde{L}}{k}\theta_j$'s, this will automatically imply that $\tilde{\gamma}$ is close to the original path γ in the Lipschitz norm.

It is clear that the parameter $\rho_j \in [0, 1]$ represents the *unsigned direction* of the j-th line segment, a_j^x, a_j^y represent the signs of x and y directions (in $\mathbb{R}^2$) in that segment, and $\tilde{L}$ is an approximation to the ℓ^1 length of the original path.

In order to get a rough idea how these parameters can be obtained from the signature of γ, we first briefly recall from [LX15] the inversion scheme for integer lattice paths. This can be decomposed into two steps:

1. Identify the unique longest square free word[2] w such that $C(w) \neq 0$. The order of the letters in w gives the directions (up to the sign) of each piece of the lattice path.
2. Move one level up in the signature to recover the sign as well as the length of each piece.

The reconstruction scheme developed in this article is close in spirit to that for the lattice paths. In what follows, we will recover these parameters in the order of ρ_j, a_j^x, a_j^y and then finally $\tilde{L}$. Among these parameters, the most difficult one to recover is the unsigned direction ρ_j. Ideally, one would like to implement the strategy for integer lattice paths to recover these ρ_j's. However, at first glance, that procedure seems to crucially depend on the very special structure of the integer lattice, and does not generalize directly to other situations. In particular, the vanishing/non-vanishing property of coefficients of square free words does not carry over to more general cases where the path can move along any direction in the plane. In order to recover the directions that are not necessarily parallel to the Euclidean axes, we use a symmetrization procedure together with a more robust notion of non-degeneracy that replaces the strict non-zero criterion in Step 1 above.

Once the directions ρ_j's are recovered, one can (similar to the lattice path case) move one level up in the signature to recover the signs a_j^x and a_j^y. Finally, the approximated length $\tilde{L}$ is obtained by a simple scaling argument.

1.3 Assumptions and notations

We now summarize the assumptions on our path γ as well as the notations we will be using in the article. We will mainly work in dimension $d = 2$. We equip $\mathbb{R}^2$ with the ℓ^1 norm, and let x and y denote the standard basis elements of $\mathbb{R}^2$.

[2] A word $w = e_{i_1} \cdots e_{i_n}$ is square free if for all $j = 1, \cdots, n-1$, we have $i_j \neq i_{j+1}$.

The notation $|\cdot|$ will have different meanings in various contexts. If $\alpha \in \mathbb{R}$, then $|\alpha|$ is the *absolute value* of α. If $\alpha = (\alpha_1, \alpha_2) \in \mathbb{R}^2$ is a vector, then $|\alpha| = |\alpha_1| + |\alpha_2|$ is its ℓ^1 norm. Finally, if w is a word, then $|w|$ denote the length of w.

Throughout, we fix our path γ that is $\mathcal{C}^1$ under *natural parametrization* and has length L. More precisely, the path

$$\gamma : [0,1] \to \mathbb{R}^2, \qquad \gamma_u = (x_u, y_u)$$

has continuous derivative $\dot{\gamma}_u = (\dot{x}_u, \dot{y}_u)$ on $[0,1]$, and satisfies $|\dot{x}_u| + |\dot{y}_u| \equiv L$ for all $u \in [0,1]$. We let δ denote the modulus of continuity of $\dot{\gamma}$, so

$$\delta(\epsilon) := \sup_{|s-t|<\epsilon} |\dot{\gamma}_s - \dot{\gamma}_t|. \tag{1.3}$$

For each integer k, we define ϵ_k and η_k to be

$$\epsilon_k := \sqrt{2}\left(\sqrt{\frac{\delta(1/k)}{L}} + \frac{1}{\sqrt{k}}\right), \qquad \eta_k := \delta(3\epsilon_k) + \frac{L}{\sqrt{k}}. \tag{1.4}$$

The definitions (1.4) are for the two dimensional case; the case for general dimension d are introduced in (5.2). For any integer k, let Δ_{k-1} be the standard simplex

$$\Delta_{k-1} := \big\{(u_1, \dots, u_{k-1}) : 0 = u_0 < u_1 < \cdots < u_{k-1} < u_k = 1\big\}. \tag{1.5}$$

In situations where there might be a confusion, We use boldface letters to denote vectors and normal letters for their components; for example we write

$$\mathbf{u} = (u_1, \dots, u_{k-1}).$$

On the other hand, we write the path as $\gamma = (\gamma^1, \gamma^2)$ as there could be no confusion to arise in this case.

Given the path γ and $\mathbf{u} \in \Delta_{k-1}$, for every $j = 1, \dots, k$, we let

$$\Delta_{\mathbf{u},j} x := x_{u_j} - x_{u_{j-1}}, \qquad \Delta_{\mathbf{u},j} y := y_{u_j} - y_{u_{j-1}} \tag{1.6}$$

denote the increments in the relevant directions in the time interval $[u_{j-1}, u_j]$, and

$$|\Delta_{\mathbf{u},j}\gamma| := |\Delta_{\mathbf{u},j} x| + |\Delta_{\mathbf{u},j} y| \tag{1.7}$$

be the magnitude of total increments. Similarly, we denote the increments of the j-th piece under standard subdivision $\mathbf{s}_k = \big\{\frac{j}{k}\big\}_{j=0}^{k}$ by

$$\Delta_j x = x_{j/k} - x_{(j-1)/k}, \qquad \Delta_j y = y_{j/k} - y_{(j-1)/k}, \tag{1.8}$$

and let $|\Delta_j \gamma| = |\Delta_j x| + |\Delta_j y|$.

1.4 Organization of the article

This article is organized as follows. In Section 2, we introduce and set up the symmetrization procedure on signatures. Section 3 is devoted to the proof of a concentration property of the symmetrized signatures with explicit quantitative estimates. These estimates will then be used in Section 4 where we give a detailed description of the reconstruction procedure of the path from the signatures. Finally, in Section 5, we give an outline of the reconstruction procedure for paths in higher dimensions.

Acknowledgements

We thank two referees for carefully reading the manuscript and providing helpful suggestions. Weijun Xu also thanks Horatio Boedihardjo for many helpful discussions.

The research of Terry Lyons is supported by EPSRC grant EP/H000100/1 and the European Research Council under the European Unions Seventh Framework Program (FP7-IDEAS-ERC) / ERC grant agreement nr. 291244. Terry Lyons acknowledges the support of the Oxford-Man Institute. Weijun Xu has been supported by the Oxford-Man Institute through a scholarship during his time as a student at Oxford. He is now supported by Leverhulme trust.

2 Symmetrization

As mentioned in the introduction, the first step in the reconstruction is to obtain the unsigned direction $\rho_j \in [0,1]$ of the j-th line segment in our piecewise linear path approximation, and the key to the accurate recovery of this parameter is the use of symmetrization procedure together with a robust notion of non-degeneracy. We first give a simple example to illustrate this.

Example 2.1. Let X be the signature of some bounded variation path $\gamma_t = (x_t, y_t), t \in [0,1]$, and we would like to recover from X the increments

$$\Delta x := x_1 - x_0 \qquad \text{and} \qquad \Delta y := y_1 - y_0,$$

where (x_0, y_0) and (x_1, y_1) are beginning and end points of γ. One could of course get the exact values of the pair $(\Delta x, \Delta y)$ directly from the first level signature X^1, but cannot proceed much further. The symmetrization method given below is more complicated, but has the advantage that it can be generalized to recover much more information beyond the increments.

The symmetrization procedure to recover the increments is as follows. For every integer n and every $\ell = 0, 1, \ldots, n$, let

$$\mathcal{S}^n(\ell) := n! \sum C(w),$$

where the sum is taken over all words w with length n that contain ℓ x's and $(n-\ell)$ y's. Then, $\mathcal{S}^n(\ell)$ has the expression

$$\mathcal{S}^n(\ell) = \binom{n}{\ell} (\Delta x)^\ell (\Delta y)^{n-\ell}.$$

Note that for each n and ℓ, the left hand side above is the information available to us (from the signature), and the right hand side is its expression. It is standard that for fixed large n, the quantity $|\mathcal{S}^n(\ell)|$ is maximized near the value ℓ^* such that

$$\frac{\ell^*}{n-\ell^*} \approx \frac{|\Delta x|}{|\Delta y|}.$$

We can then asymptotically recover the ratio $|\Delta x| : |\Delta y|$ by finding the maximizer ℓ^* of $\mathcal{S}^n(\ell)$, and this gives us the *unsigned direction* of the increment.

To recover the signs of Δx and Δy, one repeats the same trick as in the case for integer lattice paths: moving one level up and comparing the signs. Finally, the magnitude of the increment will be obtained via scaling.

The example above illustrates the case for the single-piece approximation to the path (that is, $k = 1$). In order to recover finer information of the path (for large k), instead of symmetrizing the whole signature, we divide high level signatures into k equal blocks, and symmetrize each block. We first introduce some notations before we set up the procedure.

For every integer k and n, we let $\mathcal{L}_k^n$ denote the set of multi-indices

$$\mathcal{L}_k^n := \Big\{ \boldsymbol{\ell} = (\ell_1, \dots, \ell_k) : 0 \le \ell_j \le n, \sum_{j=1}^k \ell_j = n \Big\}. \tag{2.1}$$

For every word w, we let $|w|_x$ and $|w|_y$ denote the number of letters x and y in w, respectively. For every w of length $k-1$ of the form $w = e_{i_1} \cdots e_{i_{k-1}}$ and every multi-index $\boldsymbol{\ell} \in \mathcal{L}_k^n$, we let $\mathcal{W}_k^{2n}(w, \boldsymbol{\ell})$ be the set of words

$$\mathcal{W}_k^{2n}(w, \boldsymbol{\ell}) = \Big\{ w' = w_1 * e_{i_1} * \cdots * e_{i_{k-1}} * w_k : \ |w_j|_x = 2\ell_j, |w_j|_y = 2n - 2\ell_j \Big\}. \tag{2.2}$$

A typical word $w' \in \mathcal{W}_k^{2n}(w, \boldsymbol{\ell})$ where $w = e_{i_1} \dots e_{i_{k-1}}$ has the form

$$\underbrace{*****}_{w_1} \ e_{i_1} \ \underbrace{*****}_{w_2} \ e_{i_2} \cdots\cdots e_{i_{k-2}} \underbrace{*****}_{w_{k-1}} \ e_{i_{k-1}} \ \underbrace{*****}_{w_k}.$$

Here, each w_j is a sub-word of length $2n$ with $2\ell_j$ letters x and $2n - 2\ell_j$ letters y. The two consecutive sub-words (blocks) w_j and w_{j+1} are separated by the letter e_{i_j} from w. For example, for $n = 2$ and $k = 1$, $\mathcal{L}_1^2 = \{0, 1, 2\}$, so we have

$$\begin{aligned} &\mathcal{W}_1^4(\emptyset, (0)) = \big\{yyyy\big\}, \qquad \mathcal{W}_1^4(\emptyset, (2)) = \big\{xxxx\big\}, \\ &\mathcal{W}_1^4(\emptyset, (1)) = \big\{xxyy, xyxy, xyyx, yxxy, yxyx, yyxx\big\}. \end{aligned}$$

For $n = k = 2$, the set $\mathcal{W}_2^4(x, (1, 0))$ consists of the words

$$\big\{xxyyxyyyy, xyxyxyyyy, xyyxxyyyy, yxxyxyyyy, yxyxxyyyy, yyxxxyyyy\big\}.$$

The set $\mathcal{W}_2^4(x, (1, 2))$ is similar except that one replaces the last four y's by four x's. With this definition, we introduce the *symmetrized signatures*

$$\mathcal{S}_k^{2n}(w, \boldsymbol{\ell}) := ((2n)!)^k \sum_{w' \in \mathcal{W}_k^{2n}(w, \boldsymbol{\ell})} C(w'). \tag{2.3}$$

Recall the definitions of Δ_{k-1}, $\Delta_{\mathbf{u},j} x$ and $\Delta_{\mathbf{u}_j} y$ in (1.5) and (1.6), we have the following proposition.

Proposition 2.2. *Fix integer k and n. Let $w = e_{i_1} \cdots e_{i_{k-1}}$ and $\boldsymbol{\ell} = \{\ell_1, \dots, \ell_k\} \in \mathcal{L}_k^n$. Then, the quantity $\mathcal{S}_k^{2n}(w, \boldsymbol{\ell})$ defined above has the expression*

$$\mathcal{S}_k^{2n}(w, \boldsymbol{\ell}) = \int_{\Delta_{k-1}} \prod_{j=1}^{k-1} \dot{\gamma}_{u_j}^{i_j} \prod_{j=1}^{k} \binom{2n}{2\ell_j} (\Delta_{\boldsymbol{u},j} x)^{2\ell_j} (\Delta_{\boldsymbol{u},j} y)^{2n-2\ell_j} d\boldsymbol{u}. \tag{2.4}$$

Proof. By the definition of $\mathcal{W}_k^{2n}(w, \boldsymbol{\ell})$ in (2.2), we have

$$\begin{aligned} &\sum_{w' \in \mathcal{W}_k^{2n}(w, \boldsymbol{\ell})} C(w') \\ &= \int_{\mathbf{u} \in \Delta_{k-1}} \prod_{j=1}^{k} \bigg(\frac{1}{(2\ell_j)!(2n - 2\ell_j)!} \int_{u_{j-1} < v_1^j, \dots, v_{2n}^j < u_j} d\gamma_{v_1^j}^{i_1^j} \cdots d\gamma_{v_{2n}^j}^{i_{2n}^j} \bigg) d\gamma_{u_1}^{i_1} \cdots d\gamma_{u_{k-1}}^{i_{k-1}}, \end{aligned}$$

where $\mathbf{u} = (u_1, \dots, u_{k-1})$, $u_0 = 0$, $u_k = 1$, and for each j, the word $w_j = e_{i_1^j} \cdots e_{i_{2n}^j}$ consists of $2\ell_j$ letters x and $2n - 2\ell_j$ letters y. Thus, we have

$$\int_{u_{j-1}<v_1^j,\dots,v_{2n}^j<u_j} d\gamma_{v_1^j}^{i_1^j} \cdots d\gamma_{v_{2n}^j}^{i_{2n}^j} = \big(\Delta_{\mathbf{u},j}x\big)^{2\ell_j}\big(\Delta_{\mathbf{u},j}y\big)^{2n-2\ell_j},$$

and (2.4) follows immediately. □

We will see below that these $\mathcal{S}_k^{2n}(w,\boldsymbol{\ell})$'s are the only quantities we will use to recover the unsigned directions of each piece in our piecewise linear approximation. It is the recovery of the sign of each direction that requires more information in the signature other than the $\mathcal{S}_k^{2n}$'s. We will introduce those additional quantities only when it becomes necessary.

Remark 2.3. We emphasize that (2.3) is the definition of the symmetrized signature; this information is available to us from the signature X. On the other hand, (2.4) is an expression of this quantity, and we will make use of this expression later to prove a priori bounds of the symmetrized signature.

Remark 2.4. The reason why we insert a letter e_{i_j} between every two consecutive symmetrized blocks is to let $\mathcal{S}_k^{2n}(w,\boldsymbol{\ell})$ have a closed form expression as in (2.4). This is mainly for technical convenience, and we expect results in the next section still hold true when the symmetrization is taken without using these e_{i_j}'s to separate blocks. Also, the symmetrization is taken only over even numbers of x's and y's in each block. This is to avoid cancellations of different signs inside the integration on the right hand side of (2.4).

Before we proceed, we first give a heuristic explanation how the quantities $\mathcal{S}_k^{2n}(w,\boldsymbol{\ell})$'s can help recover the unsigned directions of each segment of the path. For fixed word $w = e_{i_1} \cdots e_{i_{k-1}}$, using Proposition 2.2 and then summing over $\boldsymbol{\ell} \in \mathcal{L}_k^n$, we have

$$\sum_{\boldsymbol{\ell}\in\mathcal{L}_k^n} \mathcal{S}_k^{2n}(w,\boldsymbol{\ell}) = \frac{1}{2^k}\int_{\Delta_{k-1}} \prod_{j=1}^{k-1} \dot{\gamma}_{u_j}^{i_j} \prod_{j=1}^{k} \big((\Delta_{\mathbf{u},j}x + \Delta_{\mathbf{u},j}y)^{2n} + (\Delta_{\mathbf{u},j}x - \Delta_{\mathbf{u},j}y)^{2n}\big) d\mathbf{u}. \tag{2.5}$$

If $n \gg k$, it is natural to expect that the above integrand has magnitude of order $C_k \prod_{j=1}^k |\Delta_{\mathbf{u},j}\gamma|^{2n}$, where $|\Delta_{\mathbf{u},j}\gamma| = |\Delta_{\mathbf{u},j}x| + |\Delta_{\mathbf{u},j}y|$ is defined in (1.7). This suggests

$$\sum_{w:|w|=k-1} \sum_{\boldsymbol{\ell}\in\mathcal{L}_k^n} \mathcal{S}_k^{2n}(w,\boldsymbol{\ell}) \sim C_k \int_{\Delta_{k-1}} \prod_{j=1}^{k} |\Delta_{\mathbf{u},j}\gamma|^{2n} d\mathbf{u}. \tag{2.6}$$

By the concentration properties of the integral $\int_{\Delta_{k-1}} \prod_j |\Delta_{\mathbf{u},j}\gamma|^{2n} d\mathbf{u}$ (which will be proven and quantified in the next Section) as well as the normal approximation to binomials, it turns out that as long as we sum the $\mathcal{S}_k^{2n}(w,\boldsymbol{\ell})$'s over a very small range of $\boldsymbol{\ell} \in \mathcal{L}_k^n$ such that

$$\frac{\ell_j}{n-\ell_j} \sim \frac{|\Delta_j x|}{|\Delta_j y|}, \qquad j = 1, \dots, k, \tag{2.7}$$

then the sum will also be "close" to its possible maximum – the right hand side of (2.6). On the other hand, if any ℓ_j is away from the range in (2.7), then the sum of $\mathcal{S}_k^{2n}(w,\boldsymbol{\ell})$'s will be negligible compared to (2.6). Then, similar as Example 2.1, "observing" which range of ℓ_j maximizes the sum of $\mathcal{S}_k^{2n}(w,\boldsymbol{\ell})$'s will recover asymptotically the unsigned directions of each piece. We will precisely formulate and prove this heuristic in the next sections.

3 Concentration of symmetrized signatures

The aim of this section is to prove a quantitative statement about the concentration property of the integral

$$\int_{\Delta_{k-1}} \prod_{j=1}^{k} |\Delta_{\mathbf{u},j}\gamma|^n d\mathbf{u} \tag{3.1}$$

when $n \gg k$, where $|\Delta_{\mathbf{u},j}\gamma|$ is the magnitude of the increment of γ in $[u_{j-1}, u_j]$ as defined in (1.7). This concentration property roughly states that although the integration is taken over the whole simplex, when n is large, almost all its contribution comes from a very small subset of Δ_{k-1}. In fact, the domain of concentration is around the points $\mathbf{u} \in \Delta_{k-1}$ such that the product $\prod_j |\Delta_{\mathbf{u},j}\gamma|$ is maximized, and these maximizers cannot be far away from the standard dissection $\mathbf{s}_k = \{\frac{j}{k}\}_{j=1}^{k-1}$. A quantitative statement will be given in Proposition 3.3 below. As a consequence, we will obtain in Theorem 3.4 a precise quantitative characterization of the "heuristic" (2.6). Throughout, we assume $\gamma : [0,1] \to \mathbb{R}^2$ is at natural parametrization with respect to ℓ^1 norm, $\gamma \in \mathcal{C}^1$, and δ is the modulus of continuity of $\dot{\gamma}$.

Lemma 3.1. *Let k be an integer such that $\delta(\frac{1}{k}) < \frac{L}{2}$, where L is the ℓ^1 length of γ. Then, for every $j = 1, \ldots, k$, we have*

$$\frac{L - \delta(\frac{1}{k})}{k} \leq |\Delta_j\gamma| \leq \frac{L}{k},$$

where $|\Delta_j\gamma| = |\Delta_j x| + |\Delta_j y|$ is the magnitude of the increment of γ in $[\frac{j-1}{k}, \frac{j}{k}]$, as defined in (1.8).

Proof. The inequality $|\Delta_j\gamma| \leq \frac{L}{k}$ follows immediately from the assumption that γ is at natural parametrization.

For the lower bound, for each $j = 1, \ldots, k$, we let $I_j = [\frac{j-1}{k}, \frac{j}{k}]$. If both $\dot{x}_u$ and $\dot{y}_u$ keep their signs unchanged in the interval I_j, then we have

$$\left| \int_{I_j} \dot{x}_u du \right| = \int_{I_j} |\dot{x}_u| du, \qquad \left| \int_{I_j} \dot{y}_u du \right| = \int_{I_j} |\dot{y}_u| du,$$

and it follows immediately that $|\Delta_j\gamma| = \frac{L}{k}$. If not, then either $\dot{x}$ or $\dot{y}$ is 0 at some point u in the interval I_j. Suppose without loss of generality that $\dot{y}_u = 0$ for some $u \in I_j$, then we have

$$\sup_{u \in I_j} |\dot{y}_u| \leq \delta(\frac{1}{k}),$$

which in turn gives

$$|\dot{x}_u| \geq L - \delta(\frac{1}{k}) \geq \frac{L}{2}.$$

for all $u \in I_j$. In particular, the continuity of $\dot{x}_u$ implies that it does not change sign in I_j. Thus, we have

$$|\Delta_j\gamma| \geq \left| \int_{I_j} \dot{x}_u du \right| = \int_{I_j} |\dot{x}_u| du \geq \frac{L - \delta(\frac{1}{k})}{k}.$$

This finishes the proof of the lemma. □

Lemma 3.1 says that at $\mathbf{u} = \mathbf{s}_k = \{\frac{j}{k}\}$, the product $\prod_j |\Delta_j \gamma|$ is close to its largest possible value. On the other hand, if any u_j is far away from $\frac{j}{k}$, then $\prod_j |\Delta_{\mathbf{u},j}\gamma|$ must be small. This is the content of the following lemma.

Lemma 3.2. *Let k be an integer such that $\delta(\frac{1}{k}) < \frac{L}{2}$, and let $\boldsymbol{u} = \{u_1 < \cdots < u_{k-1}\} \in \Delta_{k-1}$. Recall from (1.4) that*

$$\epsilon_k = \sqrt{2}\Big(\sqrt{\frac{\delta(\frac{1}{k})}{L}} + \frac{1}{\sqrt{k}}\Big).$$

If $\big|u_j - \frac{j}{k}\big| \geq \epsilon_k$ for some $1 \leq j \leq k$, then

$$\prod_{i=1}^{k} \Big(|\Delta_{\boldsymbol{u},i}\gamma| / |\Delta_i \gamma| \Big) < \frac{1}{e}. \tag{3.2}$$

Proof. Since the left hand side of (3.2) is invariant under rescaling of length, we can assume without loss of generality that $L = 1$. Suppose

$$u_j - \frac{j}{k} = \epsilon$$

for some j and some ϵ. Then $u_j = \frac{j}{k} + \epsilon$, and the sum of all increments before and after the time $t = u_j$ satisfy

$$\sum_{i=1}^{j} |\Delta_{\mathbf{u},i}\gamma| \leq \frac{j}{k} + \epsilon, \qquad \sum_{i=j+1}^{k} |\Delta_{\mathbf{u},i}\gamma| \leq \frac{k-j}{k} - \epsilon. \tag{3.3}$$

Note that here we do not require ϵ to be positive. By the bound (3.3), the best possible maximum one can hope for $\prod_i |\Delta_{\mathbf{u},i}\gamma|$ is the case when we have

$$|\Delta_{\mathbf{u},i}\gamma| = \frac{1}{k} + \frac{\epsilon}{j}, \quad \forall i \leq j \qquad \text{and} \qquad |\Delta_{\mathbf{u},i}\gamma| = \frac{1}{k} - \frac{\epsilon}{k-j}, \quad \forall i \geq j+1,$$

which gives

$$\prod_{i=1}^{k} |\Delta_{\mathbf{u},i}\gamma| \leq \Big(\frac{1}{k} + \frac{\epsilon}{j}\Big)^{j} \Big(\frac{1}{k} - \frac{\epsilon}{k-j}\Big)^{k-j}.$$

Since $\delta(\frac{1}{k}) < \frac{L}{2}$, we can apply Lemma 3.1 to get

$$\prod_{i=1}^{k} \Big(|\Delta_{\mathbf{u},i}\gamma| / |\Delta_i \gamma| \Big) \leq \Big(\frac{(1+p\epsilon)^{\frac{1}{p}}(1-q\epsilon)^{\frac{1}{q}}}{1-\delta(\frac{1}{k})} \Big)^{k}, \tag{3.4}$$

where $p = \frac{k}{j}$ and $q = \frac{k}{k-j}$. Now, let

$$f(x) = (1+px)^{\frac{1}{p}}(1-qx)^{\frac{1}{q}},$$

then f is defined on the interval $[-\frac{1}{p}, \frac{1}{q}] = [-\frac{j}{k}, 1 - \frac{j}{k}]$, and has derivative

$$f'(x) = -(p+q)(1+px)^{-\frac{1}{q}}(1-qx)^{-\frac{1}{p}}x.$$

For $x \in [0, \frac{1}{q}]$, $f'(x)$ is negative and satisfies

$$|f'(x)| \geq (p+q)\big(1 + \frac{p}{q}\big)^{-\frac{1}{q}} x = (p+q)^{\frac{1}{p}} q^{\frac{1}{q}} x \geq x.$$

Similarly, for $x \in [-\frac{1}{p}, 0]$, $f'(x)$ is positive and satisfies $f'(x) \geq |x|$. Noting that $f(0) = 1$, we then deduce

$$f(x) \leq 1 - \frac{x^2}{2}, \qquad x \in [-\frac{1}{p}, \frac{1}{q}].$$

Plugging this bound into (3.4) with $x = \epsilon$ (noting that ϵ could be negative), we get

$$\prod_{i=1}^{k} \Big(|\Delta_{\mathbf{u},i}\gamma| / |\Delta_i \gamma| \Big) \leq \left(\frac{1 - \frac{\epsilon^2}{2}}{1 - \delta(\frac{1}{k})} \right)^k.$$

It is then clear that if $\frac{\epsilon^2}{2} > \delta(\frac{1}{k}) + \frac{1}{k}$, we will necessarily have

$$\prod_{i=1}^{k} \Big(|\Delta_{\mathbf{u},i}\gamma| / |\Delta_i \gamma| \Big) \leq \left(1 - \frac{1}{k} \right)^k < \frac{1}{e}.$$

The case for general length L is essentially the same except one replaces $\delta(\frac{1}{k})$ by $\delta(\frac{1}{k})/L$. □

In view of Lemma 3.2, we let E_{k-1} be the set

$$E_{k-1} = \left\{ (u_1, \cdots, u_{k-1}) : \big| u_j - \frac{j}{k} \big| < \epsilon_k, j = 1, \cdots, k-1 \right\}. \tag{3.5}$$

We can now prove the following concentration property.

Proposition 3.3. *For every integer k such that $\delta(\frac{1}{k}) < \frac{L}{2}$ and every integer n, we have*

$$\int_{\Delta_{k-1} \cap E_{k-1}} \prod_{j=1}^{k} |\Delta_{\boldsymbol{u},j}\gamma|^n d\boldsymbol{u} \geq \big(1 - e^{3k \log k - \frac{n}{2}}\big) \int_{\Delta_{k-1}} \prod_{j=1}^{k} |\Delta_{\boldsymbol{u},j}\gamma|^n d\boldsymbol{u}. \tag{3.6}$$

Proof. Let $k \geq K$, and let $\mathcal{E}_{k-1}$ denote the set

$$\mathcal{E}_{k-1} = \left\{ \mathbf{v} = (v_1, \ldots, v_{k-1}) : \big| v_j - \frac{j}{k} \big| < \frac{1}{12k^2}, \;\; j = 1, \ldots, k-1 \right\}.$$

Then, for any $\mathbf{v} \in \mathcal{E}_{k-1}$ and $j = 1, \ldots, k$, we have

$$\big| |\Delta_{\mathbf{v},j}\gamma| - |\Delta_j \gamma| \big| \leq \big| \gamma_{v_j} - \gamma_{\frac{j}{k}} \big| + \big| \gamma_{v_{j-1}} - \gamma_{\frac{j-1}{k}} \big| < \frac{L}{6k^2}. \tag{3.7}$$

Since $\delta(\frac{1}{k}) < \frac{L}{2}$, we have $L \leq 2(L - \delta(\frac{1}{k}))$, so it follows from (3.7) and then Lemma 3.1 that

$$|\Delta_{\mathbf{v},j}\gamma| \geq |\Delta_j \gamma| - \frac{L - \delta(\frac{1}{k})}{3k^2} \geq \big(1 - \frac{1}{3k}\big) |\Delta_j \gamma|.$$

Since $(1 - \frac{1}{3k})^k$ is increasing in k, multiplying over j from 1 to k, we get

$$\prod_{j=1}^{k} |\Delta_{\mathbf{v},j}\gamma| \geq \frac{2}{3} \prod_{j=1}^{k} |\Delta_j \gamma| > e^{-\frac{1}{2}} \prod_{j=1}^{k} |\Delta_j \gamma| \tag{3.8}$$

for all $\mathbf{v} \in \mathcal{E}_{k-1}$. Now, raising both sides of (3.8) to power n and using Lemma 3.2, we get

$$\prod_{j=1}^{k} |\Delta_{\mathbf{u},j}\gamma|^n < e^{-\frac{n}{2}} \prod_{j=1}^{k} |\Delta_{\mathbf{v},j}\gamma|^n, \qquad \forall \;\; \mathbf{u} \in \Delta_{k-1} \cap E_{k-1}^c, \mathbf{v} \in \mathcal{E}_{k-1}.$$

Averaging both sides above in their respective domains, and noting $\mathcal{E}_{k-1} \subset \Delta_{k-1}$, we deduce that

$$\int_{\Delta_{k-1}\cap E_{k-1}^c} \prod_{j=1}^{k} |\Delta_{\mathbf{u},j}\gamma|^n d\mathbf{u} < C_k e^{-\frac{n}{2}} \int_{\Delta_{k-1}} \prod_{j=1}^{k} |\Delta_{\mathbf{u},j}\gamma|^n d\mathbf{u}, \tag{3.9}$$

where the constant C_k is given by

$$C_k = \frac{|\Delta_{k-1} \cap E_{k-1}^c|}{|\mathcal{E}_{k-1}|},$$

and $|\cdot|$ denotes the volume of a set. Since $|\Delta_{k-1} \cap E_{k-1}^c| < |\Delta_{k-1}| = \frac{1}{(k-1)!}$ and $|\mathcal{E}_{k-1}| = (6k^2)^{-(k-1)}$, using the bound

$$\log(k!) > \int_1^k \log x dx = k\log k - (k-1)$$

and the fact that $\left(1+\frac{1}{k-1}\right)^{k-1} < e$ for all k, we obtain the estimate

$$C_k < (6ek)^{k-1}.$$

The conclusion of the theorem follows by plugging this estimate of C_k into (3.9) and noting that $(k-1)\log(6ek) < 3k\log k$. □

Proposition 3.3 (and also Theorem 3.4 below) will be useful when $n \gg k\log k$, in which case the integral $\int_{\Delta_{k-1}} \prod_{j=1}^{k} |\Delta_{\mathbf{u},j}\gamma|^k$ is concentrated on the sub-domain $\Delta_{k-1} \cap E_{k-1}$. As a consequence, the sum of the symmetrized signatures $\mathcal{S}^{2n}(w,\boldsymbol{\ell})$'s cannot be too far away from its maximal possible value as $n \to +\infty$. This is the content of the following theorem.

Theorem 3.4. *For every k such that $\delta(2\epsilon_k) < \frac{L}{6}$, there exists a word w^* of length $k-1$ such that for every n, we have*

$$\sum_{\boldsymbol{\ell}\in\mathcal{L}_k^n} \left|\mathcal{S}_k^{2n}(w^*,\boldsymbol{\ell})\right| \geq \left(\frac{3}{6^k} - \left(1+\frac{3}{6^k}\right)e^{3k\log k - n}\right) L^{k-1} \int_{\Delta_{k-1}} \prod_{j=1}^{k} |\Delta_{\boldsymbol{u},j}\gamma|^{2n} d\boldsymbol{u}, \tag{3.10}$$

where $\mathcal{L}_k^n$ is set of multi-indices as defined in (2.1).

Proof. For any word w of length $k-1$, by (2.4), we have

$$\sum_{\boldsymbol{\ell}\in\mathcal{L}_k^n} \left|\mathcal{S}_k^{2n}(w,\boldsymbol{\ell})\right| \geq \left| \int_{\Delta_{k-1}} \prod_{j=1}^{k-1} \dot{\gamma}_{u_j}^{i_j} \prod_{j=1}^{k} \sum_{\ell_j=0}^{n} \binom{2n}{2\ell_j} (\Delta_{\mathbf{u},j}x)^{2\ell_j} (\Delta_{\mathbf{u},j}y)^{2n-2\ell_j} d\mathbf{u} \right|, \tag{3.11}$$

where we have interchanged the sum over $\boldsymbol{\ell}$ and the product over j since different components of $\boldsymbol{\ell}$ are summed up independently. The integrand of the right hand side of (3.11) can be split into two parts: the product of pointwise derivatives $\dot{\gamma}_{u_j}^{i_j}$, whose "direction" i_j is given by the j-th letter in the word w, and the product of the increments $\sum_{\ell_j} \binom{2n}{2\ell_j} (\Delta_{\mathbf{u},j}x)^{2\ell_j} (\Delta_{\mathbf{u},j}y)^{2n-2\ell_j}$. For the latter one, since

$$\sum_{\ell_j=0}^{n} \binom{2n}{2\ell_j} (\Delta_{\mathbf{u},j}x)^{2\ell_j} (\Delta_{\mathbf{u},j}y)^{2n-2\ell_j} = \frac{1}{2}\left((\Delta_{\mathbf{u},j}x + \Delta_{\mathbf{u},j}y)^{2n} + (\Delta_{\mathbf{u},j}x - \Delta_{\mathbf{u},j}y)^{2n} \right),$$

which is bounded from above by $|\Delta_{\mathbf{u},j}\gamma|^{2n}$ and from below by $\frac{1}{2}|\Delta_{\mathbf{u},j}\gamma|^{2n}$, we have the bound

$$\frac{1}{2^k}\prod_{j=1}^{k}|\Delta_{\mathbf{u},j}\gamma|^{2n} \le \prod_{j=1}^{k}\sum_{\ell_j=0}^{n}\binom{2n}{2\ell_j}(\Delta_{\mathbf{u},j}x)^{2\ell_j}(\Delta_{\mathbf{u},j}y)^{2n-2\ell_j} \le \prod_{j=1}^{k}|\Delta_{\mathbf{u},j}\gamma|^{2n}. \quad (3.12)$$

Now we look at the first part of the integrand, $\prod_{j=1}^{k-1}\dot{\gamma}_{u_j}^{i_j}$. Since we hope the whole integral on the right hand side of (3.11) to be concentrated on the domain $\Delta_{k-1}\cap E_{k-1}$, we choose a word $w^* = e_{i_1}\cdots e_{i_{k-1}}$ such that

$$|\dot{\gamma}_{u_j}^{i_j}| \ge \frac{L}{3} \quad (3.13)$$

for all j and all $\mathbf{u}\in E_{k-1}$. This also guarantees that none of the $\dot{\gamma}_{u_j}^{i_j}$'s changes its sign in this domain. The main purpose of choosing w^* in this way is to ensure that the term $\prod_j \dot{\gamma}_{u_j}^{i_j}$ does not cause any degeneracy or cancellations of the integral in its domain of concentration $\Delta_{k-1}\cap E_{k-1}$. The continuity of $\dot{\gamma}$ ensures that we can always find such a word as long as $\delta(2\epsilon_k) < \frac{L}{6}$.

We now decompose the the integral on the right hand side of (3.11) into two disjoint domains: $\Delta_{k-1}\cap E_{k-1}$ and $\Delta_{k-1}\cap E_{k-1}^c$. For the first one, since the product $\prod_{j=1}^{k-1}\dot{\gamma}_{u_j}^{i_j}$ is bounded away from 0 by $(L/3)^{k-1}$ and does not change its sign in E_{k-1}, and the rest of the integrand is always positive as it only contains even powers, we can move the absolute value into the integral to get

$$\begin{aligned}&\Big|\int_{\Delta_{k-1}\cap E_{k-1}}\prod_{j=1}^{k-1}\dot{\gamma}_{u_j}^{i_j}\prod_{j=1}^{k}\sum_{\ell_j=0}^{n}\binom{2n}{2\ell_j}(\Delta_{\mathbf{u},j}x)^{2\ell_j}(\Delta_{\mathbf{u},j}y)^{2n-2\ell_j}d\mathbf{u}\Big| \\ &\ge \frac{1}{2^k}\Big(\frac{L}{3}\Big)^{k-1}\big(1-e^{3k\log k-n}\big)\int_{\Delta_{k-1}}\prod_{j=1}^{k}|\Delta_{\mathbf{u},j}\gamma|^{2n}d\mathbf{u},\end{aligned} \quad (3.14)$$

where we used the lower bound in (3.12) to reduce the integrand to $\frac{1}{2^k}\prod_j|\Delta_{\mathbf{u},j}\gamma|^{2n}$, the bound (3.13) to replace $\prod_j|\dot{\gamma}_{u_j}^{i_j}|$ by $(L/3)^{k-1}$, and Proposition 3.3 to enlarge the domain of integration to Δ_{k-1} with a compensation of the factor $1-e^{3k\log k-n}$. Note that the exponent here is $3k\log k - n$ instead of $3k\log k - \frac{n}{2}$ since the power of $|\Delta_{\mathbf{u},j}\gamma|$ is $2n$.

For the integration over the domain $\Delta_{k-1}\cap E_{k-1}^c$, it follows from $|\dot{\gamma}_{u_j}^{i_j}|\le L$, the upper bound in (3.12) and Proposition 3.3 that it is bounded by

$$e^{3k\log k-n}L^{k-1}\int_{\Delta_{k-1}}\prod_{j=1}^{k}|\Delta_{\mathbf{u},j}\gamma|^{2n}d\mathbf{u}. \quad (3.15)$$

Combining (3.14) and (3.15), we obtain (3.10) and thus finish the proof of the theorem. □

4 Reconstructing the path

We are now ready to reconstruct the path from its signature. Recall from (1.4) that

$$\eta_k = \delta(3\epsilon_k) + \frac{L}{\sqrt{k}}.$$

Our aim is to find the parameters $\rho_j \in [0,1]$, $a_j^x, a_j^y \in \{\pm 1\}$ and $\tilde{L} \in \mathbb{R}^+$ such that

$$\sup_{1\le j\le k} \sup_{u\in[\frac{j-1}{k},\frac{j}{k}]} \left|\tilde{L}\big(a_j^x\rho_j, a_j^y(1-\rho_j)\big) - \dot{\gamma}_u\right| < C\eta_k \tag{4.1}$$

for all large enough k, where C is some constant depending on γ only. In what follows, we will recover these parameters in the order of ρ_j's, a_j^x, a_j^y's, and then finally $\tilde{L}$. We will also give in Theorem 4.17 the precise value of the constant C and a quantitative characterization of how large k needs to be.

Remark 4.1. As for the magnitude of η_k, if $\dot{\gamma} \in \mathcal{C}^\alpha$ for some $\alpha \in (0,1)$, then

$$\epsilon_k \lesssim \sqrt{\frac{\delta(1/k)}{L}} + \frac{1}{\sqrt{k}} \lesssim k^{-\frac{\alpha}{2}},$$

and consequently, we have

$$\eta_k = \delta(3\epsilon_k) + \frac{L}{\sqrt{k}} \lesssim k^{-\frac{\alpha^2}{2}}.$$

The readers might wonder why we add the additional term $\frac{L}{\sqrt{k}}$ to the definition of η_k. In fact, as long as γ is not a straight line, then there exists $\lambda > 0$ such that

$$\delta(3\epsilon_k) > 3\lambda\epsilon_k \ge \frac{3\lambda}{\sqrt{k}}.$$

Thus, $\frac{L}{\sqrt{k}} < C\delta(3\epsilon_k)$, and adding this additional term $\frac{L}{\sqrt{k}}$ does not change the magnitude of η_k for large k. Since we will need a lower bound for η_k in Theorem 4.4 below, this additional term will save us from introducing the new constant λ.

4.1 The unsigned directions

We start with the recovery of unsigned directions ρ_j's. At this stage, we only use the quantities $\mathcal{S}_k^{2n}(w,\boldsymbol{\ell})$'s which can be obtained from symmetrization of the signature as in (2.3) and have an expression as in (2.4). Recall from Section 1.3 that

$$\Delta_j x = x_{j/k} - x_{(j-1)/k}, \quad \Delta_j y = y_{j/k} - y_{(j-1)/k}, \quad |\Delta_j\gamma| = |\Delta_j x| + |\Delta_j y|.$$

Since we expect ρ_j to be close to the increment of$\gamma|_{[\frac{j-1}{k},\frac{j}{k}]}$, it is natural to introduce for each j the unique real number $r_j \in [0,1]$ such that

$$|\Delta_j x| : |\Delta_j y| = r_j : (1-r_j).$$

Also recall from (3.5) the definition of E_{k-1}, which we will be frequently using throughout this section. The following two elementary lemmas will be useful in the sequel.

Lemma 4.2. *For every k such that $\delta(3\epsilon_k) < \frac{L}{2}$ and every $\boldsymbol{u} \in \Delta_{k-1} \cap E_{k-1}$, we have*

$$\sup_{1\le j\le k} \left|\frac{|\Delta_{\boldsymbol{u},j}x|}{|\Delta_{\boldsymbol{u},j}\gamma|} - r_j\right| < \frac{\eta_k}{L}. \tag{4.2}$$

Proof. Fix $\mathbf{u} \in \Delta_{k-1} \cap E_{k-1}$ and $1 \le j \le k$. Let

$$Q_j = \big[\frac{j-1}{k} - \epsilon_k,\ \frac{j}{k} + \epsilon_k\big] \cap [0,1].$$

Then, $Q_j \supset \big([\frac{j-1}{k}, \frac{j}{k}] \cup [u_{j-1}, u_j]\big)$ and $|Q_j| < 3\epsilon_k$. It suffices to prove (4.2) in the case when $|\dot{x}_v| \geq \frac{L}{2}$ for some $v \in Q_j$, for otherwise we would have $|\dot{y}_v| \geq \frac{L}{2}$ for some $v \in Q_j$ and can get (4.2) through

$$\left| \frac{|\Delta_{\mathbf{u},j} x|}{|\Delta_{\mathbf{u},j}\gamma|} - r_j \right| = \left| \frac{|\Delta_{\mathbf{u},j} y|}{|\Delta_{\mathbf{u},j}\gamma|} - (1 - r_j) \right|.$$

Since $|\dot{x}_v| \geq \frac{L}{2}$ for some $v \in Q_j$, $|Q_j| < 3\epsilon_k$, and $\delta(3\epsilon_k) < \frac{L}{2}$, we have

$$\inf_{t \in Q_j} |\dot{x}_t| > 0.$$

In particular, x is monotone in Q_j and $\dot{x}$ does not change its sign. If y is also monotone in Q_j, then $|\Delta_{\mathbf{u},j}\gamma| = L(u_j - u_{j-1})$ and $|\Delta_j \gamma| = \frac{L}{k}$. Thus, there exist $v, \tilde{v} \in Q_j$ such that

$$\frac{|\Delta_{\mathbf{u},j} x|}{|\Delta_{\mathbf{u},j}\gamma|} = \frac{|\dot{x}_v|}{L}, \qquad r_j = \frac{|\dot{x}_{\tilde{v}}|}{L},$$

and (4.2) follows since $|v - \tilde{v}| < |Q_j| < 3\epsilon_k$. If y is not monotone in Q_j, then there exists $s \in Q_j$ such that $\dot{y}_s = 0$, and we have the bound

$$\inf_{t \in Q_j} |\dot{x}_t| \geq L - \delta(3\epsilon_k).$$

This implies that

$$\frac{L - \delta(3\epsilon_k)}{L} \leq \frac{|\Delta_{\mathbf{u},j} x|}{|\Delta_{\mathbf{u},j}\gamma|} \leq 1, \qquad \frac{L - \delta(3\epsilon_k)}{L} \leq r_j \leq 1,$$

and the bound (4.2) follows immediately. □

Lemma 4.3. *Let $m \geq 1$. For every $\ell = 0, \ldots, m$ and $p \in [0,1]$ we have*

$$\binom{m}{\ell} p^\ell (1-p)^{m-\ell} \leq m e^{-\frac{mx^2}{2}},$$

where $x = \frac{\ell}{m} - p \in [-p, 1-p]$.

Proof. The lemma is clearly true if $p = 0$ or 1. For $p \in (0,1)$, using the estimate

$$\ell \log \ell - (\ell - 1) = \int_1^\ell \log x dx < \log(\ell!) < \int_1^{\ell+1} \log x dx = (\ell+1)\log(\ell+1) - \ell$$

as well as those for $(m-\ell)!$ and $m!$, we obtain the bound

$$f(x) := \binom{m}{\ell} p^\ell q^{m-\ell} < m \left[\left(\frac{p}{p+x} \right)^{p+x} \left(\frac{q}{q-x} \right)^{q-x} \right]^m,$$

where $x = \frac{\ell}{m} - p \in [-p, q]$, and $q = 1 - p$. Let

$$g(x) = (p+x)\big(\log p - \log(p+x)\big) + (q-x)\big(\log q - \log(q-x)\big).$$

It is straightforward to check that

$$g'(x) = \log\left(1 - \frac{x}{q(p+x)}\right)$$

satisfies $g'(x) \le -x$ for $x \in [0, q]$ and $g'(x) \ge |x|$ for $x \in [-p, 0]$. As a consequence, we have

$$g(x) \le -\frac{x^2}{2}, \qquad x \in [-p, q],$$

and that

$$f(x) = me^{mg(x)} \le me^{-\frac{mx^2}{2}}.$$

This finishes the proof. □

We are now ready to prove the following main theorem about the unsigned directions.

Theorem 4.4. *For any $k \ge 2$ such that $\eta_k < \frac{L}{6}$, $n = 4k^2 \log k$ and $j = 1, \dots, k$, we have*

$$\Big(\sum_{|w|=k-1} \sum_{\boldsymbol{\ell}: |\frac{\ell_j}{n} - r_j| \ge \frac{2\eta_k}{L}} |\mathcal{S}_k^{2n}(w, \boldsymbol{\ell})| \Big) \Big/ \Big(\sum_{|w|=k-1} \sum_{\boldsymbol{\ell} \in \mathcal{L}_k^n} |\mathcal{S}_k^{2n}(w, \boldsymbol{\ell})| \Big) < 16k^5 \Big(\frac{12}{k^4} \Big)^k, \tag{4.3}$$

where the sum for w is taken over all words with length $k-1$, and in the numerator, the sum for multi-indices $\boldsymbol{\ell}$ is taken over all $\boldsymbol{\ell} \in \mathcal{L}_k^n$ such that $|\frac{\ell_j}{n} - r_j| \ge \frac{2\eta_k}{L}$.

Proof. Since both the numerator and denominator on the left hand side of (4.3) scale like L^{2n+k-1}, we can assume without loss of generality that $L = 1$. Now fix n and large k whose values will be chosen later, and also fix $1 \le j \le k$. Since $|\dot\gamma| \equiv 1$, for any word w with length $k-1$, we have

$$\begin{aligned} \sum_{\boldsymbol{\ell}: |\frac{\ell_j}{n} - r_j| > 2\eta_k} |\mathcal{S}_k^{2n}(w, \boldsymbol{\ell})| &\le e^{3k \log k - n} \int_{\Delta_{k-1}} \prod_{i=1}^k |\Delta_{\mathbf{u},i}\gamma|^{2n} d\mathbf{u} \\ + \int_{\Delta_{k-1} \cap E_{k-1}} \sum_{|\frac{\ell_j}{n} - r_j| > 2\eta_k} &\binom{2n}{2\ell_j} \Big(\frac{|\Delta_{\mathbf{u},j}x|}{|\Delta_{\mathbf{u},j}\gamma|} \Big)^{2\ell_j} \Big(\frac{|\Delta_{\mathbf{u},j}y|}{|\Delta_{\mathbf{u},j}\gamma|} \Big)^{2n-2\ell_j} \prod_{i=1}^k |\Delta_{\mathbf{u},i}\gamma|^{2n} d\mathbf{u}, \end{aligned} \tag{4.4}$$

where we have split the integral into two disjoint domains and applied Proposition 3.3 to bound the one on $\Delta_{k-1} \cap E_{k-1}^c$. The range of sum is $|\frac{\ell_j}{n} - r_j| > 2\eta_k$ since we assumed $L = 1$.

Now we need to bound the integrand for the second term on the right hand side of (4.4). Fix $\mathbf{u} \in \Delta_{k-1} \cap E_{k-1}$, and let

$$p = \frac{|\Delta_{\mathbf{u},j}x|}{|\Delta_{\mathbf{u},j}\gamma|}.$$

Lemma 4.2 implies that if $|\frac{\ell_j}{n} - r_j| > 2\eta_k$, then we must have $|\frac{\ell_j}{n} - p| > \eta_k$. Thus, using Lemma 4.3, we obtain the pointwise bound

$$\binom{2n}{2\ell_j} p^{2\ell_j} (1-p)^{2n-2\ell_j} \le 2n \exp\big(-n \big| \frac{\ell_j}{n} - p \big|^2 \big) < 2ne^{-n\eta_k^2}.$$

Since this bound holds for all $\mathbf{u} \in \Delta_{k-1} \cap E_{k-1}$ (recall that p depends on $\mathbf{u}$) and all ℓ_j such that $\big|\frac{\ell_j}{n} - r_j\big| > 2\eta_k$, we get

$$\sum_{|\frac{\ell_j}{n} - r_j| > 2\eta_k} \binom{2n}{2\ell_j} \Big(\frac{|\Delta_{\mathbf{u},j}x|}{|\Delta_{\mathbf{u},j}\gamma|} \Big)^{2\ell_j} \Big(\frac{|\Delta_{\mathbf{u},j}y|}{|\Delta_{\mathbf{u},j}\gamma|} \Big)^{2n-2\ell_j} \le 2n^2 e^{-n\eta_k^2}, \quad \forall \mathbf{u} \in \Delta_{k-1} \cap E_{k-1}.$$

Now, plugging the above pointwise bound into (4.4) and summing over all words w with length $k-1$, we obtain an upper bound for the numerator in (4.4) as

$$\sum_{w} \sum_{\boldsymbol{\ell}:|\frac{\ell_j}{n}-r_j|\geq\frac{2\eta_k}{L}} \left|\mathcal{S}_k^{2n}(w,\boldsymbol{\ell})\right| < C_{n,k} \int_{\Delta_{k-1}} \prod_{i=1}^{k} |\Delta_{\mathbf{u},i}\gamma|^{2n} d\mathbf{u}, \tag{4.5}$$

where

$$C_{n,k} = 2^{k-1}\big(e^{3k\log k - n} + 2n^2 e^{-n\eta_k^2}\big). \tag{4.6}$$

Applying Theorem 3.4, we have a lower bound for the denominator $\sum_w \sum_{\boldsymbol{\ell}} |\mathcal{S}_k^{2n}(w,\boldsymbol{\ell})|$ in (4.3), which, when combined with (4.5) and (4.6), implies that the left hand side of (4.3) is bounded by

$$\frac{2^{k-1}\big(e^{3k\log k - n} + 2n^2 e^{-n\eta_k^2}\big)}{\frac{3}{6^k} - \big(1+\frac{3}{6^k}\big)e^{3k\log k - n}}.$$

The theorem then follows by taking $n = 4k^2 \log k$ and noting that $\eta_k^2 \geq \frac{1}{k}$ (recall we have set $L = 1$). □

Remark 4.5. Note that the left hand side of (4.3) is always smaller than 1, so the theorem is meaningful when the right hand side also falls below 1, which is the case when $k \geq 4$ (in fact it is already smaller than $\frac{1}{8}$ when $k = 4$).

The following easy corollary enables one to select the directions ρ_j (up to the sign) for our piecewise linear approximation.

Corollary 4.6. *[**Choosing directions**] Let $k \geq 4$ be such that $\eta_k < \frac{L}{6}$. For each $0 \leq j \leq k$, there exists $\rho_j \in [0,1]$ such that*

$$\Bigg(\sum_{|w|=k-1} \sum_{\boldsymbol{\ell}:|\frac{\ell_j}{n}-\rho_j|\leq\frac{2\eta_k}{L}} \left|\mathcal{S}_k^{2n}(w,\boldsymbol{\ell})\right|\Bigg) \Big/ \Bigg(\sum_{|w|=k-1} \sum_{\boldsymbol{\ell}\in\mathcal{L}_k^n} \left|\mathcal{S}_k^{2n}(w,\boldsymbol{\ell})\right|\Bigg) > \frac{1}{2}, \tag{4.7}$$

where in the numerator, the sum over $\boldsymbol{\ell}$ in the appropriate subset of $\mathcal{L}^{n_k}$ is implicit. Moreover, if $\{\rho_j\}_{j=1}^k$ is any set that satisfies (4.7), then we must have

$$|\rho_j - r_j| < \frac{4\eta_k}{L}$$

for all $j = 1, \cdots, k$.

Proof. The existence of the set $\{\rho_j\}$ that satisfies (4.7) follows directly by setting $\rho_j = r_j$ and applying Theorem 4.4. Conversely, if $|\rho_j - r_j| \geq \frac{4\eta_k}{L}$ for some j, then $|\frac{\ell_j}{n} - \rho_j| < \frac{2\eta_k}{L}$ implies that

$$|\frac{\ell_j}{n} - r_j| > \frac{2\eta_k}{L}.$$

By Theorem 4.4, this set of $\{\rho_j\}$ must violate (4.7). This completes the proof. □

Remark 4.7. Note that on the left hand side of (4.7), there is only information that is available from the signature of γ. If we choose $\{\rho_j\}$ according to (4.7), then Corollary 4.6 guarantees that these unsigned directions we recover from the signature must be close to the true directions $\{r_j\}$.

Remark 4.8. The readers might have noticed that the criterion of choosing the ρ_j's above involves the knowledge of η_k. This is of course not a problem if we know the modulus of continuity of $\dot\gamma$ in advance, as in this case

$$\frac{\eta_k}{L} = \frac{\delta(3\epsilon_k)}{L} + \frac{1}{\sqrt{k}}$$

is explicitly known. But even if that information is not available, we can always choose a sequence α_k which decreases to 0 slowly enough such that $\alpha_k > \frac{2\eta_k}{L}$, and replace the range of the sum in the numerator of (4.7) by

$$\boldsymbol{\ell} : |\frac{\ell_j}{n} - \rho_j| \le \alpha_k.$$

The directions $\{\rho_j\}$ chosen in this way satisfy

$$|\rho_j - r_j| < \alpha_k + \frac{2\eta_k}{L},$$

which still goes to 0 as $k \to +\infty$, but at a slower rate than η_k.

4.2 The signs

We now turn to the recovery of the sign of the direction of each piece. For this, we need to move one level up in the signatures, which requires more information than the $\mathcal{S}_k^{2n}(w, \boldsymbol{\ell})$'s.

We look at the signature at level $2nk + k$ (in addition to the level $2nk + k - 1$ before), and divide it into k blocks of size $2n$ except one of them which has size $2n + 1$, still with one letter separating consecutive blocks. More precisely, for any word $w = e_{i_1} \cdots e_{i_{k-1}}$, any multi-index $\boldsymbol{\ell} \in \mathcal{L}_k^n$, and any $1 \le i \le k$, we let $\mathcal{W}_{k,i,x}^{2n}(w, \boldsymbol{\ell})$ denote the set of words

$$w' = w_1 * e_{i_1} * \cdots * e_{i_{k-1}} * w_k$$

such that $|w_j|_y = 2n - 2\ell_j$ for every $1 \le j \le k$, $|w_j|_x = 2\ell_j$ for every $j \ne i$, but $|w_i|_x = 2\ell_i + 1$. The set $\mathcal{W}_{k,i,x}^{2n}(w, \boldsymbol{\ell})$ is different from $\mathcal{W}_k^{2n}(w, \boldsymbol{\ell})$ defined in (2.2) in that the i-th block has size $2n + 1$ instead of $2n$, and contains $2\ell_i + 1$ x's instead of $2\ell_i$. We define the set $\mathcal{W}_{k,i,y}^{2n}(w, \boldsymbol{\ell})$ in the same way except that $|w_i|_x = 2\ell_i$ and $|w_i|_y = 2n - 2\ell_i + 1$.

We then define the quantities $\mathcal{S}_{k,i,x}^{2n}(w, \boldsymbol{\ell})$ and $\mathcal{S}_{k,i,y}^{2n}(w, \boldsymbol{\ell})$ to be

$$\begin{aligned} \mathcal{S}_{k,i,x}^{2n}(w, \boldsymbol{\ell}) &= (2n+1)\big((2n)!\big)^k \sum_{w' \in \mathcal{W}_{k,i,x}^{2n}(w,\boldsymbol{\ell})} C(w'), \\ \mathcal{S}_{k,i,y}^{2n}(w, \boldsymbol{\ell}) &= (2n+1)\big((2n)!\big)^k \sum_{w' \in \mathcal{W}_{k,i,y}^{2n}(w,\boldsymbol{\ell})} C(w'). \end{aligned} \tag{4.8}$$

The aim of introducing these quantities is to recover the sign of x and y directions in the i-th piece of the path via comparison with $\mathcal{S}_k^{2n}(w, \boldsymbol{\ell})$. Similar to (2.4), we have the expressions

$$\begin{aligned} \mathcal{S}_{k,i,x}^{2n}(w, \boldsymbol{\ell}) = \int_{\Delta_{k-1}} \prod_{j=1}^{k-1} \dot\gamma_{u_j}^{i_j} \cdot \binom{2n+1}{2\ell_i+1} (\Delta_{\mathbf{u},i}x)^{2\ell_i+1} (\Delta_{\mathbf{u},i}y)^{2n-2\ell_i} \\ \prod_{j \ne i} \binom{2n}{2\ell_j} (\Delta_{\mathbf{u},j}x)^{2\ell_j} (\Delta_{\mathbf{u},j}y)^{2n-2\ell_j} d\mathbf{u}, \end{aligned} \tag{4.9}$$

and

$$\mathcal{S}^{2n}_{k,i,y}(w,\boldsymbol{\ell}) = \int_{\Delta_{k-1}} \prod_{j=1}^{k-1} \dot{\gamma}^{i_j}_{u_j} \cdot \binom{2n+1}{2\ell_i} (\Delta_{\mathbf{u},i}x)^{2\ell_i} (\Delta_{\mathbf{u},i}y)^{2n+1-2\ell_i} \prod_{j\neq i} \binom{2n}{2\ell_j} (\Delta_{\mathbf{u},j}x)^{2\ell_j} (\Delta_{\mathbf{u},j}y)^{2n-2\ell_j} d\mathbf{u}. \tag{4.10}$$

Again, we emphasize that (4.8) is the information available from the signature, while (4.9) and (4.10) are expressions for these quantities.

Unlike in choosing unsigned directions where we sum up the $\mathcal{S}^{2n}_k(w,\boldsymbol{\ell})$'s over all words w that have length $k-1$ (see Corollary 4.6), in what follows, we will use the quantities (4.8) with only one particular word w^*. We choose this word as follows. Let $\{\rho_j\}$ be the set of unsigned directions chosen according to Corollary 4.6, then for each $j = 1, \cdots, k-1$, we let

$$\begin{aligned} e_{i_j} &= x, \qquad \text{if } \rho_j \geq \frac{1}{2}, \\ e_{i_j} &= y, \qquad \text{if } \rho_j < \frac{1}{2}, \end{aligned} \tag{4.11}$$

and we set the word w^* to be

$$w^* = e_{i_1} \cdots e_{i_{k-1}}. \tag{4.12}$$

The word w^* chosen above guarantees that $\prod_j \dot{\gamma}^{i_j}_{u_j}$ is bounded away from 0 in the region $\Delta_{k-1} \cap E_{k-1}$. This is the content of the following proposition.

Proposition 4.9. *Let k be large enough such that $\eta_k < \frac{L}{32}$, and let w^* be the word chosen according to* (4.11) *and* (4.12)*. Then, for all $\boldsymbol{u} \in E_{k-1}$ and all $j = 1, \ldots, k-1$, we have*

$$|\dot{\gamma}^{i_j}_{u_j}| > \frac{L}{3}$$

and that $\dot{\gamma}^{i_j}_{u_j}$ does not change its sign in the domain E_{k-1}.

Proof. Fix k as in the assumption and $1 \leq j \leq k-1$. If $\rho_j \geq \frac{1}{2}$, then $e_{i_j} = x$, and Corollary 4.6 implies

$$r_j \geq \frac{1}{2} - \frac{4\eta_k}{L} > \frac{3}{8}.$$

Thus, there exist $v, \tilde{v} \in [\frac{j-1}{k}, \frac{j}{k}]$ such that $|\dot{x}_v| > \frac{3}{8}(|\dot{x}_v| + |\dot{y}_{\tilde{v}}|)$. Using $|\dot{y}_{\tilde{v}}| = L - |\dot{x}_{\tilde{v}}|$, we see that either $|\dot{x}_v|$ or $|\dot{x}_{\tilde{v}}|$ is bigger than $\frac{3L}{8}$. If $\mathbf{u} \in E_{k-1}$, then both $|u-v|$ and $|u-\tilde{v}|$ is smaller than $2\epsilon_k$, so it follows that (recall $\eta_k > \delta(3\epsilon_k)$)

$$|\dot{x}_{u_j}| > \frac{3L}{8} - \eta_k > \frac{3L}{8} - \frac{L}{32} > \frac{L}{3}.$$

Similarly, we have $|\dot{y}_{u_j}| > \frac{L}{3}$ for all $\mathbf{u} \in E_{k-1}$ if $\rho_j < \frac{1}{2}$. In particular, the continuity of the derivatives ensures these $\dot{x}_{u_j}$'s and $\dot{y}_{u_j}$'s do not change their signs in the domain E_{k-1}. □

Note that the word w^* chosen above has all the properties we used in Theorem 3.4. Theorem 3.4 only gives the existence of such a word, but here we choose it explicitly based on the recovery of the unsigned directions. We now determine the signs of the i-th piece as follows.

Definition 4.10. *[**Determining the signs**] Fix $k \geq 4$ such that $\delta(\frac{1}{k}) < \frac{L}{6}$, and let $n = 4k^2 \log k$. Let w^* be the word chosen according to* (4.11) *and* (4.12). *We then choose the sign $a_i^x, a_i^y \in \{\pm 1\}$ for the i-th linear piece by:*

$$a_i^x = 1, \qquad \text{if } \frac{\sum_{\boldsymbol{\ell} \in \mathcal{L}_k^n} \mathcal{S}_k^{2n}(w^*, \boldsymbol{\ell})}{\sum_{\boldsymbol{\ell} \in \mathcal{L}_k^n} \mathcal{S}_{k,i,x}^{2n}(w^*, \boldsymbol{\ell})} \geq 0,$$

$$a_i^x = -1, \qquad \text{if } \frac{\sum_{\boldsymbol{\ell} \in \mathcal{L}_k^n} \mathcal{S}_k^{2n}(w^*, \boldsymbol{\ell})}{\sum_{\boldsymbol{\ell} \in \mathcal{L}_k^n} \mathcal{S}_{k,i,x}^{2n}(w^*, \boldsymbol{\ell})} < 0.$$

The choice for a_i^y is the same except that one replaces $\mathcal{S}_{k,i,x}^{2n}(w^*, \boldsymbol{\ell})$ by $\mathcal{S}_{k,i,y}^{2n}(w^*, \boldsymbol{\ell})$.

It appears that the above choices of the signs depend on k, and the choices may be different if k changes. But it turns out that the choices above remain stable for all sufficiently large k, and they indeed give the correct signs as long as the directions are not close to degenerate. The remaining of this subsection will be devoted to the verification of this stability.

Theorem 4.11. *Let $k \geq 4$ be sufficiently large such that $\eta_k < \frac{L}{32}$, and $n = 4k^2 \log k$. If $r_i \geq \frac{2\eta_k}{L}$, then*

$$\frac{\sum_{\boldsymbol{\ell}} \mathcal{S}_k^{2n}(w^*, \boldsymbol{\ell})}{\sum_{\boldsymbol{\ell}} \mathcal{S}_{k,i,x}^{2n}(w^*, \boldsymbol{\ell})} \geq \frac{1}{6\epsilon_k L} \qquad \text{if } \Delta_i x > 0,$$

$$\frac{\sum_{\boldsymbol{\ell}} \mathcal{S}_k^{2n}(w^*, \boldsymbol{\ell})}{\sum_{\boldsymbol{\ell}} \mathcal{S}_{k,i,x}^{2n}(w^*, \boldsymbol{\ell})} \leq -\frac{1}{6\epsilon_k L} \qquad \text{if } \Delta_i x < 0.$$

Similarly, if $r_i \leq 1 - \frac{2\eta_k}{L}$, then

$$\frac{\sum_{\boldsymbol{\ell}} \mathcal{S}_k^{2n}(w^*, \boldsymbol{\ell})}{\sum_{\boldsymbol{\ell}} \mathcal{S}_{k,i,y}^{2n}(w^*, \boldsymbol{\ell})} \geq \frac{1}{6\epsilon_k L} \qquad \text{if } \Delta_i y > 0,$$

$$\frac{\sum_{\boldsymbol{\ell}} \mathcal{S}_k^{2n}(w^*, \boldsymbol{\ell})}{\sum_{\boldsymbol{\ell}} \mathcal{S}_{k,i,y}^{2n}(w^*, \boldsymbol{\ell})} < -\frac{1}{6\epsilon_k L} \qquad \text{if } \Delta_i y < 0.$$

All the sums above are taken over $\boldsymbol{\ell} \in \mathcal{L}_k^n$, and $\Delta_i x$ and $\Delta_i y$ are defined in (1.8).

Remark 4.12. The above theorem guarantees that as long as the i-th piece of γ is not too horizontal or vertical (corresponding to the assumptions $r_i \geq \frac{2\eta_k}{L}$ and $r_i \leq 1 - \frac{2\eta_k}{L}$), then the choices in Definition 4.10 do give the correct signs for all sufficiently large k. The case $r_i < \frac{2\eta_k}{L}$ and $r_i > 1 - \frac{2\eta_k}{L}$ are not covered, but since the i-th piece would be almost horizontal or vertical in that situation, the choice of the sign would not affect accuracy.

Proof. (of Theorem 4.11). We only prove the first case when $r_i \geq \frac{2\eta_k}{L}$ and $\Delta_i x > 0$; the other three cases are essentially the same. By the expressions (2.4) and (4.9), we can write

$$\sum_{\boldsymbol{\ell} \in \mathcal{L}_k^n} \mathcal{S}_k^{2n}(w^*, \boldsymbol{\ell}) = \int_{\Delta_{k-1}} \mathcal{N}(\mathbf{u}) d\mathbf{u}, \qquad \sum_{\boldsymbol{\ell} \in \mathcal{L}_k^n} \mathcal{S}_{k,i,x}^{2n}(w^*, \boldsymbol{\ell}) = \int_{\Delta_{k-1}} \mathcal{D}(\mathbf{u}) d\mathbf{u}.$$

Here, $\mathcal{N}(\mathbf{u})$ and $\mathcal{D}(\mathbf{u})$ are respectively given by

$$\mathcal{N}(\mathbf{u}) = \frac{1}{2^k} \prod_{j=1}^{k-1} \dot{\gamma}_{u_j}^{i_j} \prod_{j=1}^{k} \Big(\big(\Delta_{\mathbf{u},j} x + \Delta_{\mathbf{u},j} y\big)^{2n} + \big(\Delta_{\mathbf{u},j} x - \Delta_{\mathbf{u},j} y\big)^{2n} \Big),$$

and

$$\mathcal{D}(\mathbf{u}) = \frac{1}{2^k}\prod_{j=1}^{k-1}\dot{\gamma}_{u_j}^{i_j}\prod_{j=1}^{k}\Big(\big(\Delta_{\mathbf{u},j}x+\Delta_{\mathbf{u},j}y\big)^{2n+\delta_{i,j}}+\big(\Delta_{\mathbf{u},j}x-\Delta_{\mathbf{u},j}y\big)^{2n+\delta_{i,j}}\Big), \quad (4.13)$$

where $\delta_{i,j}=1$ if $j=i$ and 0 otherwise. We now need to estimate the ratio of two integrals, both over the domain Δ_{k-1}. Similar as before, we decompose both integrals into domains $\Delta_{k-1}\cap E_{k-1}$ and $\Delta_{k-1}\cap E_{k-1}^c$. We postpone the estimate of the latter one to the next lemma, and first consider the quantity

$$\frac{\int_{\Delta_{k-1}\cap E_{k-1}}\mathcal{N}(u)d\mathbf{u}}{\int_{\Delta_{k-1}\cap E_{k-1}}\mathcal{D}(u)d\mathbf{u}}. \quad (4.14)$$

We hope to control the ratio of the two integrals by means of a pointwise bound on

$$\frac{\mathcal{N}(\mathbf{u})}{\mathcal{D}(\mathbf{u})}=\frac{\big(\Delta_{\mathbf{u},i}x+\Delta_{\mathbf{u},i}y\big)^{2n}+\big(\Delta_{\mathbf{u},i}x-\Delta_{\mathbf{u},i}y\big)^{2n}}{\big(\Delta_{\mathbf{u},i}x+\Delta_{\mathbf{u},i}y\big)^{2n+1}+\big(\Delta_{\mathbf{u},i}x-\Delta_{\mathbf{u},i}y\big)^{2n+1}}, \quad \mathbf{u}\in\Delta_{k-1}\cap E_{k-1}. \quad (4.15)$$

For this, we first note that similar as in Proposition 4.9, the assumption $r_i\geq\frac{2\eta_k}{L}$ implies that $|\dot{x}_v|\geq 2\eta_k$ for some $v\in[\frac{i-1}{k},\frac{i}{k}]$. Also, since $\Delta_i x>0$ and $2\eta_k>\delta(\frac{1}{k})$, we actually have $\dot{x}_v\geq 2\eta_k$. If $\mathbf{u}\in E_{k-1}$, we will have $|u_i-v|<2\epsilon_k$ and hence

$$\dot{x}_{u_i}>2\eta_k-\delta(2\epsilon_k)>\eta_k. \quad (4.16)$$

In particular, this implies $\Delta_{\mathbf{u},i}x$ is positive and so is the ratio $\frac{\mathcal{N}(u)}{\mathcal{D}(u)}$ for all $\mathbf{u}\in E_{k-1}$. Since

$$\max\big\{|\Delta_{\mathbf{u},i}x+\Delta_{\mathbf{u},i}y|,|\Delta_{\mathbf{u},i}x-\Delta_{\mathbf{u},i}y|\big\}=|\Delta_{\mathbf{u},i}\gamma|,$$

we then have the bound

$$\frac{\mathcal{N}(\mathbf{u})}{\mathcal{D}(\mathbf{u})}\geq\frac{1}{|\Delta_{\mathbf{u},i}\gamma|}\geq\frac{1}{3\epsilon_k L}$$

for all $\mathbf{u}\in E_{k-1}$. Note that the choice of the word w^* guarantees both $\mathcal{N}(\mathbf{u})$ and $\mathcal{D}(\mathbf{u})$ keep their signs unchanged in the domain E_{k-1}, so the pointwise bound carries to the ratio of the integrals, which gives

$$\frac{\int_{\Delta_{k-1}\cap E_{k-1}}\mathcal{N}(\mathbf{u})d\mathbf{u}}{\int_{\Delta_{k-1}\cap E_{k-1}}\mathcal{D}(\mathbf{u})d\mathbf{u}}\geq\frac{1}{3\epsilon_k L}. \quad (4.17)$$

The proof of the theorem will be complete by combining (4.17) Lemma 4.13 below. □

We now give the lemma which allows us to replace the ratio of the integrals $\frac{\int_{\Delta_{k-1}}\mathcal{N}(\mathbf{u})d\mathbf{u}}{\int_{\Delta_{k-1}}\mathcal{D}(\mathbf{u})d\mathbf{u}}$ by the integrations over the sub-domain $\Delta_{k-1}\cap E_{k-1}$ as in (4.14).

Lemma 4.13. *Let $\mathcal{N}(\boldsymbol{u})$ and $\mathcal{D}(\boldsymbol{u})$ be as given above. Then, we have*

$$\frac{\big|\int_{\Delta_{k-1}\cap E_{k-1}^c}\mathcal{N}(\boldsymbol{u})d\boldsymbol{u}\big|}{\big|\int_{\Delta_{k-1}\cap E_{k-1}}\mathcal{N}(\boldsymbol{u})d\boldsymbol{u}\big|}<\frac{1}{3}, \qquad \frac{\big|\int_{\Delta_{k-1}\cap E_{k-1}^c}\mathcal{D}(\boldsymbol{u})d\boldsymbol{u}\big|}{\big|\int_{\Delta_{k-1}\cap E_{k-1}}\mathcal{D}(\boldsymbol{u})d\boldsymbol{u}\big|}<\frac{1}{3}. \quad (4.18)$$

Proof. Since the word w^* chosen above has all the properties we used in Theorem 3.4, the first inequality in (4.18) follows as a direct consequence of Proposition 3.3 and Theorem 3.4. For the second inequality, since it involves one more power of $|\Delta_{\mathbf{u},i}\gamma|$, we first need a modified version of Proposition 3.3.

Let $\mathcal{E}_{k-1}$ be the same set as defined in Proposition 3.3, so we have

$$\prod_{j=1}^{k} |\Delta_{\mathbf{u},j}\gamma|^{2n} < e^{-n} \prod_{j=1}^{k} |\Delta_{\mathbf{v},j}|^{2n}, \qquad \forall \mathbf{u} \in \Delta_{k-1} \cap E_{k-1}^c, \ \ \mathbf{v} \in \mathcal{E}_{k-1}.$$

On the other hand, similar as in Lemma 3.1, if $\delta(\frac{2}{k}) < \frac{L}{2}$ and $\mathbf{v} \in \mathcal{E}_{k-1}$, we will have

$$|\Delta_{\mathbf{v},i}\gamma| \geq \frac{L - \delta(\frac{2}{k})}{2k} \geq \frac{L}{4k} \geq \frac{|\Delta_{\mathbf{u},i}\gamma|}{4k}.$$

This then implies

$$|\Delta_{\mathbf{u},i}\gamma| \prod_{j=1}^{k} |\Delta_{\mathbf{u},j}\gamma|^{2n} < 4ke^{-n} |\Delta_{\mathbf{v},i}\gamma| \prod_{j=1}^{k} |\Delta_{\mathbf{v},j}\gamma|^{2n} \tag{4.19}$$

for all $\mathbf{u} \in \Delta_{k-1} \cap E_{k-1}^c$ and $\mathbf{v} \in \mathcal{E}_{k-1}$. We now arrive at the same situation as in Proposition 3.3 except that there is one more factor of $4k$ on the right hand side. Integrating both sides above in their respective domains and then enlarging $\mathcal{E}_{k-1}$ to Δ_{k-1}, we get

$$\int_{\Delta_{k-1}\cap E_{k-1}^c} \prod_{j=1}^{k} |\Delta_{\mathbf{u},j}\gamma|^{2n+\delta_{i,j}} \, d\mathbf{u} < 4ke^{3k\log k - n} \int_{\Delta_{k-1}} \prod_{j=1}^{k} |\Delta_{\mathbf{u},j}\gamma|^{2n+\delta_{i,j}} \, d\mathbf{u}, \tag{4.20}$$

where $\delta_{i,j} = 1$ if $j = i$ and is 0 otherwise. We are now ready to prove the second inequality in (4.18). By the expression (4.13), we have the pointwise bound

$$|\mathcal{D}(\mathbf{u})| \leq L^{k-1} \prod_{j=1}^{k} |\Delta_{\mathbf{u},j}\gamma|^{2n+\delta_{i,j}}, \qquad \forall \mathbf{u} \in \Delta_{k-1},$$

so it follows from (4.20) that

$$\int_{\Delta_{k-1}\cap E_{k-1}^c} \mathcal{D}(\mathbf{u}) d\mathbf{u} < 4kL^{k-1}e^{3k\log k - n} \int_{\Delta_{k-1}} \prod_{j=1}^{k} |\Delta_{\mathbf{u},j}\gamma|^{2n+\delta_{i,j}} \, d\mathbf{u}. \tag{4.21}$$

As for the integration of $\mathcal{D}(\mathbf{u})$ in $\Delta_{k-1} \cap E_{k-1}$, we first note from (4.16) that $\Delta_{\mathbf{u},i}x$ is positive with the lower bound

$$\Delta_{\mathbf{u},i}x \geq \eta_k(u_i - u_{i-1}) \geq \frac{\eta_k}{L} |\Delta_{\mathbf{u},i}\gamma|,$$

and hence

$$(\Delta_{\mathbf{u},i}x + \Delta_{\mathbf{u},i}y)^{2n+1} + (\Delta_{\mathbf{u},i}x - \Delta_{\mathbf{u},i}y)^{2n+1} \geq (\Delta_{\mathbf{u},i}x)|\Delta_{\mathbf{u},i}\gamma|^{2n} \geq \frac{\eta_k}{L} |\Delta_{\mathbf{u},i}\gamma|^{2n+1}.$$

Thus, by the choice of w^* and Proposition 4.9, we have

$$|\mathcal{D}(\mathbf{u})| \geq \frac{L^{k-1}}{6^k\sqrt{k}} \prod_{j=1}^{k} |\Delta_{\mathbf{u},j}\gamma|^{2n+\delta_{i,j}}, \quad \forall \mathbf{u} \in \Delta_{k-1} \cap E_{k-1},$$

where we have used $\eta_k \geq \frac{L}{\sqrt{k}}$. Also, since the choice of w^* and the positivity of $\Delta_{\mathbf{u},i}x$ guarantee that $\mathcal{D}(\mathbf{u})$ does not change its sign in the domain $\Delta_{k-1} \cap E_{k-1}$, we can change the order of the absolute value sign and integration to get

$$\left| \int_{\Delta_{k-1}\cap E_{k-1}} \mathcal{D}(\mathbf{u}) d\mathbf{u} \right| \geq \frac{L^{k-1}}{6^k\sqrt{k}} \int_{\Delta_{k-1}\cap E_{k-1}} \prod_{j=1}^{k} |\Delta_{\mathbf{u},j}\gamma|^{2n+\delta_{i,j}} d\mathbf{u}.$$

Applying (4.20), we can enlarge the domain of integration on the right hand side above to Δ_{k-1} and obtain

$$\left| \int_{\Delta_{k-1}\cap E_{k-1}} \mathcal{D}(\mathbf{u}) d\mathbf{u} \right| \geq \frac{L^{k-1}}{6^k\sqrt{k}} \big(1 - 4ke^{3k\log k - n}\big) \int_{\Delta_{k-1}} \prod_{j=1}^{k} |\Delta_{\mathbf{u},j}\gamma|^{2n+\delta_{i,j}} d\mathbf{u}. \quad (4.22)$$

Combining (4.21) and (4.22), and taking $n = 4k^2 \log k$, we can conclude the proof of this lemma as well as Theorem 4.11. □

So far, we have recovered the unsigned directions $\{\rho_j\}$ as well the signs a_j^x, a_j^y's. This already gives us a unit-length piecewise linear path

$$\zeta = \frac{1}{k}\big(\theta_1 * \cdots * \theta_k\big), \quad (4.23)$$

where $\theta_j = \big(a_j^x \rho_j, a_j^y(1-\rho_j)\big)$ for every $j = 1, \ldots, k$. Note that we have an abuse use of notation here as the j-th piece of θ above should really be the *path* $\theta_j t$ for $t \in [\frac{j-1}{k}, \frac{j}{k}]$ instead of the *direction* $\frac{1}{k}\theta_j$. But we choose to stay with (4.23) for notational simplicity. We now end this subsection by showing that the piecewise linear ζ is close to γ when the latter is normalized to have length 1.

Corollary 4.14. *Fix an integer $k \geq 4$ such that $\eta_k < \frac{L}{32}$. For each $j = 1, \ldots, k$, let ρ_j, a_j^x and a_j^y be obtained as in Corollary 4.6 and Definition 4.10. For each $j = 1, \cdots, k$, let*

$$\theta_j = \big(a_j^x \rho_j, a_j^y(1-\rho_j)\big).$$

Then, the signed directions θ_j's satisfy

$$\left| \theta_j - \left(\frac{\Delta_j x}{|\Delta_j \gamma|}, \frac{\Delta_j y}{|\Delta_j \gamma|} \right) \right| < \frac{12\eta_k}{L} \quad (4.24)$$

for every $j = 1, \ldots, k$. As a consequence, we have

$$\sup_{1\leq j\leq k} \sup_{u\in[\frac{j-1}{k}, \frac{j}{k}]} \big|\theta_j - \frac{\dot{\gamma_u}}{L}\big| < \frac{16\eta_k}{L}.$$

Proof. Since for each j, we have

$$\sup_{u\in[\frac{j-1}{k}, \frac{j}{k}]} \left| \frac{\dot{\gamma_u}}{L} - \left(\frac{\Delta_j x}{|\Delta_j \gamma|}, \frac{\Delta_j y}{|\Delta_j \gamma|} \right) \right| < \frac{4\eta_k}{L},$$

it suffices to prove the inequality (4.24). If $r_j \in [\frac{2\eta_k}{L}, 1 - \frac{2\eta_k}{L}]$, then by Definition 4.10 and Theorem 4.11, a_j^x and a_j^y have the same signs as $\Delta_j x$ and $\Delta_j y$, respectively. Thus, by Corollary 4.6, we have

$$\left| a_j^x \rho_j - \frac{\Delta_j x}{|\Delta_j \gamma|} \right| = |\rho_j - r_j| < \frac{4\eta_k}{L}, \quad (4.25)$$

and the same bound is true for $\left|a_j^y(1-\rho_j) - \frac{\Delta_j y}{|\Delta_j \gamma|}\right|$. This proves (4.24) when $r_j \in [\frac{2\eta_k}{L}, 1-\frac{2\eta_k}{L}]$.

If $r_j > 1 - \frac{2\eta_k}{L}$, the bound (4.25) for the x-direction is still true, but the bound for the y-direction becomes

$$\left|a_j^y(1-\rho_j) - \frac{\Delta_j y}{|\Delta_j \gamma|}\right| \leq (1-\rho_j) + (1-r_j) < \frac{8\eta_k}{L},$$

since Corollary 4.6 forces ρ_j to be bigger than $1 - \frac{6\eta_k}{L}$. Combining this bound with (4.25), we get (4.24) when $r_j > 1 - \frac{2\eta_k}{L}$. The case when $r_j < \frac{2\eta_k}{L}$ follows in the same way. This completes the proof. □

4.3 Length

We have now recovered from the signature the piecewise linear path ζ with unit length, which is shown in Corollary 4.14 to be close to the "normalized" γ. Thus, the only remaining quantity to be determined is $\tilde{L}$, which is expected to approximate the ℓ^1 length of γ. We can achieve this by a simple scaling argument.

Let ζ be the unit piecewise linear path as in (4.23). Let m be the smallest integer such that $C_\gamma(\tilde{w}) \neq 0$ for some $|\tilde{w}| = m$, where $C_\gamma(\tilde{w})$ is the coefficient of the word $\tilde{w}$ in the signature of γ. Fix that word $\tilde{w}$, and set

$$\tilde{L} := \left(\frac{C_\gamma(\tilde{w})}{C_\zeta(\tilde{w})}\right)^{\frac{1}{m}}. \tag{4.26}$$

Note that it is not obvious from the above expression that $\tilde{L}$ is always well defined, as the denominator might just be 0, or $C_\zeta(\tilde{w})$ has a different sign with $C_\gamma(\tilde{w})$ and m is even. However, it turns out that for all sufficiently large k, $\tilde{L}$ defined by (4.26) does make sense, and is in fact close to the true length L. This is the content of the following theorem.

Theorem 4.15. *Let k be large enough such that*

$$\eta_k < \frac{(m-1)!}{32L^{m-1}}|C_\gamma(\tilde{w})|. \tag{4.27}$$

Then $\tilde{L}$ introduced in (4.26) *is well defined, and satisfies*

$$|\tilde{L} - L| < \frac{32L^m\eta_k}{(m-1)!|C_\gamma(\tilde{w})|}. \tag{4.28}$$

Proof. If θ is at natural parametrization, then we have

$$\dot{\zeta}_v = \theta_j, \qquad v \in (\frac{j-1}{k}, \frac{j}{k}).$$

If γ is also at natural parametrization, then for almost every $\mathbf{u} = (u_1, \ldots, u_m) \in \Delta_m$, we have

$$\left|\frac{1}{L^m}\prod_{j=1}^m \dot{\gamma}_{u_j}^{i_j} - \prod_{j=1}^m \dot{\zeta}_{u_j}\right| \leq \frac{16m\eta_k}{L},$$

where we have used Corollary 4.14 and the fact that both $\frac{|\dot{\gamma}|}{L}$ and $|\dot{\zeta}|$ are uniformly bounded by 1. Thus, the difference between the signatures $\frac{1}{L^m}C_\gamma(\tilde{w})$ and $C_\zeta(\tilde{w})$ can be bounded by

$$\left|\frac{1}{L^m}C_\gamma(\tilde{w}) - C_\zeta(\tilde{w})\right| \leq \frac{16\eta_k}{(m-1)!L}. \tag{4.29}$$

Note that (4.27) and (4.29) together imply $C_\zeta(\tilde{w})$ has the same sign as $C_\gamma(\tilde{w})$, and $|C_\zeta(\tilde{w})| > \frac{|C_\gamma(\tilde{w})|}{2L^m}$. In particular, this shows $\tilde{L}$ in (4.26) is well defined and positive, so we have

$$|\tilde{L}^m - L^m| \geq L^{m-1}|\tilde{L} - L|. \tag{4.30}$$

On the other hand, the lower bound $|C_\zeta(\tilde{w})| > \frac{|C_\gamma(\tilde{w})|}{2L^m}$ implies

$$|\tilde{L}^m - L^m| = \left|\frac{C_\gamma(\tilde{w})}{C_\zeta(\tilde{w})} - L^m\right| \leq \frac{16L^{m-1}\eta_k}{(m-1)!|C_\theta(\tilde{w})|} < \frac{32L^{2m-1}\eta_k}{(m-1)!|C_\gamma(\tilde{w})|}. \tag{4.31}$$

Combining (4.30) and (4.31), we obtain

$$|\tilde{L} - L| < \frac{32L^m\eta_k}{(m-1)!|C_\gamma(\tilde{w})|}.$$

This proves the theorem. □

Remark 4.16. Note that both the assumption (4.27) and the bound (4.28) give the correct scaling in length, since η_k scales linearly in L, and $|C_\gamma(\tilde{w})|$ scales as L^m.

4.4 Summary of the procedure

We now summarize the inversion procedure developed in this section, and give its validity as well as stability properties. Let k be a fixed large number whose value will be specified in Theorem 4.17 below, and let $n = 4k^2 \log k$. Recall the definition of η_k from (1.4). Also recall the definitions of the set $\mathcal{L}_k^n$ from (2.1) and the symmetrized signatures $\mathcal{S}_k^{2n}$, $\mathcal{S}_{k,j,x}^{2n}$, $\mathcal{S}_{k,j,y}^{2n}$ from (2.3) and (4.8). The inversion procedure includes choosing for each $j = 1, \ldots, k$ a real number $\rho_j \in [0,1]$, $a_j^x, a_j^y \in \{-1, 1\}$ and $\tilde{L} \in \mathbb{R}^+$ in the following way.

1. For each $j = 1, \ldots, k$, choose $\rho_j \in [0,1]$ according to Corollary 4.6 such that

$$\Bigg(\sum_{|w|=k-1} \sum_{\boldsymbol{\ell}:|\frac{\ell_j}{n}-\rho_j|\leq\frac{2\eta_k}{L}} |\mathcal{S}_k^{2n}(w,\boldsymbol{\ell})|\Bigg) \Bigg/ \Bigg(\sum_{|w|=k-1} \sum_{\boldsymbol{\ell}\in\mathcal{L}_k^n} |\mathcal{S}_k^{2n}(w,\boldsymbol{\ell})|\Bigg) > \frac{1}{2}, \tag{4.32}$$

   where in the numerator the sum over $\boldsymbol{\ell}$ is restricted to the indicated subset of $\mathcal{L}_k^n$. Any set $\{\rho_j\}$ satisfying (4.32) can be used[3].

2. Now we choose a word $w^* = e_{i_1} \cdots e_{i_{k-1}}$ by setting

$$e_{i_j} = x, \qquad \text{if } \rho_j \geq \frac{1}{2};$$
$$e_{i_j} = y, \qquad \text{if } \rho_j < \frac{1}{2},$$

   and determine the signs a_j^x's by

$$a_j^x = 1, \qquad \text{if } \frac{\sum_{\boldsymbol{\ell}\in\mathcal{L}_k^n} \mathcal{S}_k^{2n}(w^*,\boldsymbol{\ell})}{\sum_{\boldsymbol{\ell}\in\mathcal{L}_k^n} \mathcal{S}_{k,j,x}^{2n}(w^*,\boldsymbol{\ell})} \geq 0;$$
$$a_j^x = -1, \qquad \text{if } \frac{\sum_{\boldsymbol{\ell}\in\mathcal{L}_k^n} \mathcal{S}_k^{2n}(w^*,\boldsymbol{\ell})}{\sum_{\boldsymbol{\ell}\in\mathcal{L}_k^n} \mathcal{S}_{k,j,x}^{2n}(w^*,\boldsymbol{\ell})} < 0,$$

[3]Note that this step uses the knowledge of $\frac{\eta_k}{L}$, which is not directly available from the signature. However, we can still circumvent the problem even if we do not know $\frac{\eta_k}{L}$, at the cost of a lower accuracy of the inversion procedure. See Remark 4.8 for a discussion on this.

where w^* is the word chosen above. The signs a_j^y's are determined in the same way except one replaces $\mathcal{S}^{2n}_{k,j,x}(w^*, \boldsymbol{\ell})$ by $\mathcal{S}^{2n}_{k,j,y}(w^*, \boldsymbol{\ell})$. These two steps already produce a piecewise linear path

$$\zeta = \frac{1}{k}\big(\theta_1 * \cdots * \theta_k\big) \tag{4.33}$$

as in (4.23), where $\theta_j = \big(a_j^x \rho_j, a_j^y(1-\rho_j)\big)$ for each $j = 1, \ldots, k$.

3. Let m be the smallest integer such that $C_\gamma(\tilde{w}) \neq 0$ for some $|\tilde{w}| = m$, and we determine the length $\tilde{L}$ by setting

$$\tilde{L} := \left(\frac{C_\gamma(\tilde{w})}{C_\zeta(\tilde{w})}\right)^{\frac{1}{m}}, \tag{4.34}$$

where θ is the path chosen in (4.33) above, and $C_\gamma(\tilde{w})$ and $C_\zeta(\tilde{w})$ are the coefficients of $\tilde{w}$ in the signature of γ and θ, respectively.

The above three steps produce a piecewise linear path $\tilde{\gamma}$ of the form

$$\tilde{\gamma} = \tilde{L}\zeta = \frac{\tilde{L}}{k}\big(\theta_1 * \cdots \theta_k\big).$$

Note that given any set $\rho_j \in [0,1]$, the choices for a_j^x and a_j^y are always well defined. For large enough k, the choices for the parameters $\{\rho_j\}$ and $\tilde{L}$ are also well defined, and the path $\tilde{\gamma}$ will turn out to be close to the original path γ in Lipschitz norm.

We first quantify how large k needs to be. All the statements before Section 4.3 (recovery of length) hold true when $\eta_k < \frac{L}{32}$. In order for $\tilde{L}$ to be well defined (Theorem 4.15), one needs η_k to satisfy the condition (4.27). But since

$$|C_\gamma(w)| \leq \frac{L^m}{m!}$$

for every m and every $|w| = m$, so the assumption (4.27) implies $\eta_k < \frac{L}{32}$. We are now ready to state our main theorem.

Theorem 4.17. *Let m be the smallest integer such that $C_\gamma(\tilde{w}) \neq 0$ for some $|\tilde{w}| = m$, and we fix that word $\tilde{w}$. Let $k \geq 4$ be large enough such that*

$$\eta_k < \frac{(m-1)!}{32L^{m-1}}|C_\gamma(\tilde{w})|,$$

and let $n = 4k^2 \log k$. Then, the above choices of parameters are well defined. In addition, when γ is at natural parametrization, we have the bound

$$\sup_{1\leq j\leq k} \sup_{u\in[\frac{j-1}{k}, \frac{j}{k}]} |\dot{\gamma}_u - \tilde{L}\theta_j| < 16\eta_k\big(1 + \frac{2L^m}{(m-1)!|C_\gamma(\tilde{w})|}\big),$$

where $\theta_j = \big(a_j^x \rho_j, a_j^y(1-\rho_j)\big)$.

Proof. By Corollary 4.14, for every $j = 1, \ldots, k$, we have

$$\sup_{u\in[\frac{j-1}{k}, \frac{j}{k}]} \big|\dot{\gamma}_u - L\theta_j\big| < 16\eta_k.$$

By the bound on $|\tilde{L} - L|$ from Theorem 4.15 and that $|\theta_j| = 1$, we get

$$|\dot{\gamma}_u - \tilde{L}\theta_j| < 16\eta_k\big(1 + \frac{2L^m}{(m-1)!|C_\gamma(w)|}\big),$$

which holds for all $u \in [\frac{j-1}{k}, \frac{j}{k}]$ and all $j = 1, \ldots, k$. This completes the proof of the main theorem. ∎

5 Higher dimensions

We now give a brief explanation of how the symmetrization procedure extends to the recovery of $\mathcal{C}^1$ paths in dimension higher than 2 from their signatures. Let $\gamma = (\gamma^1, \dots, \gamma^d)$ be a d-dimensional $\mathcal{C}^1$ path at natural parametrization. For large enough integer k and each $1 \le j \le k$, we need to reconstruct from the signature of γ a non-negative vector $\boldsymbol{\rho}_j = (\rho_j^1, \dots, \rho_j^d)$ such that $\sum_i \rho_j^i = 1$, the signs $a_j^i \in \{\pm 1\}$ for $1 \le i \le d$, and $\tilde{L} \in \mathbb{R}^+$ that approximates the ℓ^1 length L of γ.

Same as the 2-dimensional case, we still symmetrize k blocks of size $2n$ with one letter separating consecutive blocks. To set up, for every positive integer k and n, we let $\mathcal{L}_k^n$ be the set of multi-component multi-indices

$$\mathcal{L}_k^n := \big\{\boldsymbol{\ell} = (\boldsymbol{\ell}_1, \dots, \boldsymbol{\ell}_k) : \ \boldsymbol{\ell}_j = (\ell_j^1, \dots, \ell_j^d), \sum_{i=1}^d \ell_j^i = n, \forall j = 1, \dots, k\big\}.$$

For every word $w = e_{i_1} \cdots e_{i_{k-1}}$ and every $\boldsymbol{\ell} \in \mathcal{L}_k^n$, we define $\mathcal{W}_k^{2n}(w, \boldsymbol{\ell})$ to be the set of words of the form

$$\mathcal{W}_k^{2n}(w, \boldsymbol{\ell}) := \big\{w' = w_1 * e_{i_1} * \cdots * e_{i_{k-1}} * w_k : |w_j|_{e_i} = 2\ell_j^i, \forall i, j\big\},$$

where $e_1, \dots, e_d$ are standard basis elements of $\mathbb{R}^d$. Note that the assumption $\boldsymbol{\ell} \in \mathcal{L}_k^n$ together with $|w_j|_{e_i} = 2\ell_j^i$ imply $|w_j| = 2n$ for every j. Also, for every integer $1 \le p \le k$ and $1 \le q \le d$, we let $\mathcal{W}_{k,p,q}^{2n}(w, \boldsymbol{\ell})$ be the set of words

$$w' = w_1 * e_{i_1} * \cdots * e_{i_{k-1}} * w_k$$

such that $|w_j|_{e_i} = 2\ell_j^i$ if $j \ne p$ or $i \ne q$, but $|w_p|_{e_q} = 2\ell_p^q + 1$. In other words, the definition for $\mathcal{W}_{k,p,q}^{2n}(w, \boldsymbol{\ell})$ is the same as $\mathcal{W}_k^{2n}(w, \boldsymbol{\ell})$ except that $|w_p|_{e_q} = 2\ell_p^q + 1$, and as a consequence we have $|w_p| = 2n + 1$.

Similar to (2.3) and (4.8), we define the symmetrized signatures $\mathcal{S}_k^{2n}(w, \boldsymbol{\ell})$ and $\mathcal{S}_{k,p,q}^{2n}(w, \boldsymbol{\ell})$ by

$$\begin{aligned} \mathcal{S}_k^{2n}(w, \boldsymbol{\ell}) &= \big((2n)!\big)^k \sum_{w' \in \mathcal{W}_k^{2n}(w, \boldsymbol{\ell})} C(w'), \\ \mathcal{S}_{k,p,q}^{2n}(w, \boldsymbol{\ell}) &= (2n+1)\big((2n)!\big)^k \sum_{w' \in \mathcal{W}_{k,p,q}^{2n}(w, \boldsymbol{\ell})} C(w'). \end{aligned} \tag{5.1}$$

Finally, for every integer k, we define ϵ_k and η_k by

$$\epsilon_k := \sqrt{2}\bigg(\sqrt{\frac{(d-1)\delta(\frac{1}{k})}{L}} + \frac{1}{\sqrt{k}}\bigg), \qquad \eta_k := \delta(3\epsilon_k) + \frac{L}{\sqrt{k}}, \tag{5.2}$$

where L is the ℓ^1 length of γ, and δ is the modulus of continuity of $\dot{\gamma}$. Now, we let k to be a fixed integer whose value will be specified later, and let $n = 2dk^2 \log k$. Similar to the 2-dimensional case, we have the following inversion algorithm to reconstruct γ from its signature.

1. For each $1 \le j \le k$, we choose the unsigned direction

$$\boldsymbol{\rho}_j = (\rho_j^1, \dots, \rho_j^d) \quad \text{s.t.} \quad \rho_j^i \ge 0, \quad \sum_{i=1}^d \rho_j^i = 1$$

according to the criterion

$$\Big(\sum_{w:|w|=k-1}\sum_{\boldsymbol{\ell}:\sup_i|\frac{\ell^i_j}{n}-\rho^i_j|<\frac{3d\eta_k}{L}}\big|\mathcal{S}^{2n}_k(w,\boldsymbol{\ell})\big|\Big)\Big/\Big(\sum_{w:|w|=k-1}\sum_{\boldsymbol{\ell}\in\mathcal{L}^n_k}\big|\mathcal{S}^{2n}_k(w,\boldsymbol{\ell})\big|\Big)>\frac{1}{2}.$$

Any element $\boldsymbol{\rho}_j$ satisfying the above will suffice.

2. We choose the word $w^*=e_{i_1}\cdots e_{i_{k-1}}$ by setting

$$e_{i_j}=e_i,\qquad \text{if } \rho^i_j\geq\rho^q_j,\quad \forall q\neq i.$$

If there are two or more maximizers, then any of them is suitable. We then determine the signs $a^i_j\in\{\pm1\}$ by

$$a^i_j=1,\qquad \text{if}\quad \frac{\sum_{\boldsymbol{\ell}}\mathcal{S}^{2n}_k(w^*,\boldsymbol{\ell})}{\sum_{\boldsymbol{\ell}}\mathcal{S}^{2n}_{k,j,i}(w^*,\boldsymbol{\ell})}\geq 0,$$

$$a^i_j=-1,\qquad \text{if}\quad \frac{\sum_{\boldsymbol{\ell}}\mathcal{S}^{2n}_k(w^*,\boldsymbol{\ell})}{\sum_{\boldsymbol{\ell}}\mathcal{S}^{2n}_{k,j,i}(w^*,\boldsymbol{\ell})}<0,$$

where all the sums are taken over $\boldsymbol{\ell}\in\mathcal{L}^n_k$.

3. The recovery of length is exactly the same as the 2-dimensional case. Let

$$\theta_j=\big(a^1_j\rho^1_j,\ldots,a^d_j\rho^d_j\big)$$

be the unit vector (in ℓ^1 sense) obtained from the previous two steps, and let

$$\zeta:=\frac{1}{k}(\theta_1*\cdots*\theta_k).$$

In other words, ζ is a piecewise linear path whose j-th piece is in the direction θ_j and has length $\frac{1}{k}$. We also let m be the smallest integer such that $C_\gamma(\tilde{w})\neq 0$ for some $|\tilde{w}|=m$, and determine $\tilde{L}$ by

$$\tilde{L}:=\left(\frac{C_\gamma(\tilde{w})}{C_\zeta(\tilde{w})}\right)^{\frac{1}{m}},$$

where $\tilde{w}$ can be any word with length m such that $C_\gamma(\tilde{w})\neq 0$.

Following the arguments for the 2-dimensional case, we can get the following theorem, which gives the stability of the inversion algorithm for $\mathcal{C}^1$ paths in $\mathbb{R}^d$.

Theorem 5.1. *Let m be the smallest integer such that $C_\gamma(\tilde{w})\neq 0$ for some $|\tilde{w}|=m$. Let $k\geq 2d$ be large enough such that*

$$\eta_k<\frac{(m-1)!|C_\gamma(\tilde{w})|}{16(d+1)L^{m-1}},$$

and let $n=2dk^2\log k$. Then for every $1\leq j\leq k$, the above choices of the parameters $\boldsymbol{\rho}_j$, $\{a^i_j\}_{i=1}^d$ and $\tilde{L}$ are all well defined, and when γ is at natural parametrization, we have the bound

$$\sup_{1\leq i\leq d}\sup_{1\leq j\leq k}\sup_{u\in[\frac{j-1}{k},\frac{j}{k}]}|\dot{\gamma}^i_u-\tilde{L}a^i_j\rho^i_j|<8(d+1)\big(1+\frac{2L^m}{(m-1)!|C_\gamma(w)|}\big)\eta_k.$$

In other words, the reconstructed path

$$\tilde{L}\zeta=\frac{\tilde{L}}{k}\big(\theta_1*\cdots*\theta_k\big)$$

is η_k-close to the original path γ in Lipschitz norm.

Remark 5.2. The proportionality constants in Theorem 4.17 for the 2-dimensional case are better than the ones in Theorem 5.1 by simply setting $d = 2$. This is because when $d = 2$, the assumption $|\dot{x}| + |\dot{y}| \equiv L$ implies that in many cases (for example, Lemma 4.2), the same optimal estimates hold for both x and y directions, while this is not true in general case when $d \geq 3$.

References

[BG15] H. Boedehardjo and X. Geng. The uniqueness of signature problem in the non-Markov setting. *Stoch. Proc. Appl.* **125**, no. 12, (2015), 4674–4701.

[BGLY16] H. Boedihardjo, X. Geng, T. Lyons, and D. Yang. The signature of a rough path: uniqueness. *Adv. Math.* **293**, (2016), 720–737.

[Che54] K.-T. Chen. Iterated integrals and exponential homomorphisms. *Proc. London Math. Soc.* **4**, no. 3, (1954), 502–512.

[Che57] K.-T. Chen. Integration of paths, geometric invariants and a generalized Baker-Hausdorff formula. *Ann. of Math.* **65**, no. 1, (1957), 163–178.

[Che58] K.-T. Chen. Integration of paths - a faithful representation of paths by noncommutative formal power series. *Trans. Amer. Math. Soc.* **89**, no. 2, (1958), 395–407.

[Gen15] X. Geng. Reconstruction for the signature of a rough path. *Arxiv e-prints* (2015). `arXiv:1508.06890`.

[GQ16] X. Geng and Z. Qian. On an inversion theorem for Stratonovich's signatures of multidimensional diffusion paths. *Ann. Inst. Henri Poincaré B* **52**, no. 1, (2016), 429–447.

[HL10] B. Hambly and T. Lyons. Uniqueness for the signature of a path of bounded variation and the reduced path group. *Ann. of Math.* **171**, no. 1, (2010), 109–167.

[LJQ13] Y. Le Jan and Z. Qian. Stratonovich's signatures of Brownian motion determine Brownian sampel paths. *Probab. Theory Related Fields* **157**, no. 1-2, (2013), 209–223.

[LX15] T. Lyons and W. Xu. Hyperbolic development and inversion of the signature. *Arxiv e-prints* (2015). `arXiv:1507.00286`.

[Lyo98] T. Lyons. Differential equations driven by rough signals. *Rev. Mat. Iberoam.* **14**, no. 2, (1998), 215–310.

Mathematical and Oxford-Man Institutes, University of Oxford, Woodstock road, Oxford, OX2 6GG, UK.

Email: tlyons@maths.ox.ac.uk

Mathematics Institute, University of Warwick, Coventry, CV4 7AL, UK.

Email: weijun.xu@warwick.ac.uk